

Université de Montréal

**Archivage du Web organisationnel
dans une perspective archivistique**

par

Aïda Chebbi

École de bibliothéconomie et des sciences de l'information

Faculté des arts et des sciences

Thèse présentée à la Faculté des études supérieures
en vue de l'obtention du grade de Philosophiæ Doctor (Ph.D.)
en sciences de l'information

décembre, 2012

© Aïda Chebbi, 2012

Résumé

Le Web représente actuellement un espace privilégié d'expression et d'activité pour plusieurs communautés, où pratiques communicationnelles et pratiques documentaires s'enrichissent mutuellement. Dans sa dimension visible ou invisible, le Web constitue aussi un réservoir documentaire planétaire caractérisé non seulement par l'abondance de l'information qui y circule, mais aussi par sa diversité, sa complexité et son caractère éphémère.

Les projets d'archivage du Web en cours abordent pour beaucoup cette question du point de vue de la préservation des publications en ligne sans la considérer dans une perspective archivistique. Seuls quelques projets d'archivage du Web visent la préservation du Web organisationnel ou gouvernemental. La valeur archivistique du Web, notamment du Web organisationnel, ne semble pas être reconnue malgré un effort soutenu de certaines archives nationales à diffuser des politiques d'archivage du Web organisationnel.

La présente thèse a pour but de développer une meilleure compréhension de la nature des archives Web et de documenter les pratiques actuelles d'archivage du Web organisationnel. Plus précisément, cette recherche vise à répondre aux trois questions suivantes :

- (1) Que recommandent en général les politiques d'archivage du Web organisationnel?
- (2) Quelles sont les principales caractéristiques des archives Web?
- (3) Quelles pratiques d'archivage du Web organisationnel sont mises en place dans des organisations au Québec?

Pour répondre à ces questions, cette recherche exploratoire et descriptive a adopté une approche qualitative basée sur trois modes de collecte des données, à savoir : l'analyse d'un corpus de 55 politiques et documents complémentaires relatifs à l'archivage du Web organisationnel; l'observation de 11 sites Web publics d'organismes au Québec de même que l'observation d'un échantillon de 737 documents produits par ces systèmes Web; et, enfin, des entrevues avec 21 participants impliqués dans la gestion et l'archivage de ces sites Web.

Les résultats de recherche démontrent que les sites Web étudiés sont le produit de la conduite des activités en ligne d'une organisation et documentent, en même temps, les objectifs et les manifestations de sa présence sur le Web. De nouveaux types de documents propres au

Web organisationnel ont pu être identifiés. Les documents qui ont migré sur le Web ont acquis un autre contexte d'usage et de nouvelles caractéristiques. Les méthodes de gestion actuelles doivent prendre en considération les propriétés des documents dans un environnement Web.

Alors que certains sites d'étude n'archivent pas leur site Web public, d'autres s'y investissent. Toutefois les choix établis ne correspondent pas toujours aux recommandations proposées dans les politiques d'archivage du Web analysées et ne garantissent pas la pérennité des archives Web ni leur exploitabilité à long terme.

Ce constat nous a amenée à proposer une politique type adaptée aux caractéristiques des archives Web. Ce modèle décrit les composantes essentielles d'une politique pour l'archivage des sites Web ainsi qu'un éventail des mesures que pourrait mettre en place l'organisation en fonction des résultats d'une analyse des risques associés à l'usage de son site Web public dans la conduite de ses affaires.

Mots-clés : Analyse des risques, archives Web, gestion des archives, méthodes d'archivage du Web, politique d'archivage, pratiques archivistiques, approche d'archivage du Web, techniques d'archivage du Web, types des archives, Web organisationnel

Abstract

Several communities have adopted the web as a privileged space for expression and activity, where communication and documentary practices complement and enrich each other. Both in its visible and invisible dimensions, the web is a documentation vault of planetary scope, characterized not only by the sheer amount of information it contains, but by its diversity, complexity and ephemeral nature.

Most current web archiving projects focus on preserving online publications without considering their value from an archival point of view. Only a few web archiving projects target the preservation of organisational or governmental web records. The archival value of the web, especially that of organisational web, does not seem to be justly recognized, despite the continuous effort deployed by certain National Archives bodies to disseminate best practices and policies for organisational web archiving.

This thesis aims to develop a better understanding of the nature of web records and to document current archiving practices intended for organisational web. In particular, this research will look at the following three questions:

- (1) What general recommendations can be found in archiving policies for organisational web?
- (2) What are the main characteristics of web records?
- (3) Which web record-keeping practices have been deployed in Quebec organisations?

To address these questions, this exploratory and descriptive research uses a qualitative approach based on three data collection methods, namely: the analysis of a body of 55 policies and supporting documents related to organisational web archiving; the scrutiny of 11 public websites of various Quebec organisations and of a sample of 737 documents generated by those web systems; and interviews with 21 individuals that are involved in the management and archiving of those websites.

The results of this research show that the observed sites are the product of an organisation's online activity and that they simultaneously document the objectives and the occurrences of an organisation's web presence. New types of documents that are specific to

organisational web have been identified. Documents that have migrated online have acquired a different context of use and new characteristics. Hence, current document management methods must consider the unique properties of documents in a web environment.

Only a portion of the observed organisations are involved in the process of archiving their public website. Additionally, the chosen archiving strategies are not always coherent with the recommendations found in web archiving policies, and do not guarantee the sustainability of web records.

Those results led us to design a standard policy model adapted to the particular properties of web archives. This model describes the essential components of a web archiving policy and proposes a range of measures that an organisation could implement based on the results of a risks analysis of their public website's uses in a business context.

Keywords : archival practices, archives management, archives typology, risks analysis, web archiving, web archiving methods, web archiving policy, web archiving approaches; web archiving techniques, web records

Table des matières

Liste des tableaux-----	xii
Liste des figures -----	xiii
Liste des sigles-----	xiv
Remerciements-----	xvi
1. Introduction -----	19
1.1 Problématique-----	19
1.2 But et objectifs de la recherche -----	24
1.3 Questions de recherche -----	24
2. Revue de littérature-----	27
Introduction -----	27
2.1 Définitions-----	27
2.1.1 Concepts technologiques -----	27
2.1.2 Concepts archivistiques-----	29
2.2 Le Web et la production documentaire -----	30
2.2.1 Caractéristiques de la production documentaire sur le Web -----	31
2.2.1.1 Des objets documentaires faciles à produire, difficiles à définir -----	31
2.2.1.2 Des objets documentaires dynamiques, aux contours dilués-----	32
2.2.1.3 Des objets documentaires uniques et éphémères -----	33
2.2.1.4 Des objets documentaires composites difficiles à cerner dans le temps et dans l'espace -----	35
2.2.2 Critères d'identification des archives dans un environnement Web -----	37
2.2.2.1 Caractère organique et consigné-----	37
2.2.2.2 Contexte de création et d'utilisation -----	38
2.2.2.3 Valeurs archivistiques-----	39
2.2.2.4 Conditions de validité des archives -----	40
2.2.2.5 Capture et gestion dans un système d'archivage -----	43
2.3 Stratégies d'archivage du Web -----	44
2.3.1 Perspective documentaire -----	47

2.3.1.1	Modes d'archivage -----	47
2.3.1.1.1	Dépôt légal des ressources en ligne -----	47
2.3.1.1.2	Dépôt volontaire des ressources en ligne -----	49
2.3.1.1.3	Moissonnage à distance du Web public -----	50
2.3.1.1.4	Archivage rétrospectif -----	50
2.3.1.2	Approches d'archivage -----	51
2.3.1.2.1	Approche intégrale ou collecte automatique de tout le Web -----	51
2.3.1.2.2	Approche exhaustive ou collecte de l'ensemble d'un domaine Web -----	52
2.3.1.2.3	Approche sélective ou collecte ciblée -----	52
2.3.1.2.4	Approche intégrée -----	53
2.3.1.2.5	Approche collaborative ou collaboration interinstitutionnelle -----	54
2.3.1.3	Description des principaux projets d'archivage du Web -----	56
2.3.2	Perspective archivistique -----	58
2.3.2.1	Mode de gestion -----	58
2.3.2.1.1	Gestion des risques -----	58
2.3.2.1.2	Gestion du cycle de vie des archives – Records Continuum -----	60
2.3.2.2	Approches d'archivage -----	62
2.3.2.2.1	Approche orientée objet -----	62
2.3.2.2.2	Approche orientée événement -----	63
2.3.2.3	Principaux projets d'archivage du Web organisationnel dans une perspective archivistique -----	63
2.4	Techniques d'archivage du Web -----	65
2.4.1	Techniques recommandées -----	65
2.4.1.1	Archivage à la source -----	65
2.4.1.2	Archivage à distance -----	65
2.4.1.3	Archivage des transactions -----	67
2.4.2	Outils d'archivage des sites Web -----	67
2.4.2.1	Logiciels de moissonnage ou d'acquisition (aspirateurs Web) -----	68
2.4.2.2	Logiciels de stockage et de maintenance -----	68
2.4.2.3	Logiciels d'indexation et de recherche -----	68
2.5	Format d'archivage -----	69

2.6	Accès et exploitation des archives Web -----	69
2.7	Archivage du Web social -----	70
2.7.1	L'archivage du Web social : les principales initiatives -----	72
2.7.2	L'archivage de Twitter -----	73
2.7.3	L'archivage de Facebook -----	74
2.8	Cadre conceptuel pour étudier l'archivage du Web organisationnel -----	75
	Conclusion -----	78
3.	Méthodologie -----	81
	Introduction -----	81
3.1	Approche générale -----	81
3.2	Collecte des données -----	82
3.2.1	Modes de collecte des données -----	82
3.2.1.1	Analyse des politiques d'archivage du Web organisationnel et de la documentation pertinente -----	83
3.2.1.1.1	Analyse des politiques d'archivage du Web organisationnel -----	83
3.2.1.1.2	Documentation pertinente -----	85
3.2.1.2	Observation des sites Web publics -----	86
3.2.1.3	Entrevues en profondeur -----	88
3.2.2	Prétest des instruments de collecte des données -----	90
3.2.3	Chronologie de la collecte des données -----	92
3.2.4	Lien entre les questions de recherche, les modes de collecte et les instruments de collecte -----	92
3.3	Plan d'échantillonnage -----	94
3.3.1	Politiques d'archivage du Web organisationnel -----	95
3.3.1.1	Sélection des politiques -----	95
3.3.1.2	Caractéristiques des politiques -----	97
3.3.2	Sites d'étude -----	99
3.3.2.1	Sélection des sites d'étude -----	99
3.3.2.2	Caractéristiques des sites Web publics analysés -----	100
3.3.3	Participants à l'étude -----	103
3.4	Analyse des données -----	105

3.4.1	Processus d'analyse	105
3.4.2	Unité d'analyse	107
3.4.3	Chronologie de l'analyse des données	108
3.5	Qualité de la recherche	108
3.5.1	Préparation de la recherche	109
3.5.1.1	Habilités de la chercheuse	109
3.5.1.2	Conduite d'un prétest	110
3.5.1.3	Élaboration d'un protocole de recherche	110
3.5.2	Réalisation de la recherche	110
3.5.2.1	Crédibilité de la recherche	110
3.5.2.2	Stabilité des données	111
3.5.2.3	Transférabilité des résultats	112
3.5.2.4	Fiabilité de la recherche	113
	Conclusion	113
4.	Résultats	115
	Introduction	115
4.1	Analyse des politiques d'archivage du Web organisationnel	115
4.1.1	Portée et finalité des politiques analysées	116
4.1.2	Systèmes Web visés par les politiques	119
4.1.3	Intervenants impliqués dans le processus d'archivage du Web organisationnel	121
4.1.4	Unités d'archivage et critères de sélection	124
4.1.5	Stratégies d'archivage et principes de base	128
4.2	Observation des archives Web	134
4.2.1	Caractéristiques des archives Web	135
4.2.1.1	Composantes et fonctionnalités d'une page Web	136
4.2.1.2	Valeurs archivistiques et critères de validité	139
4.2.1.2.1	Valeurs archivistiques	139
4.2.1.2.2	Critères et conditions de validité	146
4.2.2	Modes de création	148
4.2.2.1	Numérisation et mise en ligne des documents	148
4.2.2.2	Création de documents par des suites bureautiques	149

4.2.2.3	Création de documents par des suites multimédias -----	150
4.2.2.4	Création des documents par des éditeurs Web-----	150
4.2.2.5	Création des documents par des systèmes de gestion de contenu -----	150
4.2.3	Typologie des archives Web -----	151
4.2.3.1	Types et catégories des documents d'un site Web organisationnel -----	152
4.2.3.2	Nouvelle typologie des archives Web -----	155
4.3	Analyse des stratégies d'archivage du Web organisationnel-----	158
4.3.1	Intervenants impliqués dans le processus d'archivage-----	159
4.3.1.1	Rôles et responsabilités des webmestres -----	161
4.3.1.2	Rôles et responsabilités des archivistes -----	162
4.3.2	Stratégies d'archivage-----	164
4.3.2.1	Approches d'archivage -----	165
4.3.2.2	Techniques d'archivage-----	167
4.3.2.3	Outils de gestion et d'archivage du Web organisationnel -----	169
4.3.2.4	Exploitation des archives Web-----	171
4.3.3	Facteurs à prendre en considération lors de la mise en œuvre d'un projet d'archivage -----	172
	Conclusion-----	176
5.	Discussion -----	179
	Introduction -----	179
5.1	Le site Web en tant que nouvel objet archivistique -----	179
5.1.1	Le site Web en tant qu'objet documentaire : les particularités -----	180
5.1.2	Le site Web en tant qu'objet archivistique : niveau de granularité à considérer --	185
5.1.3	Le site Web comme lieu de production de nouveaux types d'archives-----	188
5.2	Archivage du Web : politiques et pratiques -----	191
5.2.1	Pratiques des archives nationales -----	191
5.2.1.1	Modèles de gestion -----	191
5.2.1.2	Le statut des sites Web dans les projets d'archivage du Web menés par des archives nationales : quelques réflexions-----	195
5.2.2	Pratiques organisationnelles-----	197
5.2.2.1	Modèles de gestion -----	197

5.2.2.2	Le statut des sites Web dans les sites d'études : les différentes perceptions	-200
5.3	Politique type pour l'archivage des sites Web organisationnels	-----206
5.3.1	Caractéristiques d'une politique d'archivage	-----206
5.3.2	Composantes d'une politique type pour l'archivage du Web organisationnel	----209
5.3.3	Mesures à mettre en place	-----217
5.3.3.1	Analyse des risques	-----218
5.3.3.2	Autres mesures à mettre en place	-----222
5.3.3.2.1	Conception et mise en ligne du site Web	-----223
5.3.3.2.2	Modification, mise à jour, retrait des contenus	-----228
5.3.3.2.3	Conservation à des fins administratives et historiques	-----228
5.3.3.3	Exemples de règles de conservation	-----230
5.3.4	Limites de la politique type	-----233
6.	Conclusion de la recherche	-----235
6.1	Résumé de la recherche	-----235
6.2	Apports théoriques de la recherche	-----238
6.3	Apports méthodologiques	-----239
6.4	Apports pratiques	-----240
6.5	Limites de la recherche	-----242
6.6	Recherches futures	-----243
	Bibliographie	-----245
	Annexe 1 : Principales initiatives d'archivage du Web (mise à jour : novembre 2012)	----- i
	Annexe 2 : Liste des politiques d'archivage du Web et documents complémentaires par pays et par type d'organisme créateur (mise à jour novembre 2011)	----- xvii
	Annexe 3 : Grille d'analyse des politiques d'archivage du Web organisationnel	----- xxii
	Annexe 4 : Grille d'observation des SIW	-----xxv
	Annexe 5 : Arborescence Scrapbook – règles de nommage des pages capturées	-----xxxii
	Annexe 6 : Grille d'observation des archives Web	-----xxxii
	Annexe 7 : Guide d'entrevue	----- xxxiv
	Annexe 8 : Formulaire de consentement	-----xxxviii
	Annexe 9 : Certificat d'éthique	----- xli
	Annexe 10 : Feuille de réponse	----- xlii

Annexe 11 : Protocole d'entrevue-----	xlviii
Annexe 12 : Lettre de sollicitation -----	l
Annexe 13 : Lettre de remerciements-----	li
Annexe 14 : Guide d'entrevue (BANQ)-----	lii
Annexe 15 : Guide de codage-----	lv

Liste des tableaux

Tableau I : Description de l'approche de gestion des risques associés au site Web. -----	59
Tableau II : Lien entre les questions de recherche, les modes et les instruments de collecte des données.-----	93
Tableau III : Répartition des politiques et documents complémentaires analysés selon l'organisme créateur.-----	96
Tableau IV : Répartition des politiques et documents complémentaires selon le pays. -----	97
Tableau V : Répartition des politiques et documents complémentaires selon les sujets couverts (n=55). -----	98
Tableau VI : Catégories des sites Web observés dans les sites d'étude (n=11). -----	100
Tableau VII : Types des sites Web observés selon la technologie de développement (n=11). -	102
Tableau VIII : Répartition des participants à la recherche selon les milieux organisationnels.	104
Tableau IX : Cadres de référence cités dans les politiques d'archivage du Web organisationnel. -----	117
Tableau X : Rôles et responsabilités en matière de gestion et d'archivage des sites Web. ----	122
Tableau XI : Unités d'archivage d'un site Web organisationnel par niveau d'archivage. -----	127
Tableau XII : Exemples de valeurs archivistiques d'un site Web public et de ses composantes. -----	143
Tableau XIII : Exemples de documents de gestion par niveau de profondeur et par fréquence d'apparition. -----	154
Tableau XIV : Rôles et responsabilités des répondants quant à la gestion et l'archivage du site Web public de leur organisation. -----	160

Liste des figures

Figure 1 : Cadre de référence de la recherche.-----	77
Figure 2 : Sources, modes et instruments de collecte des données.-----	83
Figure 3 : Chronologie de la collecte des données. -----	92
Figure 4 : Chronologie de l'analyse des données.-----	108
Figure 5 : Composantes et fonctionnalités d'un document Web. -----	137
Figure 6 : Modèle enrichi de McClure et Sprehe (1998) sur l'analyse des risques – facteurs à considérer.-----	219

Liste des sigles

BAC	Bibliothèque et Archives nationales du Canada
BAnQ	Bibliothèque et Archives nationales du Québec
BnF	Bibliothèque nationale de France
INA	Institut national d'audiovisuel
IIPC	International Internet Preservation Consortium
ISO	International Organization for Standardization
OAIS	Open Archival Information System
SIW	Système d'information Web
URL	Uniform Resource Locator

À mes parents

*À celui qui m'a mise sur le chemin de cette belle
aventure, M. Jacques Grimard*

À tous mes ami(e)s

Remerciements

Je voudrais exprimer ma reconnaissance envers tous ceux qui ont, de près ou de loin contribué, à la réalisation de ce travail.

Je tiens à adresser, en tout premier lieu, mes chaleureux remerciements à ma directrice de recherche, Mme Louise Gagnon. Elle n'a pas seulement accepté de diriger cette thèse, elle n'a pas cessé de m'encourager et de me soutenir durant ces cinq dernières années. J'en profite pour lui exprimer ma profonde gratitude pour son soutien indéfectible aussi bien sur le plan scientifique qu'humain.

Je remercie très sincèrement Mme Christine Dufour, ma codirectrice de recherche. Il est très difficile de trouver des qualificatifs assez forts pour souligner sa gentillesse, sa patience et surtout sa rigueur scientifique. Merci de m'aider à surpasser plusieurs incertitudes.

Je souhaite remercier M. Yvon Lemay qui a su m'orienter et m'apporter des pistes de réflexion intéressantes dans le cadre de cette thèse. Votre compétence, clairvoyance et sagesse m'ont beaucoup appris.

Un merci tout spécial à Mme Dominique Maurel et M. Ducharme qui ne m'ont pas soutenue uniquement sur le plan moral, mais aussi sur le plan financier. Ils ont accepté très généreusement et avec beaucoup de confiance de partager avec moi leurs projets de recherche.

Cette thèse, aboutissement de sept années d'études, je la dois à des parents exceptionnels qui ont accepté avec beaucoup de compréhension et d'espoirs l'éloignement de leur enfant unique au moment où ils en ont le plus besoin. Merci maman pour tes nombreux sacrifices. Merci papa pour ton soutien et ta confiance. Merci de vouloir toujours m'apprendre que je dois voler de mes propres ailes.

J'ai quitté ma famille, mes amis et mon pays. J'ai retrouvé au Québec les meilleurs amis au monde, sans qui rien n'aurait été possible. Leila, ton amitié est le cadeau le plus précieux que m'a offert la vie. Je te dois la nouvelle Aïda. Myriam, Sophie, Aouatef et Abdel merci pour votre affection et soutien inconditionnel. Un grand merci à tous mes amis (la liste serait très très longue) et à mes collègues au doctorat pour le réconfort et les encouragements.

J'assure ma reconnaissance aux personnes dans l'ombre à l'EBSI et à l'ISD dont la contribution à mon travail est non négligeable.

J'exprime ma gratitude pour tous les participants à cette recherche et les organismes qui l'ont subventionnée : l'EBSI, la FAS, le FQRSC et la mission universitaire de Tunisie à Montréal.

Je ne saurai terminer sans une pensée pour toi, qui n'a pas pu voir l'aboutissement de mon travail. M. Grimard vous me manquez.

1. Introduction

1.1 Problématique

La convergence entre l'informatique et les moyens de télécommunication marque fortement la société moderne. Cette fusion favorise, entre autres, l'amélioration des capacités de stockage des données, la facilité des échanges informationnels et l'évolution des pratiques documentaires (Bergeron, 1992; Duranti, 2001). Les moyens traditionnels d'édition et de communication sont de plus en plus délaissés au profit de nouvelles formes de communication « virtuelle » basées notamment sur des technologies Web très évolutives, mais encore instables (Duranti, 2004: 77; Taniar et Rahayu, 2004; Umar, 2005).

La richesse des fonctionnalités et les options de personnalisation offertes par les technologies Web séduisent aussi bien le grand public que le monde organisationnel. Le déploiement des systèmes d'information basés sur des technologies Web, appelés aussi systèmes d'information Web (SIW), dans la gestion des processus organisationnels est désormais perçu comme un avantage concurrentiel (Isakowitz et al., 1998; Stephens, 2001). Les organisations implantent massivement ces technologies et les exploitent pour s'assurer une présence dans le monde virtuel (Garon, 2003; Heinze et Hu, 2006; Huang et al., 2002; Liao et al., 2006; Thibodeau, 1999-2000). Ainsi, l'information nécessaire à la prise de décision et à la gestion est de plus en plus créée, recherchée, traitée, partagée, échangée et diffusée sur l'intranet ou l'extranet d'une organisation (Frost, 2001; Heinze et Hu, 2006). Plusieurs documents destinés aux partenaires et à la clientèle sont publiés sur le site Web officiel de l'organisation. Des services, produits et transactions sont, dans certains cas, offerts exclusivement sur le site Web de l'organisation (Liao et al., 2006; Umar, 2005). Plus encore, un « guichet unique », ou portail, fédère les services et produits offerts par des unités et des organismes collaborant ensemble (McDonald, 2005; Parent et Boulet, 1999a). En ce sens, plusieurs organismes à caractère public ou gouvernemental adoptent les technologies Web dans le souci d'assurer des services de qualité, de réduire la complexité des procédures administratives, de faciliter l'accès

à l'information et de faire preuve de transparence et d'imputabilité (Cunningham et Phillips, 2005).

L'expansion de l'usage des technologies Web, aussi bien dans la sphère publique qu'organisationnelle, accroît de manière considérable le volume et la complexité de la masse documentaire (Duranti, 2004: 75). En effet, de nombreux supports, médias et ressources d'information migrent vers le Web. D'autant plus, la publication des documents et l'édition de pages Web ne sont pas contrôlées par une instance officielle et ne requièrent pas nécessairement la maîtrise des techniques d'édition Web. De plus, il est possible de connecter plusieurs objets documentaires de nature et d'origines distinctes, d'assembler des fragments d'information stockés dans différents espaces documentaires et d'interagir avec le contenu publié. Les contours spatio-temporels d'un document sur le Web sont flous, sa forme se dissocie de son contenu et son accessibilité dépend d'une myriade d'applications et de progiciels (Boullier et Ghitalla, 2004; Leleu-Merveil, 2004; Liu, 2004).

Les transformations induites par les technologies Web tant au niveau des contenus qu'au niveau des pratiques documentaires rendent l'identification des nouvelles formes documentaires ardue, la maîtrise de certains contenus, un véritable défi, et augmentent, par la même occasion, le risque de pertes irréversibles de documents à valeur permanente (Barry, 2004; Day, 2003a; Duranti, 2004: 76-77; Harries, 1999; LeFugy, 2001; Shepherd et al., 2004).

Les technologies Web soulèvent aussi des questionnements sur l'adéquation des pratiques traditionnelles de gestion face aux nouveaux documents auxquels ces technologies donnent naissance (Duranti, 2001, 2004; Gagnon, 2000, 2002; Hedstrom, 2002; Liu, 2004). Plusieurs réflexions, recherches et expériences dans le domaine des sciences de l'information, empruntant des approches multidisciplinaires, tentent de développer une meilleure compréhension des caractéristiques de la production documentaire dans un contexte Web, et ce, afin d'améliorer le repérage, le traitement et la préservation du patrimoine documentaire en ligne (Duranti et al., 2002; McDonald, 2005; Parent et Boulet, 1999a; Pédaque, 2003; Shepherd et al., 2004).

Peu d'études empiriques portent sur l'analyse des caractéristiques et de la typologie des archives dans un environnement numérique et, plus particulièrement, dans un environnement

Web (Bourhis, 2000-2001; Duranti, 2004; Gagnon, 2002; Öberg et Borglund, 2006; Zeller, 2004). De nombreux écrits cherchent, toutefois, à redéfinir certains concepts archivistiques dans le cadre d'« un nouveau paradigme informationnel » (Duranti, 2001; Gilliland-Swetland, 2000; Grimard, 2005-2006; Yusof et Chell, 2002). D'autres réflexions soutiennent que les assises conceptuelles de la discipline archivistique restent inébranlables dans un contexte marqué par la prédominance de l'outil technologique (Hedstrom, 2002; Sénécal, 2005; Thibodeau, 1999-2000; Zwarich et Park, 2008). Des travaux plus récents s'intéressent aux nouvelles pratiques et techniques de préservation des archives numériques, champ de spécialisation émergeant dans le domaine des archives (Bourhis, 2000-2001; Bütikofer et al., 2006; Chabin, 2006 et al.; Hedstrom, 2002).

Malgré le foisonnement des efforts et des réflexions sur les archives numériques, aucune étude empirique n'a été menée, à ce jour, pour examiner les caractéristiques et les types d'archives véhiculées grâce à des systèmes d'information Web. En effet, le Web est souvent perçu comme un média complémentaire pour la diffusion de l'information et non comme outil de production documentaire résultant de l'usage des technologies Web dans la conduite des affaires d'une organisation (Harries, 1999). Les versions validées et officielles des documents Web sont censées être produites et conservées dans d'autres espaces institutionnels (Cunningham et Phillips, 2005; Forde, 2007; McDonald et al., 2005; Phillips, 2003; Stephens, 2001). La rareté des cas de litiges relatifs à l'accessibilité et au contenu des documents diffusés sur un site Web, le besoin non encore exprimé de se servir des sites Web organisationnels pour mener des recherches et la rareté de cadres de référence institutionnels quant à la gestion des sites Web en tant que documents d'archives peuvent également expliquer pourquoi les sites Web organisationnels ne sont pas souvent considérés comme ayant une valeur archivistique (Barry, 2004; Day, 2003a; Harries, 1999; McClure et Sprehe, 1998).

Des projets de collecte et de préservation des sites Web et des ressources en ligne ont été entrepris depuis les débuts du Web suite aux travaux d'Internet Archive (Brown, 2006). Les bibliothèques nationales ont également emboîté le pas (Bachimont et al., 2005; Gharsallah, 2002; McDonald, 2005). Les travaux de ces institutions de mémoire ont permis l'extension du dépôt légal aux ressources en ligne (publications en ligne, sites Web, blogues, etc.) et

l'expérimentation de nouvelles méthodes et techniques de capture et de préservation des documents (Deegan et Tanner, 2002; Masanès, 2006). Les enjeux sont de taille au point où, suite à quelques expérimentations, la coopération à l'échelle nationale et internationale entre les institutions de mémoire s'est avérée indispensable pour surmonter les nombreuses difficultés tant méthodologiques que techniques liées à l'archivage du Web (Chaimbault, 2008; Masanès, 2006; Verheul et al., 2006).

La littérature dans le domaine des archives reste silencieuse quant aux pratiques organisationnelles en matière de gestion des archives Web, qu'il s'agisse de sites Web et des documents qui y sont diffusés ou de la documentation relative à leur gestion.

« There are not many studies that consider Web records from a records management perspective, which would provide a systematic approach to the organization and management of the Web as records. » (Zwarich et Park, 2008: 3)

Plusieurs facteurs pourraient expliquer le manque d'information sur les pratiques d'archivage du Web dans un contexte organisationnel. Tout d'abord, la gestion et la maintenance des sites Web et de leurs contenus sont souvent confiées à une équipe de spécialistes en technologie de l'information, voire les créateurs des contenus Web. Ensuite, les archivistes et les gestionnaires des documents ne sont pas, dans la plupart des cas, impliqués dans ce genre de tâches. Enfin, les archivistes et gestionnaires de documents peuvent également ne pas posséder les compétences, ni les connaissances nécessaires pour participer à la gestion des sites Web. Aussi, l'archivage d'un site Web peut ne pas constituer une priorité pour l'organisation, le manque de ressources matérielles et financières pouvant en expliquer les raisons (Barry, 2004; Cunningham et Phillips, 2005; Phillips, 2003; Willer, 2008).

Plusieurs rapports d'enquête et études sur les initiatives d'archivage du Web sont publiés ces dernières années (Gomes et al., 2011; Gregory, 2009; Internet Memory Foundation, 2010; NDSA, 2012). Les bibliothèques nationales et archives nationales s'y intéressent plus assidument (Gomes et al., 2011: 413; Gregory, 2009: 10-11; Internet Memory Foundation, 2010: 5). On dénote une prépondérance de l'archivage du Web par des institutions universitaires dans une perspective documentaire (NDSA, 2012: 2, 17). L'implication des institutions archivistiques dans les projets d'archivage du Web se démarque par sa timidité

malgré la publication par des archives nationales avant-gardistes (ex. : Australie, Canada, États-Unis, Nouvelle-Zélande) de politiques et lignes directrices pour l'archivage du Web organisationnel (Beagrie et al., 2008; Khouaja et Couture, 2003-2004). Quelques institutions d'archives nationales procèdent à la collecte et ensuite à la diffusion des archives du Web gouvernemental à l'instar des Archives nationales du Royaume-Uni et de Bibliothèque et Archives Canada (BAC). Toutefois, les institutions ayant un programme actif d'archivage du Web n'investissent pas toujours dans l'archivage de leur propre site Web et encore moins comme étant un document d'archives à gérer (NDSA, 2012: 3). Le statut des documents capturés n'est pas clairement établi puisque, dans bien des cas, les organismes agissent sans cadre de référence, c'est-à-dire sans avoir défini au préalable une politique d'archivage (NDSA, 2012: 18).

Quelques organisations développent des politiques d'archivage de leurs sites Web, notamment les institutions universitaires (Beagrie et al., 2008). Ces politiques proposent des approches largement fondées sur un modèle développé depuis les débuts du Web, le modèle de gestion des risques des sites Web de McClure et Sprehe (1998). Les méthodes et les techniques d'archivage restent similaires à celles utilisées par les bibliothèques nationales (Brown, 2006). De surcroît, ces politiques ne proposent pas d'outils, ni de méthodes facilitant l'identification des archives Web dans un contexte où les frontières, le type et la forme d'un document ne sont pas faciles à définir. Or, il est évident que les usages et les techniques du Web ont fortement évolué. De même, les archives requièrent un mode de gestion qui diffère de celui des publications. L'adéquation des pratiques préconisées dans les politiques d'archivage du Web organisationnel aux caractéristiques des archives Web, d'une part, aux réalités documentaires de l'organisation, d'autre part, reste à évaluer.

Notre recherche vise à développer une meilleure compréhension de la nature des archives Web, à décrire les pratiques actuelles d'archivage du Web organisationnel et à proposer une politique type qui rappelle les principes de gestion et les mesures à mettre en place pour la gestion des archives Web. Ces archives sont un témoin unique des activités en ligne des organisations et une composante importante de la mémoire organisationnelle et sociétale. Les pratiques de gestion doivent garantir la préservation de traces fiables et authentiques des

activités organisationnelles, s'adapter aux évolutions technologiques ainsi qu'aux caractéristiques des objets documentaires à préserver (Duranti, 2004).

1.2 But et objectifs de la recherche

Cette recherche vise à explorer et à décrire des pratiques de gestion documentaire émergentes et à les comparer aux pratiques préconisées dans des cadres de référence archivistiques. Plus précisément, le projet de recherche a pour but de développer une meilleure compréhension de la nature des archives Web et de documenter la démarche de leur archivage dans un contexte organisationnel du point de vue des acteurs impliqués dans ce processus. Les résultats empiriques sont à comparer aux recommandations émanant des institutions archivistiques avant-gardistes dans le domaine des archives ou ayant un programme d'archivage du Web bien établi, et ce, afin de parvenir à établir un modèle de politique qui prend appui sur les cadres de référence existants tout en considérant les besoins à satisfaire et les ressources disponibles dans une organisation.

Les objectifs spécifiques de notre recherche sont de :

- Analyser les politiques d'archivage du Web organisationnel pour en dégager une politique type;
- Dresser un portrait des types et des caractéristiques des archives Web;
- Décrire les pratiques d'archivage du Web organisationnel.

1.3 Questions de recherche

Des objectifs de recherche découlent les questions de recherche suivantes :

1. Que recommandent en général les politiques d'archivage du Web organisationnel?
2. Quelles sont les principales caractéristiques des archives Web?
3. Quelles pratiques d'archivage du Web organisationnel sont mises en place dans des organisations au Québec?
 - a. Quels sont les principaux intervenants impliqués dans le processus d'archivage du Web organisationnel?

- b. Quelles sont les approches et les techniques d'archivage du Web employées dans un contexte organisationnel?
- c. Quels facteurs influencent la mise en œuvre d'une politique d'archivage du Web organisationnel?

Afin de répondre à ces questions, nous avons mené une étude de nature descriptive et exploratoire basée sur une approche qualitative dans 11 organismes au Québec. Nous avons examiné les caractéristiques techniques et documentaires des sites Web publics de ces organismes et interviewé des acteurs impliqués dans leur processus d'archivage. Nous avons comparé les résultats obtenus aux recommandations suggérées dans 33 politiques d'archivage du Web. Cette démarche de recherche nous apprend davantage sur les caractéristiques des archives dans un environnement numérique réseauté, sur les méthodes et techniques d'archivage mises en place dans des milieux organisationnels, et enfin, sur les facteurs qui dictent les choix actuels. Une politique type a pu être établie sur la base des résultats obtenus.

Nous avons défini, dans ce premier chapitre, la problématique de recherche, ainsi que les questions et objectifs de recherche qui en découlent. La revue de la littérature, la méthodologie utilisée, les résultats obtenus, la discussion de ces derniers et la conclusion de la recherche sont développés dans les chapitres suivants.

2. Revue de littérature

Introduction

Ce chapitre présente une recension des écrits et les résultats de travaux pertinents à notre recherche. À partir des écrits en systèmes d'information, sciences de l'information et archivistique, nous présentons dans la première partie les définitions des concepts à l'étude. Nous identifions dans la deuxième partie les principales caractéristiques de la production documentaire sur le Web et les critères d'identification des archives dans cet environnement. Nous décrivons dans la dernière partie les modes, les approches, les techniques ainsi que les acteurs impliqués dans l'archivage du Web selon deux perspectives : la perspective documentaire et la perspective archivistique.

2.1 Définitions

Dans cette section, nous présentons, du général au particulier, les définitions adoptées pour deux principales catégories de concepts qui seront abordés dans le cadre de cette recherche, soit les concepts technologiques et les concepts archivistiques.

2.1.1 Concepts technologiques

Internet

« Réseau informatique mondial constitué d'un ensemble de réseaux nationaux, régionaux et privés, qui sont reliés par le protocole de communication TCP-IP » (OQLF, 2007a).

Intranet

« Réseau informatique privé, à l'intérieur d'une organisation, qui utilise les protocoles de communication et les technologies du réseau Internet » (OQLF, 2007b).

Extranet

« Réseau informatique sécurisé, généralement constitué d'une partie de l'intranet d'une entreprise ou d'une organisation communiquant à travers le réseau Internet, qui est accessible à une clientèle externe ciblée devant utiliser un mot de passe » (OQLF, 2005).

Web

Ensemble de serveurs donnant accès à une collection de sites Web grâce au protocole HTTP et au mode d'adressage universel URL (Collin, 2002; Foo et Lim, 1997: 168).

Dans sa portion publique, les documents sur le Web sont repérables et indexés par les robots et moteurs de recherche. Lorsque les documents ne peuvent pas être indexés (ex. : site Web non référencé ou à accès restreint), localisés (ex. : documents dynamiques) ou non indexables (animations flash, cadres, etc.), la portion du Web qui les héberge est dite profonde ou invisible.

Le Web dynamique est constitué de pages Web interactives, générées automatiquement à partir d'une base de données ou d'un système de gestion de contenus. La structure d'un site Web dynamique permet également de recueillir des rétroactions directes sur le contenu présenté.

Système d'information

Ensemble finalisé et organisé de ressources humaines, matérielles et informationnelles permettant d'acquérir, de traiter, de stocker, et de diffuser des informations dont a besoin une organisation pour une gestion et un contrôle efficaces de ses activités (Laudon et Laudon, 2001: 8; Reix, 2004: 3; Rivard et Talbot, 2001: 20).

Système d'information Web

Système d'information basé sur les technologies Web (Butler, 2003: 210; Detlor, 2000: 3; Dufour, 2003: 8; Holk, 2002: 2; Isakowitz et al., 1998: 79; Wang, 2001: 385).

Site Web

Terme générique pour désigner un site Web public, un intranet ou un extranet. Un site Web est constitué d'un ensemble de pages Web interreliées sous la responsabilité d'une même entité, offrant des informations ou des services accessibles en ligne grâce à un navigateur Web (Collin, 2002; Dufour, 2003: 8-9; McClure et Sprehe, 1998: 3; Stephens, 2001: 64; Zwarich et Park, 2008: 1).

Page Web

Unité élémentaire d'un site Web créée grâce à un assemblage d'objets informationnels à l'aide d'un langage de balisage, consultée sur un même écran d'affichage sans besoin de navigation à l'extérieur de cette dernière (Collin, 2002; Dufour, 2003: 9).

Document

« Un document est constitué d'information portée par un support. L'information y est délimitée et structurée, de façon tangible ou logique selon le support qui la porte, et elle est intelligible sous forme de mots, de sons ou d'images » (*Loi concernant le cadre juridique des technologies de l'information*, chap. C-1.1, art. 3).

Document Web

Contenu hypertextuel ou hypermédia diffusé sur le Web offrant une information complète sur un sujet donné (Dmitriev et al., 2005; Thelwall, 2004: 28).

2.1.2 Concepts archivistiques*Archivage*

Ensemble d'actions et de procédures permettant, grâce à un contrôle efficace et systématique de la création, de la réception, de la conservation, de l'utilisation et du sort final des documents, la gestion du cycle de vie des documents présentant des utilités administratives ou légales et constituant une source d'information et de témoignage sur l'organisme producteur (AFNOR, 2004: 84; Chabin, 2007: 35-51; Chabin et al., 2006: 23; McDonald et al., 2005: 10; Couture, 1999: 112-114; Rousseau et Couture, 1994: 102).

Archivage du Web

Ensemble d'actions et de procédures permettant la capture, l'acquisition, le traitement, la conservation et la préservation de documents Web (Brodie, 1997; Brown, 2006).

Lorsqu'il s'agit de publications en ligne, l'archivage réfère plutôt à la collecte, à l'acquisition des documents diffusés sur le Web. Dans ce cas, il est synonyme de dépôt légal du Web (Beagrie, 2003; Chaimbault, 2008; L'Hostis et al., 2006; Masanès, 2006).

Lorsque ce terme est associé aux documents à valeur archivistique, il renvoie à la gestion et à la préservation des archives créées ou diffusées grâce à un système d'information Web (Harries, 1999; Khouaja et Couture, 2003-2004; Stephens, 2001).

Archives

« L'ensemble des documents, quelle que soit leur date ou leur nature, produits ou reçus par une personne ou un organisme pour ses besoins ou l'exercice de ses activités et conservés pour leur valeur d'information générale » (*Loi sur les archives*, chap. A-21.1, art. 2).

Archives Web

Documents issus ou relatifs à un système d'information Web, produits dans le cadre des activités d'un organisme donné y compris le site Web, les documents diffusés sur ce site Web ainsi que la documentation relative à leur gestion (Barry, 2004: 27; Cleroult, 2006: 13; McClure et Sprehe, 1998: 3; Stephens, 2001: 67).

Valeurs archivistiques

La notion de valeur archivistique renvoie à la finalité même d'un document d'archives. Dans son double emploi, elle désigne la valeur primaire et la valeur secondaire d'un document d'archives.

Valeur primaire : « finalité du document d'archives fondée sur ses utilités premières, sur les raisons pour lesquelles il existe. La valeur primaire réfère à la preuve [administrative, légale ou financière] » (Couture, 1999: 113).

Valeur secondaire : « finalité du document d'archives fondée sur ses utilités secondes ou scientifiques. La valeur secondaire est liée à l'information et au témoignage » (Couture, 1999: 114).

2.2 Le Web et la production documentaire

Le Web constitue un réservoir d'artefacts documentaires aussi bien diversifiés qu'instables et un milieu de communication où les notions d'espace et de temps n'ont plus la même signification (Leleu-Merviel, 2004). Phénomène documentaire, social et technologique, le Web devrait être préservé afin de témoigner de l'évolution des pratiques de communication qu'il a révolutionnées et des formes documentaires qu'il a engendrées au risque de perdre des fragments importants de notre mémoire collective (UNESCO, 2003). Malgré la prépondérance de cet outil technologique sur d'autres moyens de communication, la masse documentaire que véhicule le Web se caractérise par sa fragilité. La mémoire du Web devrait être conservée pour les besoins actuels et futurs (Deegan et Tanner, 2002: 26; Gladney, 2007: 3; UNESCO, 2003).

Dans les pages qui suivent, nous présentons les principales caractéristiques de la production documentaire sur le Web ainsi que les caractéristiques essentielles pour identifier, dans cet environnement, les documents à valeur archivistique.

2.2.1 Caractéristiques de la production documentaire sur le Web

2.2.1.1 Des objets documentaires faciles à produire, difficiles à définir

L'accessibilité des technologies Web et l'absence d'une instance de contrôle qui régule et évalue les contenus mis en ligne sont parmi les facteurs clés qui ont permis la production et la diffusion d'une masse prolifique et diversifiée de documents sur le Web (Liu, 2004). En effet, la technologie Web permet d'assembler des objets documentaires de nature distincte, stockés dans des fichiers et des serveurs disparates sous un même format de présentation (Bachimont, 2000: 11; Toms, 2001: 20). Ces éléments sont accessibles à partir d'une même adresse URL ou produits à la demande dans le cas d'un contenu affiché sur un site Web dynamique par exemple (Boullier et Ghitalla, 2004; Gharsallah, 2002: 63-68; Masanès, 2000).

Le Web ravive les discussions autour du concept de document, de sa permanence et de ses valeurs. La littérature en sciences de l'information, en systèmes d'information et en technologies de l'information foisonne de réflexions sur les composantes d'un document Web, les formats de documents sur le Web et les caractéristiques distinctives des documents Web (Leleu-Merviel, 2004; Pédaque, 2006). Toutefois, il reste difficile d'établir une définition générique du document Web (Thelwall, 2004: 27).

Les conceptions de la notion de document sont toujours conditionnées par les techniques et les moyens disponibles de l'époque (Lainé-Cruzel, 2005; Levy, 1994; Schamber, 1996). Pour identifier de nouvelles formes documentaires, on se réfère souvent aux attributs de base d'un document (contenu, structure et forme). Cette approche permet de définir des formes documentaires homogènes au niveau du contenu ou similaires au niveau de la forme (Crowston et Williams, 1999; Lafouge et al., 2002). Dans un contexte Web, des formes documentaires traditionnelles (livre, journal, film, musique, objet tridimensionnel, catalogue de bibliothèque, etc.) coexistent avec de nouvelles formes documentaires propres à cet environnement (page d'accueil, foire aux questions, blogue, Wikitionnaire, etc.). Les caractéristiques de ces dernières

sont largement dépendantes des caractéristiques du médium qui les incorpore (Shepherd et al., 2004).

2.2.1.2 Des objets documentaires dynamiques, aux contours dilués

La logique du Web est fondée sur trois principes : connectivité, interactivité et fluidité des contours spatio-temporels des contenus véhiculés. Ces principes bouleversent les caractéristiques essentielles des documents dans un environnement traditionnel, soit la fixité, la stabilité et la matérialité.

Un document Web résulte souvent d'une association entre des fragments documentaires hétérogènes : texte, vidéo, image, son, graphique, etc. Il n'existe pas de cadre normatif ou de règles qui dictent les limites des composantes d'un document Web. De même, la persistance des relations entre ces composantes et l'accessibilité au contenu dépendent de la stabilité des liens internes et externes, des choix éditoriaux du concepteur et des avancées technologiques (Leleu-Merveil, 2004).

Un document Web n'est jamais fini. Il est constamment en construction. Sa connectivité avec d'autres documents et sa nature dynamique assurent une mise à jour continue d'une portion ou de la totalité de son contenu. La fixité de ses contenus, la stabilité de sa forme et le fonctionnement de ses liens ne sont pas toujours garantis, ce qui confère au document Web son caractère instable (Masanès, 2005: 73).

La connectivité et l'interactivité d'un document Web rendent ardue la définition de ses contours spatio-temporels (Koehler, 1999) et, par conséquent, son identification comme un objet fixe et délimité (Brown, 2006: 25). En s'appropriant les caractéristiques du médium qui l'incorpore, le document sur le Web devient de plus en plus polymorphe, complexe et évolutif, aux contours indéfinis et poreux. La prolifération de l'usage des liens a dilué les frontières d'un document Web. Thelwall recense six types de liens qui peuvent figurer sur un document Web :

1. Les liens internes ou externes à partir d'une même page Web (*Hyperlink*);
2. Les liens entre les mêmes pages d'un site Web (*Inlink*);
3. Les liens vers les pages d'un autre site Web (*Outlink*);
4. Les liens vers un autre niveau d'une même page Web (*Selflink*);
5. Les liens entre deux sites Web (*Interlink*);

6. Les liens présents dans deux pages Web différentes pointant vers une même troisième page (*Co-linked*) (Thelwall, 2004: 5-6).

Cette variété de liens ne peut que favoriser l'ubiquité des documents Web et leur délocalisation (duplication sur plusieurs serveurs, différents points d'accès). Les frontières spatio-temporelles d'un document Web sont difficiles à délimiter (Koehler, 1999, 2004; Leleu-Merveil, 2004; Thelwall, 2004).

2.2.1.3 Des objets documentaires uniques et éphémères

Le volume d'information sur le Web ne cesse de croître. Netcraft chiffre à plus de 644 millions le nombre de sites Web actifs (463 millions le mois d'août 2011)¹. Un milliard d'individus sont actifs sur Facebook selon les statistiques du mois d'octobre 2012². D'après le *Twitter Facts and Figures* (2012), les internautes se sont échangés plus de 750 gazouillis par seconde, soit 340 millions gazouillis par jour (statistique du mois de mars 2012)³. Entre décembre 2004 et décembre 2006, le nombre de blogues dans le monde est passé de 5,4 à 63,1 millions, pour atteindre 150 millions en 2010, selon Technorati⁴.

Le Web ne dispose pas encore des fonctionnalités ou des propriétés essentielles au maintien de la permanence des documents qu'il véhicule (Bachimont et al., 2005: 5). Certains documents sur le Web encourent le risque de disparaître à tout jamais sans laisser de traces (Zeldman, 2002). En effet, la durée de vie moyenne d'une page Web est limitée entre 75 et 100 jours et 50% des pages mises en ligne disparaissent après deux ans de vie sur la toile (Brown, 2006: 25; Gharsallah, 2002; Koehler, 1999).

Outre les contraintes techniques, les documents Web produits dans un contexte organisationnel se perdent faute de ressources financières, de manque d'expertise professionnelle, de cadre légal ou de politiques documentaires adaptés aux enjeux de la

¹ Netcraft (2012). *March 2012 web server survey*. Adresse URL: <http://news.netcraft.com/archives/2012/03/05/march-2012-web-server-survey.html> (Page consultée le: 24 novembre 2012).

² Facebook (2012). *Newsroom: key facts*. Adresse URL: <http://newsroom.fb.com/Key-Facts> (Page consultée le: 24 novembre 2012).

³ Website monitoring (2012). *Twitter 2012 – Facts and Figures (infographic)*. Adresse URL: <http://www.website-monitoring.com/blog/category/infographics/> (Page consultée le: 24 novembre 2012).

⁴ Technorati (2011). *Technorati's State of the Blogosphere*. Adresse URL: <http://technorati.com/social-media/article/state-of-the-blogosphere-2011-introduction/> (Page consultée le: 24 novembre 2012).

préservation du patrimoine documentaire en ligne (Cunningham et Phillips, 2005; Harries, 1999; McDonald et al., 2005).

Fondés sur la théorie des genres, les travaux de Yates et Orlikowski sur l'évolution de la lettre d'affaires avec la messagerie électronique (Yates et Orlikowski, 1992) démontrent aussi que les genres documentaires se transforment graduellement en réponse à des changements sociaux et organisationnels. Le genre représente une forme d'action communicative, socialement reconnue, et utilisée par les mêmes membres d'une communauté dans un but social particulier (Orlikowski et Yates, 1994). Chaque genre est caractérisé par ses attributs de contenu, de forme et de but :

« [...] a genre may be characterized by having similar *content* and *form* where content refers to themes and topics and form refers to “observable physical and linguistic features”. Genre may also be characterized by *purpose* and *form*, where purpose refers to the socially recognized communicative purpose, and form may include “[the] communication medium, as well as structural and linguistic features.” » (Shepherd et Watters, 1999: 1)

Les genres évoluent également au rythme de l'évolution des pratiques organisationnelles, des relations sociales et des supports de communication dans l'organisation (Karjalainen et al., 2000: 7; Orlikowski et Yates, 1994; Oystein et Päivärinta, 2005; Tyrväinen et Päivärinta, 1999; Yates et Summer, 1997). Un nouveau genre émerge quand il atteint une forme et une structure stables et lorsqu'il est ancré dans les pratiques de communication d'une communauté (Alberts, 2009; Dillion et Gushrowski, 2000).

Maintes études, bien ancrées dans la théorie du genre, affirment que l'appropriation des fonctionnalités et propriétés d'un nouveau médium ou média de communication (Web, collecticiel, courrier électronique, etc.) occasionne la transformation de genres existants voire l'émergence de nouveaux genres de documents (Alberts, 2009; Crowston et Williams, 1999; Watters et Shepherd, 1997; Yates et Summer, 1997).

Chabin confirme, dans une perspective archivistique, cette hypothèse en donnant l'exemple de trois groupes de « nouveaux documents à archiver » issus du monde numérique, à savoir :

1. Les documents sous forme numérique dont la filiation avec des documents traditionnels est facile à établir, tant par leur finalité que par leur forme;

2. Les documents numériques qui se présentent dans une forme renouvelée;
3. Les nouveaux documents émergents qui semblent ne se rattacher à aucun type d'archives connu et dont le destin est à inventer, les sites Web font partie de ce groupe de documents (Chabin, 2000: 131).

Le Web en tant que technologie d'information confère de nouvelles propriétés aux documents qu'il véhicule (Alberts, 2009; Crowston et Williams, 1999). L'intégration des technologies Web dans les pratiques de communication provoque même l'émergence de formes documentaires originales et uniques n'ayant pas d'équivalents dans d'autres environnements documentaires (Oystein et Päivärinta, 2005; Shepherd et Watters, 1998, 1999). Shepherd et ses collaborateurs (2004), dont les travaux s'appuient sur la théorie des genres, ont conclu que des genres uniques émergent sur la toile mondiale. Ils ont identifié une nouvelle classe de genres sur le Web, les cybergenres, tels que les foires aux questions (FAQ) et les pages d'accueil (Dillion et Gushrowski, 2000; Shepherd et Watters, 1998, 1999).

Selon une étude menée par Shepherd et Watters (1998, 1999), les genres, lorsqu'ils existent sur le Web, se distinguent par leurs attributs de forme, de contenu et de fonctionnalités, les fonctionnalités étant des caractéristiques héritées du Web : « Cybergenres are characterized by the triple, <content, form, functionality>, where functionality refers to capabilities available in the new medium. » (Shepherd et Watters, 1999: 1)

Il est toutefois important de signaler que les genres n'évoluent pas tous au même rythme. Certaines mutations peuvent réduire un genre à un état latent, le rendre flou, difficile à reconnaître, aux multiples usages et objectifs, et provoquent même son extinction s'il est socialement rejeté ou substitué par un autre genre mieux adapté aux besoins informationnels d'une communauté (Beghtol, 2001: 19; Bergquist et Ljungberg, 1999; Yates et Orlikowski, 1992; Yoshioka et al., 2001). Cette évolution incontrôlée des genres ne facilite pas la reconnaissance des nouveaux documents issus de l'usage des technologies Web.

2.2.1.4 Des objets documentaires composites difficiles à cerner dans le temps et dans l'espace

Le Web abonde en documents composites ce qui ne facilite pas le choix d'une unité d'archivage (Bachimont et al., 2005: 7-8; Bonnenfant, 2008: 39; Chaimbault, 2008; INA, 2003;

Lainé-Cruzel, 2005; Lemaître, 2001). La délimitation des contours des objets à archiver dans le temps et dans l'espace est essentielle lors de la définition du périmètre d'archivage. Dans une perspective documentaire, la sélection des ressources à archiver peut être basée, entre autres, sur le nom du domaine Web, le sujet couvert, la langue de publication et le format (Bachimont et al., 2005: 7). Les sites Web peuvent être capturés avant ou après une importante mise à jour ou encore avant leur retrait s'ils sont produits dans le cadre d'un événement ponctuel.

Dans une perspective archivistique, le site Web, les documents diffusés sur le Web et la documentation Web peuvent être considérés des unités d'archivage (Cleroult, 2006). Un site Web organisationnel pourrait être archivé avant une refonte majeure ou suite à une mise à jour importante. L'analyse des politiques d'archivage du Web organisationnel de même que l'observation des pratiques sur le terrain nous aidera à préciser les critères de choix de « la matière archivable » (Cleroult, 2006).

Le statut des sites Web archivés devrait aussi être précisé. Dans une perspective documentaire, les instantanés successifs d'un site Web sont considérés des versions d'un même produit documentaire. Dans une perspective archivistique, ces instantanés représentent-ils les différents états d'un « document d'archives évolutif » ou représentent-ils eux-mêmes des documents d'archives différents les uns des autres? Nous sommes face à une question sensible en archivistique : quand est-ce qu'un site Web atteint un statut stable fixe et fini dans le temps et l'espace? Quel statut auraient les instantanés (statut de copie, de version ou de document ayant le caractère fini et complet au moment de sa capture)? Comment peut-on évaluer des « documents d'archives inachevés »? Les instantanés sont-ils la seule trace de l'évolution du site Web et de ses usages dans le temps? Le Web bouleverse réellement les caractéristiques discriminantes qui définissent un document d'archives. Nous espérons que les données empiriques et les politiques d'archivage seront en mesure de nous éclairer sur ce point.

En résumé, le Web regorge aussi bien de formes documentaires issues des environnements traditionnels, dont les propriétés sont maintenues ou légèrement transformées, que de nouvelles formes documentaires originales, produits uniques de cet environnement virtuel (Andrieu, 2005: 38; Bachimont et al., 2005).

Le Web anime de nouvelles réflexions sur la notion de document. Le document sur le Web échappe à tout contrôle (conception, diffusion, durée de vie, frontières, etc.). L'originalité, la connectivité et l'interactivité des documents sur le Web remettent en question des modèles traditionnels pour caractériser et reconnaître certaines formes documentaires. Compte tenu de ce qui précède, on aura compris que les archives du Web ne se laissent pas facilement circonscrire.

2.2.2 Critères d'identification des archives dans un environnement Web

Les archives sont des documents, c'est-à-dire une information consignée sur un support, mais ce n'est pas n'importe quel document qui peut accéder au statut d'archives (Couture, 1996). Les documents d'archives possèdent un statut qui leur est propre.

Nous partons de l'hypothèse que l'identification des caractéristiques propres aux archives contribue non seulement à les distinguer des autres documents sur le Web, mais aussi à définir les exigences relatives à leur gestion dans un tel environnement (Duranti, 2004; Gagnon, 2002; Öberg et Borglund, 2006; Zeller, 2004).

Les caractéristiques distinctives des archives sont exposées sous ce point.

2.2.2.1 Caractère organique et consigné

L'organicité est l'une des caractéristiques uniques aux archives. En effet, les archives sont produites naturellement dans l'exercice d'une activité :

« [...] tant qu'il n'est pas appréhendé dans sa dimension de trace d'une activité, de preuve ou de témoignage d'un acte ou d'un évènement que l'on peut localiser et dater, un document ne peut pas être qualifié d'archives. » (Pipon et Pontier, 2004: 51)

C'est ce lien avec une activité, une action, un contexte particulier de création qui permet de déterminer la provenance des archives (McDonald et al., 2005: 10). Les archives se caractérisent également par leur appartenance à un fonds documentaire. Les archives de même provenance constituent un fonds d'archives qui diffère de la documentation (ensemble de documents triés et classés) et de la collection d'une bibliothèque (réunion artificielle d'un ensemble de documents dans un objectif précis) (Couture, 1996; Gagnon, 2000; Henry, 2000; Hofman, 1996; Turnbaugh, 2002). Elles sont toujours articulées de manière organique au sein d'un dossier ou d'un fonds d'archives qui leur procure un caractère indivisible (Öberg et

Borglund, 2006: 64). L'appartenance des documents à un fonds d'archives permet une interprétation juste et complète de leur contenu (Chabin, 2000: 15; Pipon et Pontier, 2004: 51). Les archives doivent être considérées par rapport à l'ensemble auquel elles appartiennent (le dossier, le fonds d'archives) pour servir pleinement leur fonction d'information et de témoignage, d'où l'importance de conserver toute l'information contextuelle qui détermine le lien entre les archives et l'activité qui les génère (Hofman, 1996: 118-120).

L'information organique doit être enregistrée à des fins de conservation dans un espace institutionnel (Rousseau et Couture, 1994: 45). Cet enregistrement assure le transfert de la responsabilité du document du contrôle individuel au profit du contrôle institutionnel (Parent et Boulet, 1999a: 7; Rousseau et Couture, 1994: 44-45). Ces enregistrements institutionnels (*records*) deviennent les traces des activités de l'organisation (Nahuet, 2009; Upward, 2005). Les documents numériques peuvent être enregistrés de manière automatique lorsque les fonctionnalités des logiciels de gestion documentaire sont intégrées aux applications de gestion des processus organisationnels. Des métadonnées sont alors attribuées au document enregistré afin de gérer son cycle de vie (Parent et Boulet, 1999a, 1999b).

2.2.2.2 Contexte de création et d'utilisation

Les archives sont porteuses d'information bien particulière, produite dans un contexte spécifique : « la production d'information organique et consignée donne naissance aux archives » (Rousseau et Couture, 1994: 45). Cette information est générée ou reçue dans le cadre des activités courantes d'une personne physique ou morale qui désire laisser les traces de ses activités, conclure une transaction et en tirer un profit économique ou moral, défendre ses droits et témoigner pour soi-même ou pour autrui (Chabin, 2000: 148-149; Duranti et al., 2002: 17; Turbaugh, 2002: 32). Néanmoins, seule la volonté du producteur du document de considérer un ensemble de documents comme la trace de son activité donne naissance aux archives. Cela n'exclut pas l'éventualité d'occasionner la constitution de documents d'archives (le cas des archives orales ou le cas de la reconstitution d'un fonds d'archives abandonné), ce que Chabin qualifie « d'archives provoquées » (Chabin, 2000: 36).

Le contexte de création et d'usage est cet ensemble d'informations qui renseignent sur le lien entre, d'une part, le document et son environnement administratif et, d'autre part, le

document et les autres documents produits dans le même contexte (Öberg et Borglund, 2006: 57, 63). Les métadonnées permettent de décrire le contexte de création, le contenu et la structure d'un document d'archives ainsi que son mode de gestion (Chabin, 2004; ISO, 2001b: 21; 15489-1, 2001; Nahuet, 2009; Parent et Boulet, 1999a).

Ainsi, la nature de l'information contenue dans les archives leur confère un statut particulier qui diffère de l'œuvre littéraire, de l'écrit scientifique ou du reportage factuel (Rousseau et Couture, 1994: 12-13).

2.2.2.3 Valeurs archivistiques

Les archives se distinguent aussi par les valeurs qu'elles présentent. Les archives sont une preuve, une trace d'une transaction, d'une action, d'une décision (*evidence of transaction*). Il s'agit de l'utilité première des archives ou valeur primaire, c'est-à-dire, de la raison de création et de l'utilisation d'un document. Les archives sont un support à la décision, un instrument de travail et de contrôle ainsi qu'un moyen de justification des droits. Les archives sont aussi une source d'information, un témoignage sur des faits, des actions et des décisions. Il s'agit, dans ce cas, de la valeur secondaire ou l'utilité scientifique et informationnelle des archives (Rousseau et Couture, 1994; Sénécal, 2005).

Dans un article paru dans la revue *Archives*, Ducharme identifie neuf critères intellectuels qui s'appliquent à tout document d'archives (Ducharme, 2000-2001: 26), dont il faut tenir compte lorsqu'on évalue les archives numériques. Ces critères sont inspirés de cinq principes universels pour déterminer la valeur des archives et sur lesquels repose l'évaluation archivistique (Couture, 1999: 110-112). Nous les présentons dans un ordre différent de celui proposé par l'auteur afin d'énoncer d'abord l'intérêt que les archives présentent pour l'organisme créateur et, ensuite, pour la société :

1. Faculté de témoigner des activités de l'organisation ou d'un individu;
2. Faculté d'informer l'organisation ou l'individu sur des dossiers d'importance;
3. Pertinence de l'information;
4. Nature des documents : unicité, rareté, authenticité, etc.;
5. Complément d'autres types de documents du fonds d'archives;
6. Provenance (importance du créateur, etc.);

7. Utilisation potentielle des archives informatiques pour la recherche;
8. Secteur d'activités identifié comme étant représentatif de la société;
9. Intérêt national de l'information consignée sur support informatique (Ducharme, 2000-2001: 25).

L'auteur souligne le caractère arbitraire de ces critères et la difficulté à les mettre en œuvre. Les critères identifiés par Ducharme se justifient dans un contexte Web. Il sera intéressant de les comparer aux motifs d'archivage que nous risquons d'identifier suite à l'analyse des politiques d'archivage du Web organisationnel et de l'observation des pratiques dans les sites d'étude. Les sites Web témoignent des activités en ligne des organisations qui composent la société (Barry, 2004; Day, 2003b; Stephens, 2001). Les archives du Web, malgré le peu de recherches et d'études qui exploitent leur potentiel de témoignage, constituent des matériaux précieux pour écrire notre histoire contemporaine (Reilly, et al., 2003; Rygiel et al., 2005).

L'évaluation, en tant qu'intervention archivistique, permet de préciser le statut d'un site Web organisationnel et de ses composantes et de préciser les valeurs qu'il présente pour l'organisation et la société tout au long de son cycle de vie (Couture, 1999: 112). Certes les finalités de création d'un site Web est un élément important à considérer lorsqu'il s'agit de définir son sort final, l'appréhension de ses utilisations ou exploitations possibles est aussi déterminante.

2.2.2.4 Conditions de validité des archives

L'archivité, ou *recordness*, d'un document ne peut être acquise sans répondre à certaines conditions de validité ou critères de qualité (Hofman, 1996). Selon la norme ISO 15489-1, un document d'archives, pour être valide, doit être authentique, fiable, intègre et exploitable (ISO, 2001a: 7-8).

Un document qualifié d'archives doit donc pouvoir démontrer son/sa :

1. Authenticité : un document d'archives est authentique si l'on peut démontrer qu'il est bien ce qu'il prétend être et qu'il a été effectivement produit ou reçu au moment où il prétend l'avoir été (ISO, 2001a: 7). Les travaux du projet *International Research on Permanent Authentic Records in Electronic Systems* (InterPARES) établissent les

critères garantissant le caractère authentique d'un document d'archives, et ce, dès sa création. Ces critères sont au nombre de quatre, soit :

- o identification de manière certaine la personne physique ou morale à l'origine du document et les personnes qui concourent à son élaboration (destinataire, rédacteur et émetteur),
- o assignation de la/les date(s) de création (date de l'élaboration, de réception et de classement) et date(s) de transmission du document,
- o indication du sujet du document,
- o indication sur le lien d'archivage (code de classement, numéro d'identification du dossier) (Duranti, 2003; Maftai, [s.d.]).

Comme dans le cas d'une signature manuscrite traditionnelle, l'apposition d'une signature électronique sur le document permet à une personne de s'identifier et/ou de marquer son consentement par rapport à un contenu.

L'assignation des métadonnées complètes et suffisantes est un autre moyen pour assoir et maintenir l'authenticité des archives numériques (Nahuet, 2009). En plus des métadonnées descriptives du contenu du document, les métadonnées contextuelles indiquent la provenance du document; les métadonnées de gestion décrivent les actions apportées sur le document et, enfin, les métadonnées techniques contribuent à sa pérennisation en documentant sa technologie de développement.

2. **Fiabilité** : un document d'archives est réputé fiable s'il constitue une source d'information complète et exacte sur les activités dont il témoigne : « Un document fiable est un document dont le contenu peut être considéré comme la représentation complète et exacte des opérations, des activités ou des faits qu'il atteste, et sur lequel on peut s'appuyer lors d'opérations, d'activités ou de faits ultérieurs » (ISO, 2001a: 8). Le contrôle des procédures de création, de modification, de stockage et d'élimination concourt au respect de cette condition.
3. **Intégrité** : l'intégrité d'un document est le fait qu'il n'ait subi aucune modification ou altération qu'elle soit accidentelle ou intentionnelle, depuis qu'il a été créé (ISO, 2001a: 8). Trois critères cumulatifs sont nécessaires pour garantir l'intégrité, à savoir : la lisibilité du document, la stabilité de son contenu informationnel, la traçabilité des opérations sur ce document (Duranti, 2003).

Dans un environnement papier, l'intégrité est garantie par le support, alors que, dans un environnement numérique, la garantie d'intégrité est assurée par des moyens cryptographiques : l'apposition d'une signature électronique a pour effet de sceller numériquement le contenu d'un document et de pouvoir déceler toute altération du document même minime. Il est possible, par le contrôle de la signature, de démontrer l'intégrité du document présenté. Toutefois, si elle permet de les déceler, la signature électronique n'interdit pas d'éventuelles modifications du document (ex. : en cas de besoin de migration des données). L'intégrité des documents numériques peut être assurée, dans ce cas-là, en documentant les ajouts et les annotations apportées aux documents après leur création (traçabilité des opérations sur le document). Les circonstances des modifications apportées à un document doivent également être documentées et leurs auteurs clairement identifiés. L'intégrité pour un document numérique s'applique plus à son contenu qu'à sa structure et sa forme :

« Quand on se réfère au document électronique, on le considère essentiellement complet et inaltéré, si le message qu'il est censé communiquer pour accomplir son but est inaltéré. Ceci implique que son intégrité physique, telle que le nombre exact des chaînes binaires, peut être modifiée, à condition que l'organisation du contenu, les annotations nécessaires, ainsi que les éléments de la forme documentaire restent les mêmes. » (Maftai, [s.d.]

L'intégrité est aussi assurée par le choix de supports de stockage sécurisés permettant des contrôles d'intégrité des données (ex. : technologies WORM de stockage non réinscriptibles) et l'utilisation de formats de données ouverts.

4. Exploitabilité : un document d'archives réputé exploitable peut être en tout temps localisé, récupéré, communiqué et interprété (Duranti et al, 2002; McDonald et al., 2005: 11). Pour assurer l'exploitabilité des documents, il est nécessaire d'adopter des formats normalisés et de préserver les liens contextuels entre les documents tout en ayant la possibilité de situer chaque document dans son contexte individuel (contexte documentaire, contexte technologique, contexte procédural, contexte de la provenance et contexte juridique-administratif) (Maftai, [s.d.]; ISO, 2001: 8).

Ces conditions de validité confèrent aux archives leur caractère original (Öberg et Borglund, 2006: 63). Cette caractéristique est toutefois difficile à maintenir et à démontrer dans

un environnement numérique. En effet, les moyens de restitution et de conservation des archives ne présentent plus les conditions suffisantes pour préserver leur caractère original (Chabin, 2005: 156; Liu, 2004). Certains auteurs vont même jusqu'à réfuter l'existence d'un original dans le cas des archives numériques. Duranti conclut que, dans un univers numérique, il est impossible de conserver des documents numériques, il est seulement possible de préserver la capacité de les reproduire (Duranti, 2003).

L'authenticité, la fiabilité, l'intégrité et l'exploitabilité des documents d'archives, plus particulièrement des archives numériques, peuvent être protégées par des politiques et des procédures de contrôle propres à chacune des conditions de validité d'un document d'archives.

2.2.2.5 Capture et gestion dans un système d'archivage

Les archives se distinguent aussi par leur mode de gestion. Afin de protéger leurs valeurs primaire et secondaire et de respecter leurs conditions de qualité (authenticité, fiabilité, intégrité et exploitabilité), les archives doivent être gérées tout au long de leur cycle de vie (création, utilisation, conservation, exploitation) en les intégrant dans les systèmes archivistiques de l'organisation tels que le plan de classification et le calendrier de conservation. La capture des archives dans un système d'archivage dès leur création permet non seulement de préserver leurs conditions de validité (Bearman et Hedstrom, 2000), mais également de retracer les diverses versions d'un document et de comprendre les raisons pour lesquelles ces archives ont été créées (Öberg et Borglund, 2006).

L'usage des technologies de l'information et des télécommunications provoque la dématérialisation de l'information (Yusof et Chell, 1998: 102). Seule la capture d'un document dans le système archivistique de l'organisation justifie ses valeurs possibles et protège ses qualités :

« L'enregistrement d'un document ou d'une information dans un espace électronique institutionnel constitue, en soi, une reconnaissance de sa valeur administrative, [d'autant plus] qu'en l'absence d'une telle valeur on ne prendrait pas la peine de réaliser les opérations nécessaires à son enregistrement, si peu onéreuses ou si automatiques soient-elles. » (Parent et Boulet, 1999b: 7)

Le système d'archivage assure la capture des archives dès leur création et facilite leur gestion tout au long de leur cycle de vie (Bearman et Hedstrom, 2000; ISO, 2001a:8-9;

McDonald et al., 2005: 12). La norme française AFNOR NF Z42-013 transposée en norme ISO (ISO 14641-1 : 2012 - Archivage électronique - Partie 1 : spécifications relatives à la conception et au fonctionnement d'un système d'informations pour la conservation d'informations électroniques) recommande les mesures suivantes afin de respecter les exigences fonctionnelles d'un système d'archivage :

- Définir une politique d'archivage;
- Décrire de manière précise le dossier technique du système;
- Tenir un journal des évènements;
- Tenir un journal du cycle de vie;
- Dupliquer les objets archivés au moins sur deux sites distants (AFNOR, 2009).

En résumé, les archives se caractérisent essentiellement par la nature de l'information qu'elles véhiculent, par leur contexte de création et d'utilisation, par les valeurs qu'elles acquièrent tout au long de leur cycle de vie (Chabin, 2000; Couture, 1996; Gagnon, 2002; Rousseau et Couture, 1994). Ces caractéristiques, propres aux archives, permettent de les situer par rapport à d'autres objets documentaires et conditionnent, par conséquent, leur mode de gestion.

2.3 Stratégies d'archivage du Web

Le Web supplante divers supports documentaires. On assiste à une migration progressive de formes documentaires traditionnelles vers la toile mondiale (des encyclopédies, des journaux, des publications scientifiques, des émissions radiophoniques et télévisées, etc.). Plusieurs éditeurs, créateurs de contenus et organismes favorisent cet outil comme moyen de diffusion et de communication de leur production documentaire. Mains organismes, publics et privés, offrent des produits et des services exclusivement sur le Web.

Depuis les débuts du Web, divers programmes et projets visant la sauvegarde des traces du Web ont été entrepris aussi bien par des organismes privés que par des institutions de mémoire (bibliothèques nationales, archives nationales, services d'information, etc.). Ces initiatives d'archivage du Web se démultiplient et se diversifient. De nombreuses institutions de mémoire tentent de préserver des fragments importants et représentatifs du Web et de dépasser

une simple logique de « stockage » pour l'inclure dans une démarche de « patrimonialisation », de valorisation et d'exploitation des contenus archivés (Dumont, 2005).

Des efforts de coopération sont investis à l'échelle internationale pour déterminer les critères de sélection, les procédés de capture et les modes d'accès les plus appropriés aux contenus Web capturés. De nouvelles approches et techniques de gestion documentaire largement dépendantes de la nature des ressources en ligne sont alors expérimentées. L'extension du dépôt légal aux publications en ligne a officialisé l'intégration des ressources Web aux collections et fonds documentaires existants. Dès lors, la recherche des schémas de gestion et des cadres de préservation les mieux adaptés aux propriétés de la production documentaire issue des technologies Web devient un souci majeur pour les institutions de mémoire.

Nouvelles venues en bibliothèques et autres services d'information, les archives du Web nous renseignent, à titre d'exemple, sur l'impact d'une campagne électorale en ligne sur les choix des électeurs, sur les pratiques des adolescents sur Internet ou encore sur le rôle de YouTube ou DailyMotion dans la provocation de mouvements d'opinion. Le Web véhicule des traces uniques de phénomènes technologiques, sociaux, économiques et culturels du monde virtuel. Ces traces permettent de préserver une mémoire individuelle et collective d'une expérience sans précédent vécue par toute une génération (Daffini, 2010) et devraient être ainsi préservées au risque de perdre des fragments importants de notre mémoire collective (UNESCO, 2003).

Malgré la prépondérance des technologies Web sur d'autres moyens de communication, la masse documentaire que véhicule le Web se caractérise par sa fragilité. De même, cette masse se distingue par sa complexité et son hétérogénéité et, surtout, sa capacité à évoluer sans cesse (flux RSS, commentaires des internautes, partage et réutilisation des contenus, etc.). Cette masse grandissante échappe à tout contrôle, car « la logique du Web n'inclut pas la persistance ni la pérennité, mais l'échange et la reprise » (Bachimont et al., 2005: 5). La mémoire du Web devrait donc être conservée pour les besoins actuels et futurs (Gladney, 2007: 3; UNESCO, 2003).

Quelle unité documentaire faut-il considérer lors de la collecte des sites Web? Quelles sont les frontières spatio-temporelles d'une page Web? Quels critères de sélection faut-il considérer? Quels types de sites Web sont les plus représentatifs de notre société (site Web académique, gouvernemental, associatif, etc.)? À quel niveau de profondeur et à quelle fréquence de collecte faut-il capturer le tissu de liens dans un site Web? À quel moment de son évolution faut-il conserver un site Web? Quel niveau de pertes d'information est-il possible d'accepter? Est-ce que le Web social devrait aussi être conservé? Qui devrait avoir la charge de collecter et de conserver la mémoire de notre Web contemporain? Où se distinguent les frontières entre une publication et un document d'archives sur le Web? Quels savoirs et savoir-faire faut-il acquérir pour réaliser cette mission d'archivage du Web? Comment assurer la pérennité et l'exploitabilité des archives Web dans le temps? Voilà quels sont les types de questions à considérer lorsqu'il s'agit d'initier un projet d'archivage du Web, car, dans les termes de Bachimont et ses collaborateurs, l'archivage du Web n'est pas un acte de conservation, mais plutôt un acte de constitution de la mémoire documentaire en ligne (Bachimont et al., 2005). Il serait nécessaire de cerner plus précisément les attentes des futurs chercheurs et historiens, car ces informations donneraient des orientations pour constituer les collections du Web à constituer (Daffini, 2010).

Les initiatives actuelles d'archivage du Web opèrent selon différents modes de collecte inspirés des pratiques traditionnelles ou complètement réinventées. Plusieurs organismes de mémoire, notamment les bibliothèques nationales, dans une logique de continuité des collections, recueillent par dépôt légal des ressources sur le Web. D'autres, en l'absence d'un cadre légal réglementant l'acquisition des documents Web, doivent reconsidérer leurs méthodes (Gomes et al., 2011).

Deux principales perspectives d'archivage du Web se dégagent au travers l'analyse de la littérature et de la documentation des principaux projets de préservation du Web (voir Annexe 1 : Principales initiatives d'archivage du Web (dernière mise à jour : novembre 2012), p. i) :

- La perspective documentaire : cette perspective est la plus répandue. Le document Web, à cause de sa nature publique, tend à être défini comme une publication et non comme un document d'archives. C'est la perspective dominante malgré le manque de support légal et les enjeux liés à la préservation et à l'accès aux contenus capturés.

- La perspective archivistique : cette perspective transparaît dans quelques projets (archivage du Web gouvernemental, archivage du Web universitaire). Les projets sont peu documentés et se distinguent par leur rareté. Par contre, force est de constater l'abondance des politiques d'archivage et des principes directeurs pour la gestion et la conservation du Web organisationnel.

Nous présentons dans les sections suivantes les principales stratégies d'archivage du Web de même que les techniques et outils employés pour capturer, gérer et conserver les matériaux capturés selon ces deux perspectives.

2.3.1 Perspective documentaire

2.3.1.1 Modes d'archivage

Les initiatives actuelles opèrent selon différents modes de collecte des documents Web inspirés des pratiques traditionnelles ou complètement réinventées. Plusieurs organismes de mémoire, notamment les bibliothèques nationales, dans une logique de continuité de leurs collections et de leur mission recueillent, par dépôt légal, des ressources sur le Web. D'autres, en l'absence d'un cadre légal réglementant l'acquisition des documents Web, réinventent leurs méthodes ou font appel à des moyens traditionnels de développement des collections.

2.3.1.1.1 Dépôt légal des ressources en ligne

Les institutions de mémoire, que représentent, entre autres, les grandes bibliothèques, ont une longue tradition de mission de collecte, de conservation et de diffusion du patrimoine documentaire, composante de l'héritage culturel et scientifique d'un pays, au travers de politiques documentaires. Cette mission patrimoniale est opérationnalisée grâce à des mécanismes d'accroissement des collections tels que le dépôt légal, le don, l'échange, le versement, l'achat.

Le dépôt légal a pour objectif la conservation de tous les documents publiés dans un pays ou sur un pays. Il s'agit d'une incitation à faire connaître et à déposer une publication dans une institution désignée, le plus souvent la bibliothèque nationale d'un pays (Chaimbault, 2008: 11-13). Le dépôt légal est un mécanisme qui permet la collecte d'une quantité exhaustive et représentative de documents qui offre aux utilisateurs le matériel nécessaire pour appréhender et analyser valablement une époque, un phénomène, un événement, etc. (Gharsallah, 2002: 41).

La législation du dépôt légal a évolué notamment sous l'influence des technologies de l'information afin de prendre en compte de nouveaux types de documents et de nouvelles modalités techniques de collecte des documents (Chaimbault, 2008: 13; Gatenby, 2004; Lemaître, 2001: 37-39). Dans de nombreux pays, la loi sur le dépôt légal a été revue et étendue aux documents numériques incluant les sites Web. Il est donc possible de procéder par collecte automatique à distance ou à la source des documents en ligne à l'aide d'un robot moissonneur (*harvester*) ou par dépôt en ligne par le producteur ou le diffuseur des contenus en ligne (Gharsallah, 2002; Masanès, 2005).

Grâce à l'extension du dépôt légal en 1992 aux documents numériques, Bibliothèque et Archives nationales du Québec (BAnQ) procède depuis 2001 à la collecte des publications en ligne produites par divers organismes gouvernementaux et publics au Québec. Après une phase de collecte rétrospective des publications en ligne entre 2001 et 2004, la collecte s'est systématisée. Les publications soumises au dépôt légal doivent être signalées par l'organisme producteur au début de leur vie active par le biais d'un formulaire électronique en ligne (formulaire de dépôt d'une monographie, d'un périodique ou d'un numéro de périodique). Le document est ensuite capturé directement sur le site de l'organisme une fois que ce dernier en autorise l'accès, et ce, grâce aux logiciels MemoWeb et HTTrack. Les publications en ligne sont ensuite décrites, indexées et intégrées au catalogue IRIS (BAnQ, 2001; Léger, 2003). La recherche dans le catalogue est possible en texte intégral, par titre, auteur ou sujet. Il est également possible de repérer une publication archivée à l'aide du cadre de classification des Publications Gouvernementales du Québec (PGQ). Les organismes dépositaires peuvent inclure sur leurs sites Web un lien vers la publication archivée à BAnQ. La collecte est mensuelle ce qui permet de récupérer les diverses éditions d'une même publication. Le format de préservation choisi est le format PDF puisque les publications gouvernementales sont produites dans ce format. De plus, ce format, devenu une norme internationale (ISO 32000-1:2008 - Document management -- Portable document format - Part 1), est admis comme format pérenne de conservation (BAnQ, 2007).

Bibliothèque et Archives Canada (BAC) a initié, en 1994, le projet PPPE : projet pilote des publications électroniques visant la préservation du patrimoine canadien en ligne. Par le biais du dépôt légal électronique, BAC collecte « toute source d'information concernant les

Canadiens, ou produite par les Canadiens qu'ils soient sur le territoire ou ailleurs. » (Lemaître, 2001: 55). Le document *Politiques et directives relatives aux publications électroniques diffusées en réseau* (1998) spécifie un système à trois niveaux pour recueillir les publications électroniques :

1. Le niveau "Archivé" correspond aux publications intégrées dans la collection de la bibliothèque;
2. Le niveau "Versé" correspond aux publications destinées à être intégrées dans la collection de la bibliothèque, mais faisant encore objet de négociation avec des éditeurs ou autres institutions d'archivage;
3. Le niveau "Lié" correspond aux publications conservées sur un autre serveur, mais accessibles par redirection de liens URL (BAC, 1998; Chebbi, 2007-2008; Lemaître, 2001: 56).

2.3.1.1.2 Dépôt volontaire des ressources en ligne

Le dépôt légal n'est pas toujours le principal mode de collecte des documents Web par les bibliothèques nationales. Bien que plusieurs pays ont tout récemment étendu leur loi sur le dépôt légal aux publications en ligne (ex. : en France), dans d'autres cas, la loi sur le dépôt légal n'a pas été révisée. Par conséquent, elle n'inclut pas les publications en ligne ou les sites Web (ex. : en Australie).

Afin de détourner ces limites légales qui menacent la survie d'une portion importante du patrimoine documentaire d'un pays, quelques institutions à vocation patrimoniale font appel à d'autres modes de collecte, notamment le dépôt volontaire des documents Web par leur créateur en attendant l'évolution du cadre légal (cas de la State Library of Tasmania, de la Finnish University Library et de la National and University Library of Slovenia).

La National library of the Netherlands (Pays-Bas), dépourvue de tout support législatif en ce qui concerne le dépôt légal des publications en ligne, a mis en œuvre le programme « e-depot » dans l'objectif d'inciter les éditeurs de contenus, notamment ceux bien établis tels qu'Elsevier Science et Kluwer Academic, à fournir régulièrement sur une base volontaire des copies de leurs publications électroniques ou bases de données. L'exploitation des documents

archivés doit faire, dans ce cas, l'objet d'ententes spéciales (Chaimbault, 2008; Lemaître, 2001: 85-86).

2.3.1.1.3 *Moissonnage à distance du Web public*

Conscientes de la fragilité et de l'éphémérité des documents Web, certaines institutions ont procédé à des collectes automatisées de portions importantes du Web public (Day, 2003b). L'exemple par excellence est celui d'Internet Archive qui procède à la capture de portraits réguliers de tout le Web mondial (*snapshot*). Internet Archive indique avoir mis à disposition des chercheurs 80 téraoctets de contenus Web archivés (2,7 milliards de captures) obtenus entre mars et décembre 2011. Le 25 octobre 2012, Internet Archive a fêté le cap des 10 pétaoctets de matériel capturé sur le Web⁵.

La France a également adopté temporairement ce mode de collecte. La Bibliothèque nationale de France (BnF) a procédé à l'archivage des sites Web français depuis 1999 malgré les limites d'accès aux contenus archivés imposées par les cadres législatifs. Cette institution a limité l'accès aux archives collectées jusqu'à la promulgation de la loi DADVSI en 2006 (Chaimbault, 2008; Gharsallah, 2002; Illien et Games, 2006; L'Hostis, 2006). Aujourd'hui, les contenus archivés sont consultables dans les locaux de la BnF.

Plusieurs autres institutions opèrent selon ce mode de collecte. Elles limitent également l'accès en ligne aux collections archivées par respect du droit de la propriété intellectuelle et du droit d'auteur (cas de l'University of Heidelberg et de la National Library of Sweden).

2.3.1.1.4 *Archivage rétrospectif*

Des collectes rétrospectives sont également effectuées pour combler les brèches dans les archives Web de différentes institutions de mémoire. La Library of Congress a fait appel aux services d'Internet Archive pour récupérer des morceaux du Web américain ou de sites Web relatifs à des événements particuliers essentiels aux recherches de la clientèle de cette institution et importants pour l'histoire des États-Unis. D'autres organismes font appel à des tiers-archivistes (Archive-It, European Archive, HenzoWeb) faute de moyens techniques (serveurs de

⁵ GNT-Génération Internet (01 Novembre 2012). *The Internet Archive : 10 millions de gigaoctets archivés*. Adresse URL : <http://www.generation-nt.com/the-internet-archive-wayback-machine-actualite-1646242.html> (Page consultée le: 11 novembre 2012).

stockage, robots de collecte, etc.) ou d'expertise dans le domaine de l'archivage des documents Web (Brown, 2006). Internet Archive et la fondation European Archive offrent souvent du soutien technique et technologique à diverses institutions patrimoniales (Chebbi, 2009).

Quelques institutions impliquent les internautes dans leurs projets d'archivage du Web. Elles invitent leur public à suggérer des sites d'intérêt à archiver (Bibliothèque nationale de Singapour, le consortium des bibliothèques du Royaume-Uni, la bibliothèque nationale universitaire de Zagreb). À titre d'exemple, la Library of Congress a mis sur pied, en collaboration avec Internet Archive, le programme K12 où les enseignants et les élèves sont invités à soumettre des suggestions de sites Web qu'eux considèrent d'intérêt scientifique et historique (Archive-it, [s.d.]).

D'autres institutions prévoient l'achat d'archives Web à l'instar de l'University of Heidelberg ou la prise en charge de sites Web après le décès de leur concepteur (le projet AfterLife).

2.3.1.2 Approches d'archivage

La portée et l'étendue des collectes du Web rejoignent des objectifs fort différents. Certains projets visent la collecte de portions importantes du Web, d'autres se limitent à un périmètre bien délimité ou à un type de contenu particulier. La section, ci-après, présente les principales approches pour l'archivage des ressources en ligne.

2.3.1.2.1 Approche intégrale ou collecte automatique de tout le Web

L'approche intégrale vise une collecte automatique de tout le Web, plus précisément de tout ce qui est accessible aux robots d'archivage (*crawlers* ou *harvesters*). En effet, les techniques actuelles restent limitées quant à l'archivage du Web profond et du Web dynamique. Internet Archive a procédé depuis les débuts du Web à des collectes périodiques et fréquentes de la toile mondiale (Masanès, 2005). Malgré les limites de cette approche, le matériel collecté offre une image représentative du Web en surface à chaque moment de capture. La démarche est considérée objective puisqu'aucun critère de sélection ou limite ne sont imposés au moment de la collecte. Le matériel reste toutefois parsemé de trous (images manquantes, liens dysfonctionnels) (Day, 2003a: 39; Gharsallah, 2002; Masanès, 2005).

2.3.1.2.2 *Approche exhaustive ou collecte de l'ensemble d'un domaine Web*

L'approche exhaustive permet de collecter un périmètre bien défini du Web sans critères supplémentaires de sélection. Souvent, cette approche est adoptée pour capturer l'ensemble d'un domaine Web (ex. : .ca), ou d'un sous domaine Web (ex. : .qc.ca). Elle repose sur les mêmes fondements que l'approche intégrale et présente les mêmes limites. Elle garantit aussi la collecte d'une image représentative du domaine Web au moment de sa collecte. La National Library of Sweden est parmi les premières bibliothèques nationales ayant initié un projet d'archivage du Web suédois, le projet KulturarW3, en adoptant cette approche (Gharsallah, 2002; Lemaître, 2001).

L'approche intégrale et l'approche exhaustive sont techniquement faisables. Par contre, elles sont consommatrices de temps et de ressources. Elles se heurtent souvent aux problématiques liées au Web profond, aux limites des robots de collecte et aux lois en matière de propriété intellectuelle et de droit d'auteur (Brown, 2006: 30; Chaimbault, 2008; Charlesworth, 2003; Day, 2003a: 39).

2.3.1.2.3 *Approche sélective ou collecte ciblée*

L'approche sélective repose sur le moissonnage (*harvesting*) d'une catégorie de sites Web (Web gouvernemental, Web politique, etc.) ou d'un format particulier de documents (documents statiques, documents sonores en ligne, documents textuels, blogues, etc.). Tel qu'expliqué par Brown (2006), le processus de sélection repose sur un ensemble d'étapes successives. Tout d'abord, il faut définir une politique d'archivage qui doit préciser le contexte de la collecte, la méthode de sélection choisie et les critères de sélection. Les finalités de la collecte doivent être en accord avec la mission de l'organisme, les politiques en place et les collections conservées (Brown, 2006: 26). Cette démarche implique ensuite une définition plus fine des contours des ressources à archiver et un choix réfléchi des périodes et des fréquences de collecte (Brown, 2006: 28; Day, 2003a: 39).

Masanès distingue trois principales méthodes de sélection : (1) méthode centrée sur les domaines Web, (2) méthode centrée sur les genres Web, et (3) méthode centrée sur un événement particulier (Masanès, 2005: 75). Brown en identifie deux autres : méthode centrée sur un thème ou un sujet et méthode centrée sur des créateurs particuliers (Brown, 2006: 31).

Certaines institutions de mémoire procèdent à la collecte de territoires Web bien définis tels que le Web régional (Hallgrímsson et Bang, 2003; Gharsallah, 2002; Région wallonne, 2009) ou le Web municipal (Haettiger, 2003a, 2003b). Le projet hollandais Archipol ainsi que le projet *Political Communication Web Archiving Project* (PCWA) initié par le Center for Research Libraries (CRL) aux États-Unis ont pour objectifs la collecte de sites Web politiques (CRL, 2002-2005).

La Library of Congress collecte les sites Web relatifs à des événements ponctuels, mais importants, qui marquent l'histoire des États-Unis tels que : les sites relatifs aux élections présidentielles de 2000 et 2002, les sites Web documentant les événements du 11 septembre ou les Jeux olympiques (Beagrie, 2003). La BnF a entrepris des projets similaires visant la capture des contenus véhiculés sur le Web durant les élections législatives et présidentielles en France, de même que les élections régionales et européennes (Bachimont et al., 2005; Illien et Game, 2006; Masanès, 2005).

Plusieurs groupes de recherche développent progressivement des collections disciplinaires de contenus Web en lien avec leur domaine de recherche. Le Groupe d'études sur l'usage des TIC dans les migrations a mis en œuvre un projet pour collecter les sites Web sur les migrations et les mouvements sociaux (Groupe d'études sur l'usage des TIC dans les migrations, 2005; Rygiel et al., 2005). Le projet Digital Archives for Chinese Study (DACHS) en est un autre exemple. Ce projet est mené par l'University of Heidelberg et a pour objectif l'aspiration de ressources utiles à l'étude de la culture et des phénomènes sociaux en Chine véhiculés par le Web chinois (Brown, 2006: 31; Chaimbault, 2008: 27).

2.3.1.2.4 *Approche intégrée*

Les acteurs impliqués dans les projets d'archivage du Web expérimentent en continu différentes approches et techniques (Gladney, 2007: 3). Chaque approche présente des avantages et des limites. Les projets les plus récents adoptent une approche intégrée qui consiste à combiner plusieurs de ces approches d'archivage, ce qui permet de pallier aux limites de chacune et de constituer des archives plus complètes et de qualité (Chaimbault, 2008: 27; Gharsallah, 2002).

L'exemple le plus représentatif est celui de l'expérience française. La BnF a mené depuis 1994 plusieurs expérimentations avant de choisir une approche intégrée (Lemaître, 2001: 129-131). En effet, l'équipe responsable de l'archivage du Web établit une liste exhaustive d'URL à partir d'un premier balayage du Web français avec un robot de moissonnage. Cette liste est complétée par les suggestions d'un groupe de bibliothécaires veilleurs qui surveillent le Web et rapportent les coordonnées des sites présentant un intérêt pour la collection de la BnF. Une collecte automatique du Web français est réalisée grâce à des robots de collecte. Le matériel récolté est contrôlé et trié afin de détecter des contenus manquants (images, liens corrompus, pages Web manquantes, etc.) et d'éliminer les contenus qui ne représentent aucun intérêt pour la collection. Suite aux vérifications, des collectes manuelles et automatiques sont lancées pour compléter les collections initiales. Cette démarche est complétée par le dépôt légal des publications électroniques (Beagrie, 2003; Illien et Game, 2006; Masanès, 2000).

2.3.1.2.5 *Approche collaborative ou collaboration interinstitutionnelle*

L'approche collaborative consiste à former des alliances stratégiques et à développer des collaborations fructueuses avec les différents milieux documentaires et les acteurs clés de l'archivage du Web, et ce, à l'échelle nationale, régionale ou internationale (Beagrie, 2003; Chebbi, 2009). Des modèles coopératifs se dessinent progressivement. L'échange des expériences respectives favorise l'enrichissement de la réflexion sur les pratiques d'archivage du Web tant sur le plan théorique que pratique.

La littérature professionnelle sur le sujet et la documentation offerte sur les différentes initiatives d'archivage du Web nous permettent de distinguer deux modes de collaboration, soit la coopération entre institutions de mémoire et acteurs privés de l'archivage du Web, soit la collaboration entre institutions de mémoire à l'échelle régionale et internationale.

L'échange d'expertises et le soutien technique ainsi que logistique qu'offre Internet Archive aux institutions de mémoire sont un exemple de modèle de coopération publique privée (Illien, 2009). Internet Archive assiste plusieurs bibliothèques nationales et organismes de recherche lors de la mise en œuvre de leurs propres projets d'archivage. Les travaux d'Internet Archive et ses développements techniques ont permis la sauvegarde du Web dès ses débuts. Grâce à cette initiative, des collectes rétrospectives du Web sont possibles. Internet Archive met

à la disposition de plusieurs institutions des outils et logiciels d'archivage et offre un accès public au matériel archivé.

Le partage de responsabilités d'archivage de certains contenus est une autre forme de coopération interinstitutionnelle. La BnF et l'Institut National de l'Audiovisuel (INA) coordonnent leurs projets d'archivage en fonction des contenus qui leur incombent de conserver dans une logique de continuité des collections. Ainsi, l'INA est responsable de l'archivage du patrimoine audiovisuel français en ligne, la BnF du Web français (Chebbi, 2009; Illien et Game, 2006).

L'UK Web Archiving Consortium (UKWAC) est une alliance qui réunit six institutions œuvrant dans des domaines différents (domaines médical, scientifique et culturel). Ces institutions se partagent la mission d'une collecte exhaustive de portions complémentaires du Web national britannique selon une approche thématique basée sur le champ d'expertise des institutions suivantes :

- Le Wellcome Trust est responsable du Web médical;
- Le JISC (Joint Information Systems Committee) collecte les sites Web relatifs à la recherche et à l'enseignement;
- La National Library of Wales collecte le domaine Web gallois;
- La National Library of Scotland est responsable du patrimoine écossais en ligne;
- The National Archives préservent le Web gouvernemental;
- La British Library se charge du domaine Web britannique (Brown, 2006; Chaimbault, 2008: 39-40; Verheul et al., 2006).

Un modèle similaire est développé à l'échelle régionale. Le Nordic Web Archive (NWA), projet initié en 2000, regroupe plusieurs bibliothèques scandinaves (la Royal Library in Copenhagen, la Helsinki University Library of Finland, la National and University Library of Iceland, la National Library of Norway et la Royal Library in Stockholm). L'objectif principal de l'initiative est de réfléchir à l'indexation et la diffusion du matériel archivé. Les échanges sont fructueux : un système a été développé facilitant un accès centralisé au matériel archivé par les bibliothèques participantes et offrant des options de recherche avancées (IIPC, 2004-2008; Illien, 2009).

L'initiative australienne est fondée aussi sur une approche collaborative. En effet, la bibliothèque nationale se partage la mission de l'archivage du Web avec les archives nationales en fonction de leurs mandats respectifs (NAA, 2007). La collecte du Web organisationnel relève de la responsabilité des archives nationales alors que la capture des publications en ligne est de la responsabilité de la National Library of Australia (NLA) qui chapeaute un réseau de bibliothèques provinciales assurant des collectes bien ciblées du domaine Web australien. Seules les publications présentant un intérêt scientifique ou culturel sont collectées une fois les autorisations de leurs auteurs obtenues (Beagrie, 2003; Day, 2003b; Lemaître, 2001; McDonald, 2005).

La NLA s'est rapidement rendu compte que cette approche ne permet pas la préservation d'un matériel représentatif du patrimoine australien en ligne. La bibliothèque nationale a donc conclu en 2007 une entente avec Internet Archive pour procéder à un moissonnage exhaustif et automatique du Web australien.

Au plan international, l'International Internet Preservation Consortium, l'IIPC, forme un lieu d'échange et de réflexion sur les travaux et projets liés à l'archivage et à la préservation des ressources Web (IIPC, 2004-2008). Les membres partagent leurs expériences dans l'objectif de développer des méthodologies, des outils, des techniques et des standards communs pour la préservation du Web (Illien, 2009). Un transfert des connaissances et de compétences s'effectuent aussi entre les institutions membres de l'IIPC, l'exemple de BnF et BAnQ qui ont signé en 2010 une entente de collaboration, entre autres, dans le domaine l'archivage du Web et du dépôt légal numérique (BAnQ, 2010).

2.3.1.3 Description des principaux projets d'archivage du Web

L'annexe 1 (p. i) présente une recension des initiatives d'archivage du Web les plus documentées qu'elles soient dans une perspective documentaire ou archivistique. Les ressources ont été initialement identifiées à partir du portail australien sur l'archivage du Web, le portail Preserving Access to Digital Information (PADI) et d'une bibliographie complémentaire sur les projets d'archivage développée par les responsables de l'initiative autrichienne (Aschenbrenner, 2004). Ce tableau a été enrichi lors de la recension des écrits et suite à un processus de recherche et de veille sur le Web.

Le tableau définit le projet, les modes d'archivage utilisés pour l'acquisition des documents Web, l'institution responsable de la collecte et de la conservation des contenus Web, la nature des contenus archivés et, enfin, les approches et techniques d'archivage choisies. Nous avons également distingué les projets qui relèvent du domaine des archives de ceux qui doivent découler de l'application de la loi sur le dépôt légal (un astérisque a été ajouté au début du nom du projet). Nous nous sommes basée sur la nature des contenus archivés et non sur l'organisme responsable de l'archivage car, dans bien des cas, les sites Web collectés, en application de la loi sur le dépôt légal, sont plutôt du domaine des archives et non des publications en ligne, comme pour le cas du projet CyberCemetery (University of North Texas Libraries, 2009).

De l'analyse de ces initiatives se dégage des paramètres à considérer avant d'entreprendre des captures massives ou sélectives des archives Web. Nous avons identifié trois principaux paramètres :

- Contenu : ce paramètre est déterminé en fonction de la politique documentaire de l'institution responsable de l'archivage (sites Web, fichiers Web, Web invisible, Web public, fichiers en streaming, etc.).
- Fréquence ou périodicité des captures : ce paramètre est subordonné à la spécificité et au niveau de stabilité du document à archiver. Même si ce critère est difficile à évaluer, les robots moissonneurs permettent de collecter des informations sur la longévité et la fréquence de mise à jour d'un site Web.
- Niveau de profondeur : la profondeur délimite le nombre d'unités à collecter en fonction des contenus documentaires visés. À titre d'exemple, les robots de collecte pourraient être programmés pour aspirer un domaine Web ou des pages Web bien particulières (Bonnenfant, 2008: 39-42; Brown, 2006: 33-39).

Même s'il est aujourd'hui possible de paramétrer les robots de collecte et les diriger vers des territoires Web bien spécifiques, les frontières d'un site Web ou d'un périmètre de collecte restent difficiles à définir. La nature hypertextuelle du Web entrave la définition de frontières claires de collecte. De plus, les critères de sélection ne sont pas évidents à établir. De même, les techniques actuelles ne permettent pas la reconstitution de documents Web complets et authentiques. La préservation à long terme et l'accessibilité (visualisation et recherche) au matériel archivé constituent un réel défi pour les spécialistes de l'information.

2.3.2 Perspective archivistique

2.3.2.1 Mode de gestion

2.3.2.1.1 *Gestion des risques*

Malgré leur succès, les systèmes Web n'incluent pas les fonctionnalités nécessaires à la bonne gestion et conservation de la mémoire documentaire d'une organisation (Barry, 2004: 29; Phillips, 2003: 46; Rein et al., 1997: 88). Plusieurs auteurs suggèrent l'inclusion de modules de gestion documentaire lors du développement ou de l'implantation d'un SIW afin de maîtriser le cycle de vie des documents qui y sont produits (Balasubramanian et Bashia, 1998: 112; Barry, 2004; Frost, 2001: 34-36).

L'usage des technologies Web dans les organisations provoque un accroissement exponentiel du volume des documents produits (Cruse et al., 2003: 6; Cunningham et Phillips, 2005: 303; Phillips, 2003: 44). Ces derniers sont de plus en plus complexes, instables, fragiles et éphémères (Koehler, 1999; Huang et al., 2008). Les nouvelles propriétés héritées par le document sur le Web laissent entrevoir des problèmes de gestion et de conservation de taille liés principalement à :

- La permanence et la préservation des documents;
- L'intégrité des documents;
- La conception d'outils adéquats pour la recherche et le repérage de l'information (Foo et Lim, 1997).

Conscientes des risques auxquels s'expose une organisation en absence d'un cadre de gestion des documents Web, plusieurs institutions d'archives nationales proposent des modèles de politique d'archivage du Web, des recommandations de meilleures pratiques pour contrôler la création, la gestion et la préservation des sites Web. Le modèle de gestion des risques a été proposé par McClure et Sprehe en 1998 et adopté par de nombreuses institutions archivistiques comme cadre théorique pour la gestion, l'évaluation et l'archivage des sites Web. Ce modèle propose que les organisations doivent procéder à l'évaluation des objectifs, de la complexité, de la visibilité et de la fréquence de mise à jour de leurs sites Web (intranet, extranet ou site Web public) afin de déterminer le niveau de vulnérabilité des documents qui y sont publiés et établir en conséquence les critères de sélection des contenus à archiver, déterminer la fréquence et la

profondeur des collectes à réaliser ainsi que les méthodes et techniques d'archivage à adopter (McClure et Sprehe, 1998).

La gestion des risques consiste à analyser la vulnérabilité des contenus Web dans le temps en se basant sur certains facteurs liés aux caractéristiques du SIW. L'unité d'analyse pourrait être le site Web au complet, une portion du site Web consacrée à des fonctions ou processus organisationnels précis ou encore un ensemble de pages Web sur un sujet donné. La gestion des risques facilite, entre autres, l'identification des acteurs organisationnels impliqués dans la création, la diffusion et la gestion des contenus Web. Elle permet également d'établir une cartographie des contenus à capturer et à retenir pour une conservation à long terme de même que les activités et les services fournis ou documentés par le site Web.

Les résultats et les informations colligés orientent les décideurs vers un choix rationnel et approprié des approches et modèles d'archivage. Selon le niveau du risque que présente un site Web, des mesures particulières sont recommandées. Le **Tableau I** **Erreur ! Source du renvoi introuvable.** (p. 59) présente le modèle de gestion des risques établi dans le rapport de McClure et Sprehe (1998).

Tableau I : Description de l'approche de gestion des risques associés au site Web.

NIVEAU DU RISQUE	MESURES RECOMMANDÉES
<p><u>Faible</u>, si :</p> <ul style="list-style-type: none"> • L'organisation contrôle et gère efficacement les droits d'accès et d'édition des contenus Web. • Le site Web présente principalement des documents statiques, diffusés également sur d'autres supports documentaires. • Les contenus affichés sont déjà identifiés et intégrés dans un système d'archivage. • Les documents publiés ou diffusés sur le site Web présentent un risque de revendications ou de contestations minime, voire nul. 	<p>Mode de gestion des contenus Web similaires à des documents sur support traditionnel :</p> <ul style="list-style-type: none"> • Gestion du cycle de vie des documents publiés ou diffusés sur le site Web. • Identification et capture des documents Web dans un système d'archivage. • Évaluation et tri des documents Web (application d'un calendrier de conservation). • Contrôle continu de la vulnérabilité redditionnelle du site Web.

NIVEAU DU RISQUE	MESURES RECOMMANDÉES
<p><u>Moyen</u>, si :</p> <ul style="list-style-type: none"> • Les mesures de contrôle des droits de gestion et d'édition des contenus Web sont inadéquates. • Les documents présents sur le site sont de plus en plus de nature dynamique. • Les documents publiés ou diffusés sur le site ne sont pas tous identifiés et capturés dans un système d'archivage. • Les documents publiés ou diffusés sur le site Web présentent un risque de revendications ou de contestations. 	<p>Des mesures supplémentaires sont à appliquer dont :</p> <ul style="list-style-type: none"> • Maintien d'un registre historique : garder la traçabilité de toutes les modifications apportées à la structure ou au contenu du site Web. • Capture de vues (ou d'instantanés) périodiques du site Web.
<p><u>Élevé</u>, si :</p> <ul style="list-style-type: none"> • Les mesures de contrôle des droits d'accès et d'édition sont défaillantes. • Les contenus présents sur le site sont principalement sous format natif Web, de nature dynamique et interactive, souvent générés à la volée. • Les documents affichés ne sont pas intégrés dans un système d'archivage. • Les contenus publiés sont souvent l'objet de contestations ou de revendications. 	<p>Dans ce cas, il faut :</p> <ul style="list-style-type: none"> • Revoir les procédures de gestion documentaire. • Maintenir un registre historique. • Réaliser des captures de vues fréquentes du site. • Capturer et intégrer le registre historique et les captures de vue dans le système d'archivage.

2.3.2.1.2 Gestion du cycle de vie des archives – Records Continuum

À l'ère du numérique et des technologies de l'information, les institutions archivistiques ne peuvent plus se confiner à un rôle de gardien de la mémoire (*custody age*). Elles doivent être, plus que jamais, proactives et innovantes. Ces institutions doivent orienter, conseiller et assister les organisations dans la mise en œuvre des meilleures pratiques de gestion de leur mémoire documentaire (*postcustody age*). Les documents doivent être contrôlés et gérés depuis leur création et tout au long de leur cycle de vie (McKemish, 2001; Upward, 2005). La mission de l'archiviste ne consiste plus en un traitement de masse d'un volume croissant des archives. Avec la profusion de l'usage des technologies de l'information et les impacts qui y sont associés sur la mémoire organisationnelle, l'archiviste doit se concentrer davantage sur l'identification, au sein de la masse documentaire produite, des documents présentant une valeur archivistique. Il doit, notamment, veiller à ce que des métadonnées soient associées au contenu afin de bien documenter le contexte et la provenance de l'information (Henry, 2000: 570-571), à ce que les

documents produits respectent les règles et standards en cours et à ce que des mesures pour leur préservation à long terme soient définies à l'avance.

Cette conception du rôle proactif de l'archiviste se confirme à l'ère du numérique et s'avère être à la base du modèle du Records Continuum, concept développé et théorisé par les Australiens où l'action de l'archiviste s'inscrit dans un continuum d'espace-temps. Archivistes et gestionnaires de documents administratifs (*records managers*) doivent collaborer et intervenir dès la création du document, voire même dès la conception des systèmes d'information afin de prévoir les fonctionnalités et exigences de gestion à opérationnaliser (Bearman et Hedstrom, 2000; Upward, 2005).

Ce modèle dépasse la dichotomie instaurée par le modèle du cycle de vie des archives où l'on spécifie un terrain et une période d'intervention, un espace-temps, pour les gestionnaires de documents administratifs (gestion des archives courantes et intermédiaires) et un terrain et une période d'intervention pour les archivistes (gestion des archives définitives). Les archives numériques requièrent une intervention en amont et un contrôle continu de leur cycle de vie (McKemmish, 2001). Nous avons choisi l'usage du terme archiviste pour désigner ces deux groupes de professionnels de la gestion des documents et des archives.

Frank Upward critique la vision cyclique sur laquelle est fondée la théorie des trois âges. Il nous propose, dans le modèle du Records Continuum, une nouvelle présentation du cycle de vie des archives où l'espace-temps représente plutôt un ensemble de fonctions qui se réalisent tout au long du cycle de vie du document pour une finalité unique : la conservation des traces des activités de l'entité créatrice à des fins organisationnelles et sociétales. Les fonctions de création, d'enregistrement, d'organisation et d'exploitation forment les quatre dimensions concentriques sans véritable début ni fin du modèle du Records Continuum. Bien entendu, le processus de gestion documentaire s'enclenche au moment de la création d'un document et son enregistrement sur un support dans un système. À partir de ce moment, la gestion de l'enregistrement doit être systématique. Certains enregistrements évolueront et accéderont à d'autres statuts (document, document d'archives et archives), continueront à tisser des liens avec les autres documents créés par une unité pour former ainsi une composante de la mémoire organisationnelle et possiblement de la mémoire sociétale.

Au prisme du Records Continuum, la gestion du cycle de vie du document s'allonge sur l'axe de la mémoire. Le document évolue à travers un ensemble de statuts qui nous indique ce qu'il documente, son espace de conservation et les valeurs qu'il acquiert auprès de différents acteurs. Le modèle est plus approprié pour représenter le cycle de vie des documents numériques dans les organisations (Ducharme, 2008) car il ne se limite pas à la présentation des phases de vie d'un document uniquement en fonction de leur fréquence d'usage et de leurs espaces de conservation.

Cette vision du continuum dans les valeurs et les actions correspond aux prémisses de l'archivistique intégrée développée au Québec (Couture, 1999) et nous semble intéressante dans l'étude du cycle de vie des sites Web. Des choix documentaires s'imposent alors dès leur création afin de pouvoir maîtriser leur évolution :

« The need for managing records from the beginning of the records life cycle will become evident when managing Web site records, as metadata that is assigned to document quickly becomes invaluable for tracking, managing, storing, and preserving Web site records. » (Phillips, 2003: 47)

Le Secrétariat du Conseil du Trésor du Canada (2012) nous propose une nouvelle représentation du cycle de vie d'un site Web basée sur son évolution dans l'espace-temps :

- *Contenu Web actuel* : désigne l'information à jour accessible en ligne;
- *Contenu Web archivé* : désigne l'information qui n'est plus d'actualité, mais qui est toujours disponible sur le site Web aux fins de consultation ou pour situer le contenu actuel dans son contexte;
- *Contenu Web existant* : désigne l'information qui a été révisée ou remplacée par du contenu Web plus actuel et qui a été retirée du site pour être conservée dans un autre espace aux fins de la gestion des documents.

2.3.2.2 Approches d'archivage

Nous décrivons dans cette section les deux principales approches d'archivage des sites Web dans une perspective archivistique qui se sont dégagées de la recension des écrits.

2.3.2.2.1 Approche orientée objet

L'approche orientée objet consiste à produire périodiquement des portraits du site Web, à gérer de manière individuelle les documents qui le composent et à consigner les modifications

qui l'affectent dans un journal historique des modifications⁶. Elle est recommandée lorsqu'il s'agit d'un site Web statique constitué essentiellement de pages HTML ou de documents PDF (Brown, 2006; Brügger, 2005; Khouaja, et Couture, 2003-2004).

2.3.2.2.2 *Approche orientée événement*

Cette approche consiste à documenter les événements et les transactions ayant lieu sur le site Web. C'est une approche appropriée pour l'archivage des sites Web dynamiques ou transactionnels. Elle est opérationnalisée grâce à la capture des données enregistrées sur les profils d'utilisateurs, sur les transactions réalisées, sur les mécanismes de recherche utilisés ainsi que sur les scripts et les programmes sollicités dans un journal des transactions⁷ (Brown, 2006; Brügger, 2005; Khouaja, et Couture, 2003-2004).

2.3.2.3 **Principaux projets d'archivage du Web organisationnel dans une perspective archivistique**

Les travaux d'archivage du Web organisationnel sont peu documentés. Seulement quelques initiatives identifiées dans l'Annexe 1 (p. i) sont à visée archivistique (ex. : archives du Web du gouvernement du Canada, archives Web du gouvernement central du Royaume-Uni). Ce sont principalement des archives nationales qui procèdent à la collecte des sites Web des organismes gouvernementaux ou publics (BAC, 2007-2008; Cleroult, 2006; Racine, 2007) ou des institutions universitaires qui archivent leurs sites Web départementaux et quelques sites Web étudiants (ex. : The University of Melbourne, 2008a; University of Toronto Libraries, 2009).

Les travaux actuels sont encore à un stade expérimental. Ils sont focalisés sur la maîtrise des techniques d'archivage. Les critères de sélection ne sont pas clairement définis. L'accent est mis sur la valeur patrimoniale et informationnelle des sites Web. Or, il est évident que les sites Web organisationnels revêtent une valeur administrative et légale (Barry, 2004: 27-28; Cruse et

⁶ Le journal historique renseigne sur les éléments suivants : titre de l'élément affiché sur le site Web, numéro de version de l'élément affiché, le ou les auteur(s) du contenu, la liste des hyperliens auxquels il pointe, date du premier affichage, date de dernière modification, date de remplacement ou de retrait de l'élément affiché, son sort final après son remplacement ou retrait (NAA, 2001a: 28).

⁷ Le journal des transactions consigne, entre autres, les informations suivantes : date de l'événement ayant lieu sur le site Web, information sur l'utilisateur (adresse IP, navigateur, identification), les contenus affichés pour l'utilisateur y compris les scripts d'exécution (NAA, 2001a: 29-30).

al., 2003; Lefurgy, 2001; Pennock et al., 2006) en plus d'une valeur de témoignage et d'information (Cunningham et Phillips, 2005: 302).

Archives des sites des Premiers ministres français et CyberCemetery sont deux autres exemples d'archivage du Web gouvernemental dans une perspective archivistique. La première initiative documente les activités des gouvernements français depuis celui d'Alain Juppé jusqu'au gouvernement de Dominique Villepin (Premier ministre, 1996). La deuxième initiative, entreprise par University of North Texas Libraries (2009), archive les sites Web de nombreuses institutions publiques aux États-Unis, notamment les différentes versions du site de la Maison-Blanche (1994-2001).

Deux cas d'expériences organisationnelles sont documentés dans la littérature archivistique : le projet de la Goddard Library et l'archivage du Web à l'University of Melbourne. Le premier projet vise la collecte de la documentation technique produite sur le site intranet du Goddard Space Flight Center. La documentation collectée (captures du site et de documents individuels) présente pour les créateurs et les archivistes de l'organisme une valeur scientifique et une valeur documentaire (Senserini et al., 2004). L'archivage a été réalisé à distance grâce à un robot de collecte bien que les fichiers constituant le site Web soient accessibles à la source. Le but était d'avoir un portrait représentatif des documents tels qu'ils sont affichés et partagés sur le site. Les archives constituées représentent une source d'information sur les activités du centre. Elle est également une source d'information utile pour des décisions futures (planification de travail, mise en œuvre d'un projet).

Le projet de l'University of Melbourne a été lancé en 2002. Il vise la collecte du Web de l'université comme un patrimoine documentaire et culturel de l'organisation (Nicholls et Williams, 2002). La collecte porte sur les éléments statiques uniquement. Une liste préliminaire de critères permettant la sélection des contenus à archiver a été définie. Il s'agit plus précisément d'une liste de questions à adresser aux créateurs et aux gestionnaires des contenus. Les réponses conditionnent les choix de l'équipe d'archivage (contenus à archiver et techniques d'archivage). Cette équipe est constituée des archivistes et des webmestres de l'université.

2.4 Techniques d'archivage du Web

Le choix d'une technique d'archivage dépend de la nature des ressources en ligne à archiver (contenus statiques ou contenus dynamiques) et de l'approche choisie. Trois principales techniques documentées dans la littérature sur l'archivage du Web sont opérationnalisées par les bibliothèques nationales ou dans un contexte organisationnel.

2.4.1 Techniques recommandées

2.4.1.1 Archivage à la source

Un archivage à la source (*server side archiving* ou *direct transfer*) permet de récupérer manuellement les fichiers constitutifs d'un site Web à même le serveur qui héberge son contenu. C'est une technique appropriée pour la collecte verticale (en profondeur) d'une portion bien précise du Web dans une optique très sélective. Cette technique garantit la constitution d'archives de qualité notamment lorsqu'il s'agit de capturer des sites Web statiques (Brown, 2006: 46, 48). Onéreuse en temps et en ressources, cette technique s'avère pertinente pour des collectes ponctuelles et ciblées (Masanès, 2005, 2006). L'archivage à la source exige également une démarche de négociation avec le créateur ou l'hébergeur du contenu afin d'obtenir les autorisations de collecte et les clés d'accès aux fichiers originaux.

2.4.1.2 Archivage à distance

L'archivage à distance (*client side archiving* ou *remote harvesting*) vise la capture et le stockage de documents diffusés sur le Web à partir d'un poste client. Cette technique est utile notamment pour un moissonnage automatique de portions importantes du Web. On utilise principalement un robot de collecte ou aspirateur (logiciel d'indexation Web, ou *crawler*, transformé en un robot de collecte de fichiers Web, ou *harvester*, en respectant des paramètres spécifiques prédéfinis). Dans une entrevue accordée dans le cadre des inter-ateliers organisés par l'INA, Thomas Drugeon explique les principales étapes de la chaîne technique de l'archivage à distance (INA, 2003: 4-10).

Dans une première étape, l'équipe responsable de l'archivage fixe l'étendue de l'archivage et identifie le périmètre à archiver. Cette équipe exerce aussi une veille

documentaire afin de faire le suivi des sites à aspirer et découvrir de nouveaux sites. Les sites à archiver sont ensuite enregistrés dans un plan d'aspirations qui définit aussi l'enchaînement ou l'ordonnement des aspirations des sites (Kavcic-Colic et Grobelnik, 2004: 5).

Dans une deuxième étape, un robot d'aspiration est lancé sur une machine consacrée à l'archivage du Web. Il s'identifie auprès de l'ordonnanceur des aspirations. Il lui fournit des informations telles que son adresse, ses capacités et spécialisations (robot spécialisé dans la capture des statistiques, des thèses, etc.). L'ordonnanceur envoie au robot un ordre d'aspiration avec des paramètres assez précis sur un site à archiver (niveaux de profondeur, nombre de pages et paramètres divers). S'il dispose des ressources suffisantes pour appréhender le site à archiver, le robot d'aspiration accepte l'ordre et met en place sa propre stratégie d'aspiration (parcourir le site, appréhender les formats techniques, temporiser l'aspiration entre chaque page, etc.) (Chen et al., 2008: 4-6).

Par la suite, le robot va aspirer ou capturer le site à une fréquence donnée et à des moments particuliers. Il parcourt l'ensemble des liens du site et rapatrie son contenu et des informations utiles pour les prochaines aspirations (fréquence de mise à jour du site, frontière du site, etc.). Il analyse en parallèle le site et communique ensuite le résultat de ces travaux à l'ordonnanceur d'aspirations afin que ce dernier puisse mettre à jour le calendrier des aspirations. En cours d'opérations, l'ordonnanceur peut aussi interroger le robot sur son état, sur des statistiques techniques ou d'autres paramètres. L'ordonnanceur autorise par la suite le versement des pages Web aspirées dans un entrepôt de sites Web archivés. Les contenus sont conservés dans leur format d'origine, avec ou sans relativisation de liens (transformations des URL en PURL) ou convertis vers des formats normés (Cruse et al., 2003: 45-46). Une signature est assignée aux contenus archivés qui va permettre, lors d'une prochaine aspiration, de vérifier si une page est déjà, ou pas, aspirée et indexée (Kavcic-Colic et Grobelnik, 2004: 3).

Lors de la définition de la stratégie d'archivage (critères de sélection, fréquences des captures, niveau de profondeur), l'équipe responsable de l'archivage peut choisir entre deux solutions pour garantir un accès pérenne à l'archive :

1. Mise en place d'une couche d'émulation : il s'agit de transformer, au moment de la consultation, les formats conservés dans un format normé qui correspond à l'époque de leur création.
2. Migrations successives : il s'agit de convertir, au moment de la capture, tous les formats archivés dans un format normalisé et reconvertir après une période de temps l'ensemble des archives conservées.

Largement adoptée, cette technique d'archivage à distance est tout à fait appropriée pour des projets adoptant une approche intégrale ou exhaustive. Elle est aussi utile pour réaliser un instantané d'un site ou d'un domaine Web à des moments critiques de son cycle de vie (refonte majeure, mise à jour, migration vers d'autres interfaces, etc.). Elle peut également être déployée pour des collectes bien ciblées (Andrieu, 2005). Cette technique ne requiert pas d'importants investissements, notamment au niveau des compétences du personnel. Elle pourrait par contre être consommatrice de temps et d'espaces de stockage (Brown, 2006: 50-58). Le matériel récupéré pourrait souffrir de contenus incomplets, redondants ou illisibles (Masanès, 2005, 2006). Cette méthode permet toutefois de rendre compte de l'aspect visuel et structurel d'un site Web dans le temps (Marill et al., 2004; Masanès, 2005).

2.4.1.3 Archivage des transactions

L'archivage des transactions est une technique peu documentée. Elle permet de retracer et d'enregistrer les différentes actions qu'un usager entreprend lors d'une session de navigation sur un site Web. Elle permet aussi de capturer les contenus générés à la volée suite à la consultation d'une base de données en ligne ou la saisie d'un formulaire en ligne (Chebbi, 2007-2008). Cette technique est fondée sur une approche d'archivage orientée événement (Brown, 2006: 61). C'est une technique très appropriée pour la capture des transactions en ligne. Toutefois, elle pourrait présenter des contraintes légales et éthiques si elle est réalisée à l'insu de l'internaute et du fournisseur de contenu (Brown, 2006: 61-64).

2.4.2 Outils d'archivage des sites Web

L'archivage d'un document Web requiert la mise à contribution de plusieurs solutions logicielles que nous présentons dans la section suivante. Plusieurs logiciels libres et gratuits

sont proposés par l'IIPC. La description de ces logiciels de même que leur documentation technique est disponible sur le site Web de l'IIPC (2012b).

2.4.2.1 Logiciels de moissonnage ou d'acquisition (aspirateurs Web)

Il s'agit de robots de moissonnage de masse qui fonctionnent comme un robot d'indexation Web. Ils permettent la collecte d'un ensemble de sites Web selon des paramètres particuliers (ex. : niveau de profondeur, inclusion ou exclusion de certains types de liens, autorisation de capture). Il suffit de leur indiquer une adresse URL de départ et de programmer les paramètres de collecte. Heritrix, développé par Internet Archive en coopération avec la Bibliothèque nationale de Finlande est souvent utilisé par les bibliothèques nationales lorsqu'elles ne disposent pas d'un logiciel maison pour réaliser leurs collectes (Brown, 2006: 197-201; Cruse et al., 2003: 47; IIPC, 2004-2008; Marill et al., 2004). Des logiciels tels que Snag-it ou l'extension Scrapbook de Firefox permettent de récupérer pour des besoins strictement individuels une page ou un site Web et de les rendre accessibles hors ligne. Les logiciels PageVault et WebCapture permettent la réalisation de ce genre de captures (Fitch, 2003).

2.4.2.2 Logiciels de stockage et de maintenance

Il s'agit d'une catégorie de logiciels facilitant le contrôle et la gestion du calendrier des collectes et le contrôle de la qualité de l'archive constituée. Web Curator Tools et NetarchiveSuite en sont des exemples (Brown, 2006; Cruse et al., 2003: 84; IIPC, 2004-2008).

2.4.2.3 Logiciels d'indexation et de recherche

Le matériel collecté doit être catalogué et indexé à des fins de recherche et d'accès. NutchWax (Nutch with Web Archive eXtensions), Wayback ou WERA (WEb aRchive Access) sont des logiciels qui permettent l'indexation des contenus collectés de même que la recherche dans un entrepôt d'archives Web. Ces logiciels autorisent aussi la navigation entre les différentes versions d'une page/site Web (Brown, 2006; Cruse et al., 2003: 49; IIPC, 2004-2008).

2.5 Format d'archivage

En mai 2009, le format de fichier pour l'archivage de l'Internet (WARC) a été approuvé comme norme internationale : la norme ISO 28500 : 2009. Le format WARC est une extension du format de fichier ARC, développé et utilisé à l'origine par Internet Archive.org, et ce, depuis 1996. Cette norme définit un format de fichier conteneur qui regroupe les divers objets documentaires capturés dans un seul fichier long accompagnés de leurs métadonnées (Groupe de coordination 46, 2009).

Le format WARC autorise également l'enregistrement des entêtes de requête HTTP, des métadonnées, l'attribution d'un identifiant pour chaque fichier capturé ainsi que la gestion des doublons. En offrant cette option de personnalisation, ce format facilite donc la gestion des nombreuses versions archivées d'un même site Web tout en le distinguant de sa version en ligne. Ce format est adopté par plusieurs institutions de mémoire (ex. : BnF).

Certaines organisations ont adopté le modèle OAIS (Open Archival Information System) pour intégrer les archives du Web à leurs collections numériques existantes (ex. : bibliothèque nationale suisse, bibliothèque royale des Pays-Bas). L'OAIS, une norme ISO (ISO 14721), constitue un modèle conceptuel qui encadre la gestion, l'archivage et la préservation à long terme de documents numériques. Il ne suggère pas la conception d'un système d'archivage particulier, mais spécifie de manière très générale l'architecture logique et les fonctionnalités d'un système d'archivage (modèle fonctionnel). Il focalise sur la nature des données que l'on souhaite conserver (modèle d'information) et décrit également les responsabilités liées à la mise en place de ce système d'archivage (Bermès, 2009: 213-214). Selon ce cadre de référence, un système d'archivage est constitué de producteurs, d'utilisateurs et de décideurs qui s'échangent de l'information. L'information ou l'objet à archiver est présentée sous forme de paquet d'information. Ce dernier contient l'objet que l'on souhaite archiver, sa représentation et les métadonnées nécessaires à sa préservation.

2.6 Accès et exploitation des archives Web

Les initiatives actuelles d'archivage du Web se centrent sur les problématiques de capture et de stockage des contenus en ligne. Elles se heurtent également à des contraintes

juridiques et éthiques quant à la diffusion en ligne des sites archivés. À titre d'exemple, la législation britannique et australienne oblige à obtenir l'accord des ayants droit d'un site avant sa capture alors que la législation française ne demande pas l'accord, mais limite l'accès aux documents archivés. Ces archives restent ainsi dans l'ombre : « *dark archives* » (Nelson et al., 2007). Les contenus archivés sans l'accord du créateur sont accessibles selon des conditions particulières. À ce titre, les archives du Web français sont consultables uniquement dans les locaux de la BnF par des chercheurs accrédités (BnF, 2012a).

Les contenus collectés sont indexés de manière automatique par URL, par date de capture ou en texte intégral. La recherche en texte intégral génère beaucoup de bruit ou un silence (ex. : NLA, 2007). Dans la plupart des interfaces de recherche des archives Web, pour accéder aux archives en ligne d'un site Web, il faut connaître son URL (ex. : Internet Archive, 2012b). L'internaute pourrait limiter sa recherche à un intervalle précis de dates de collecte (seule option avancée offerte à date). Pour palier aux limites actuelles au niveau de la recherche dans les collections Web constituées, certains organismes ont opté pour les solutions suivantes :

- Offrir une liste des noms des organismes dont le site Web a été archivé (ex. : BAC, 2007-2008);
- Proposer un accès selon le type de contenus archivés (fichiers vidéos, fichiers sonores, livres et documents historiques numérisés, musique et concerts) (Internet archive, 2012b);
- Suggérer des collections thématiques de leurs archives. À titre d'exemple, la BnF présente des « parcours guidés » de sa collection des archives Web qui permettent de découvrir des sélections de sites archivés sur des thèmes donnés ou liés à des événements marquants. Le parcours intitulé *Cliquer, voter : l'Internet électoral* propose une sélection illustrée et commentée des sites des acteurs, observateurs et témoins des campagnes électorales (élections présidentielles et législatives de 2002 et 2007, élections régionales et européennes de 2004) (BnF, 2012b). De la même manière, les archives nationales de Grande-Bretagne proposent également une sélection de sites Web qui documente la réaction du gouvernement britannique face à des événements majeurs (TNA, [s.d.]c).

2.7 Archivage du Web social

Le Web 2.0 est né d'une évolution dans la manière dont les internautes font usage du Web. En effet, les technologies du Web 2.0 offrent cette possibilité de trouver, organiser, partager ou créer de l'information d'une manière à la fois personnelle, mais globalement

accessible (Stenger et Coutant, 2010). Le Web 2.0 représente également une nouvelle forme de réseaux d'interaction sociale. Sous cette appellation, nous pouvons regrouper, entre autres, les cybercarnets ou blogues, les sites de partage de contenus multimédias (ex. : YouTube, Flickr), les sites de réseautage (ex. : Facebook, LinkedIn) et les sites de microblogage (ex. : Twitter).

Sur le plan de la mémorisation, le Web 2.0 apporte son lot de défis : accroissement du volume des contenus partagés, mise à jour et agrégation continues des informations, diversité des formats et des plateformes de diffusion, etc. Les applications du Web 2.0 permettent de gérer la temporalité des contenus (ex. : conservation de l'historique des changements et des mises à jour des pages d'un Wiki, archives en ligne d'un blogue, téléchargement en local d'un profil Facebook), mais elles offrent rarement des fonctions de conservation à long terme (Bermès et Oury, 2009). Bien de questionnements se posent face à la mémorisation de ces contenus. Quels sont les critères de sélection qu'il faut prendre en considération pour identifier des matériaux les plus représentatifs de notre histoire contemporaine écrite et exprimée en partie sur le Web social? Quelles méthodes faut-il utiliser pour capturer ces flux d'information? Comment peut-on traiter et organiser un volume vertigineux d'information?

Dans un contexte organisationnel, l'archivage du Web social est bien plus problématique que l'archivage du Web public du fait que les contenus sont souvent hébergés à l'externe. L'organisme perd ainsi tout contrôle sur les contenus diffusés sur les différentes applications du Web 2.0. Les organisations, en devenant plus friandes de ces applications (ex. : usage du Web 2.0 dans les pratiques d'embauche, pour la surveillance des activités des employés, à des fins de marketing ou pour l'interaction avec les clients), s'exposent à de nombreux risques en diffusant leurs contenus sur le Web 2.0 tels que le risque de divulgation des renseignements confidentiels, risque de publication de contenus contraires à la loi ou les atteintes au droit d'auteur (CEFRIO, 2012). L'information diffusée sur les plateformes du Web 2.0 est liée aux processus d'affaires ou activités en ligne de l'organisation. Les archivistes n'ont pas à ignorer la valeur archivistique possible que pourraient présenter ces contenus (ARMA International). Une action proactive s'impose; ces contenus doivent être intégrés dès leur création dans un système d'archivage.

« It is not appropriate for a records professional to ignore the use of web 2.0 applications within their organisation. Records professionals need to integrate web 2.0 applications into their records management programs rather than

alienating it. A proactive approach will help to reduce the amount of information that may otherwise be lost when using web 2.0 tools. » (NSW, 2009b: 7)

Nous décrivons, dans ce qui suit, les méthodes et techniques possibles pour l'archivage de quelques exemples d'applications du Web 2.0 les plus connues et utilisées, soit les blogues, les gazouillis et les pages Facebook.

2.7.1 L'archivage du Web social : les principales initiatives

Les projets d'archivage des blogues ne sont pas aussi nombreux et bien documentés que les projets d'archivage du Web public. La recension des écrits permet le repérage de quelques initiatives :

- *Library of Congress*⁸ : cette institution archive depuis 2007 une collection de 133 blogues professionnels concernant le domaine légal et juridique (*Legal Blawgs Web Archive*) (Library of congress, 2011).
- *Bibliothèque nationale de France (BnF)*⁹ : en collaboration avec l'Association pour l'Autobiographie et le Patrimoine Autobiographique, la BnF collecte deux fois par an, depuis 2007, une liste de journaux personnels et littéraires. La collection compte quelque 400 sites et vise à être représentative de l'histoire du « diarisme en ligne ». La BnF propose, pour explorer cette collection, un parcours guidé intitulé (*S'écrire en ligne*) (disponible sur les postes de la salle de lecture) (BnF, 2009).
- *National Library of Australia et partenaires*¹⁰ : PANDORA, les archives Web australiennes, comprennent une collection très sélective de 461 blogues académiques. La recherche dans PANDORA peut être limitée aux blogues (NLA et partners, [s.d.]).
- *Internet Archive* : grâce à une collaboration avec la Library of Congress et la BnF, Internet Archive offre depuis le 19 janvier 2011 un accès en ligne à la collection "*Jasmine Revolution*"¹¹, une collection d'archives Web (235 ressources) constituée principalement de blogues (95 blogues) qui documentent le soulèvement populaire de 2011 en Tunisie (Internet Archive et archive-It, 2011-2012).

Le moissonnage à distance, la méthode la plus utilisée pour archiver des blogues, a ses limites. Nous avons constaté, en naviguant dans les archives des blogues de la Library of Congress ou celles de la National Library of Australia, que certains fragments des blogues n'ont pas pu être capturés. De même, la visualisation de quelques éléments de la version archivée

⁸ Adresse URL: <http://lcweb2.loc.gov/diglib/lcwa/html/lawlb/lawlb-overview.html>

⁹ Adresse URL: <http://blog.bnf.fr/lecteurs/index.php/2009/04/06/du-journal-intime-au-blog/>

¹⁰ Adresse URL: <http://pandora.nla.gov.au/subject/29>

¹¹ Adresse URL: <http://www.archive-it.org/collections/2323>

d'un blogue requiert l'installation de plugiciels ou des versions particulières de navigateurs Web.

D'autres méthodes, à un stade encore expérimental, sont possibles pour l'archivage des blogues (Capra et al., 2008; Kelly et Marieke, 2010; University of London Computer Centre et University of Bath, 2008), soient :

1. Production d'une version statique du contenu d'un blogue;
2. Production d'une copie de sauvegarde d'un blogue grâce à une option d'archivage offerte par l'application Web ou le navigateur Web utilisé;
3. Migration du blogue vers une autre plateforme pour maintenir ou enrichir ses fonctionnalités.

Ces méthodes présentent des avantages et des inconvénients ce qui les rend complémentaires.

2.7.2 L'archivage de Twitter

Le 14 avril 2010, Twitter a fait un don de ses gazouillis (*tweets*) publics de 2006 à 2010 à la Library of Congress, environ 5 téraoctets de contenus (Raymond, 2010). Les données relatives à un compte Twitter, les gazouillis supprimés de même que les liens vers des photos et sites Web n'ont pas été transférés à la Library of Congress. De même, l'exploitation de ces archives serait limitée à des études à but non commercial (Campbell et Dulabahn, 2010). Ce corpus permettrait à la Library of Congress d'explorer de nouvelles approches de préservation des collections numériques. Les archives de Twitter sont d'une valeur culturelle, scientifique et historique inestimable, car, elles sont dans quelques occasions, une source primaire sur des événements de la vie quotidienne, mais aussi des événements marquants de notre histoire contemporaine (ex. : Tsunami au Japon, révolutions dans le monde arabe).

« For historians, Twitter provides direct witness accounts of events in real time. It also serves as a virtual timeline of communications about events, people and places. This provides an enormous amount of raw unmediated primary source material for historical research. » (Campbell et Dulabahn, 2010)

En plus de la modalité du don ou du dépôt volontaire par le fournisseur du service Web 2.0, plusieurs applications API (Application Programming Interface)¹² permettent aux abonnés de Twitter de télécharger les traces de leurs activités sur Twitter, telles que Tweetnest¹³ et Thinkupapp¹⁴ et l'application API de Twitter. Internet Archive, par le biais de son service Archive-It, offre également la possibilité d'archiver les gazouillis publics (Internet Archive, 2012a). Google propose, quant à lui, l'application *Google Replay*¹⁵ pour chercher des gazouillis sur un événement (Twitter blog, 2010).

Les archives de Twitter ne sont pas encore exploitées. Des recherches sont en cours afin de rendre accessibles ces contenus au monde de la recherche (Campbell et Dulabahn, 2010).

2.7.3 L'archivage de Facebook

Facebook est un réseau social qui a eu beaucoup de succès auprès des internautes. Une page Facebook d'un utilisateur nous permet d'avoir accès à ses informations personnelles (nom et prénom, date de naissance, adresse, numéro de téléphone, courrier électronique, occupation, passe-temps, etc.), à sa liste d'amis et groupes favoris, aux informations, vidéos, photos ou documents de tout genre qu'il partage ou qu'il aime (ou n'aime pas), et, enfin, à ses commentaires (ce qu'il pense des publications des autres usagers voire même ses propres publications). Facebook offre à ses usagers, entre autres, un service de messagerie électronique, un moteur de recherche (contacts, messages, etc.), un service de discussion instantanée et plusieurs autres applications. Comment peut-on capturer une myriade de contenus et d'applications, aussi variées que différentes? Car, en plus de ses caractéristiques techniques uniques, Facebook est un espace où naissent des phénomènes sociaux singuliers (ex. : *e-revolution*).

McCown et Nelson (2009) mentionnent quelques solutions telles que la capture manuelle (téléchargement en local d'une photo) ou automatique avec, par exemple, les

¹² Une application API est une interface contenant les fonctions nécessaires au développement d'applications. Elle permet l'interaction des programmes les uns avec les autres, de manière analogue à une interface homme-machine et rend possible, en conséquence, l'interaction entre un homme et une machine (Wikipédia, 2012).

¹³ Adresse URL de l'application : <http://pongsocket.com/tweetnest/>

¹⁴ Adresse URL de l'application : <http://thinkupapp.com/>

¹⁵ Adresse URL de l'application : <http://googleblog.blogspot.com/2010/04/replay-it-google-search-across-twitter.html#/2010/04/replay-it-google-search-across-twitter.html>

extensions *Scrapbook* ou *ArchiveFacebook* de FireFox. Les notifications Facebook peuvent être acheminées par courrier électronique préservant ainsi les traces individuelles des changements qui affectent une page Facebook. Une autre option offerte par Facebook consiste à sauvegarder en local une copie de son profil Facebook sous forme d'un dossier en format compressé envoyé par messagerie électronique. Enfin, McCown et Nelson (2009) suggérèrent les applications API développées par Facebook pour extraire des données particulières à partir d'une page Facebook (ex. : `Friends.get` pour avoir la liste d'amis ou `photos.get` pour télécharger les photos partagées sur une page Facebook). Les deux chercheurs ont développé eux-mêmes un prototype d'application *Facebook Archiver* qui utilise une version modifiée de *Scrapbook* capable d'exécuter des requêtes AJAX. Archive-It, service d'archivage d'Internet Archive, pourrait également archiver des pages Facebook grâce à une syntaxe particulière¹⁶. Il est à noter que depuis 2010, Facebook bloque son site en entier aux robots de collecte.

Les exemples de méthodes que nous avons présentées pour archiver le Web social produisent, dans la plupart des cas, des archives incomplètes, dépourvues de leur aspect interactif, complètement décontextualisées ou qui ne reflètent pas l'expérience vécue par l'internaute. Ces solutions se heurtent également à la nature fermée de Facebook. La capture des profils d'utilisateurs à l'insu des détenteurs de comptes Facebook constitue une transgression aux règles de confidentialité et un non-respect des données privées des utilisateurs de Facebook.

En résumé, l'archivage du Web 2.0 présente de nouvelles difficultés d'ordre technique et remonte à la surface les enjeux de l'archivage du Web liés à la protection de la vie privée des utilisateurs du Web.

2.8 Cadre conceptuel pour étudier l'archivage du Web organisationnel

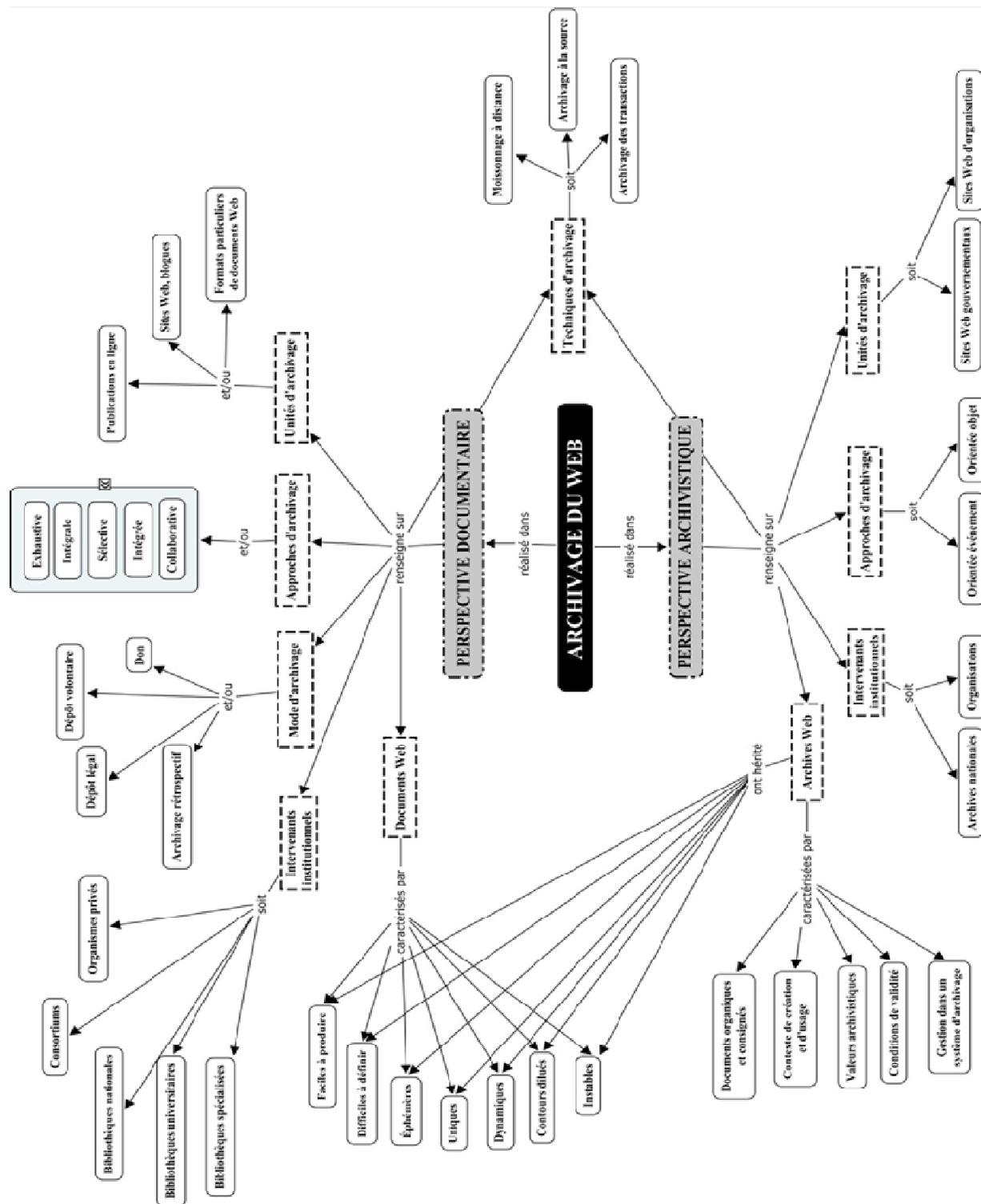
L'analyse des écrits nous a amenée à : (1) décrire les caractéristiques de la production documentaire sur le Web, (2) rappeler les principales caractéristiques distinctives des archives qui permettent de les identifier dans un environnement Web et (3) identifier les principales

¹⁶ La syntaxe d'archivage des pages Facebook d'Archive-it est disponible à l'adresse suivante : <https://webarchive.jira.com/wiki/display/ARIH/Archiving+Social+Networking+Sites+with+Archive-It> (Consulté le 14 avril 2012).

approches et techniques d'archivage du Web, selon une perspective documentaire et une perspective archivistique. Les résultats de cette recension révèlent que les pratiques d'archivage du Web dans un environnement organisationnel sont peu documentées dans la littérature archivistique. Ce constat nous confirme l'intérêt d'aller sur le terrain afin de se renseigner sur les composantes des SIW et des documents qui sont archivés, sur les intervenants responsables de l'archivage du Web organisationnel et sur les stratégies d'archivage adoptées. Ceci nous offre la possibilité de vérifier, d'une part, l'adéquation des stratégies d'archivage mises en œuvre aux caractéristiques des archives Web et, d'autre part, comparer les pratiques recommandées dans les politiques d'archivage que nous souhaitons analyser à celles développées dans les sites d'étude.

La cartographie des stratégies actuelles pour l'archivage du Web nous aide donc à forger un cadre de référence dont la structure est basée sur nos questions de recherche et dont les éléments indiqués nous guident dans notre analyse des données recueillies sur les pratiques d'archivage du Web dans des milieux organisationnels (Figure 1, p. 77).

Figure 1 : Cadre de référence de la recherche.



Une stratégie d'archivage doit définir :

- Les contenus devant être archivés : il s'agit de spécifier le niveau de granularité à partir duquel il faut procéder à l'archivage du site Web (ex. : domaine et sous-domaine du site Web), les types et les caractéristiques des documents à archiver.
- Le mode d'archivage à adopter : il s'agit d'indiquer la façon dont on va procéder pour capturer les documents Web (ex. : dépôt légal, don).
- Les approches d'archivage à mettre en œuvre : il s'agit de préciser la manière dont le processus d'archivage sera opérationnalisé (ex. : approche intégrale, approche sélective, approche collaborative).
- Les techniques d'archivage à utiliser : il s'agit de définir le procédé technique qui permet de capturer/acquérir des documents Web afin de les intégrer dans une collection ou un fonds, les stocker, les traiter et les exploiter par la suite (ex. : archivage à la source, moissonnage à distance).
- Les intervenants : il s'agit d'indiquer les entités responsables de l'archivage du Web (ex. : bibliothèques nationales, archives nationales, services d'archives).

Notre cadre de référence permet, dans un premier temps, d'étudier les stratégies d'archivage recommandées dans les politiques d'archivage du Web organisationnel et celles mises en œuvre dans les milieux d'étude. Cette analyse nous aide à relever, par la suite, les similitudes ou les particularités des documents d'archives produits dans un environnement Web par rapport à ceux considérés comme des publications. Nous partons de l'hypothèse qu'un document Web, qu'il soit une publication ou un document d'archives, présente les mêmes caractéristiques techniques. Toutefois, son mode d'archivage pourrait différer selon son contexte de production et les exigences relatives à sa gestion (ex. : un document d'archives, produit de la conduite d'une activité, doit satisfaire certaines conditions telles que l'authenticité, l'intégrité, la fiabilité et l'exploitabilité). De nouvelles façons de faire donc peuvent émerger suite à l'analyse des données réalisées dans le cadre de cette étude empirique.

Conclusion

La conservation du Web est une entreprise complexe sur tous les plans. La diversité des formats, la connectivité des objets et l'interactivité des contenus rendent difficile toute tentative de fixer une masse imposante, mouvante, instable et protéiforme dans le temps et dans l'espace. Les méthodes en cours se heurtent aux limites imposées par le Web. Les projets d'archivage actuels se basent essentiellement sur des captures automatisées d'un Web de surface du côté

client. Les approches et techniques en usage ne permettent pas la constitution d'une mémoire fidèle et exacte du Web. Dès lors, des choix s'imposent et des concessions sont à accepter (Gladney, 2007: 3).

Les institutions de mémoire développent des partenariats avec d'autres organismes afin de surmonter les défis techniques, technologiques et méthodologiques liés à l'archivage du Web. La mise en œuvre d'un projet d'archivage requiert d'importantes ressources organisationnelles, la mise en commun d'expertises complémentaires et un investissement important en temps et en coût (Chapman, 2006; Lavoie, 2006: 111-112, 122).

L'archivage du Web dans une perspective archivistique reste peu documenté. Les écrits actuels ne renseignent pas suffisamment sur les pratiques d'archivage dans un contexte organisationnel et sur l'articulation des rôles et responsabilités entre les organisations et les archives nationales dans la constitution d'une mémoire du Web organisationnel.

Dans cette section, nous avons présenté les particularités des ressources Web et nous avons rappelé les caractéristiques des archives Web. Nous avons ensuite décrit les stratégies d'archivage visant à les contrôler. Cette recension des écrits nous a permis de dresser le cadre conceptuel sur lequel nous nous basons pour étudier l'archivage du Web dans un contexte organisationnel.

Le chapitre qui suit décrit la méthodologie utilisée pour cette recherche ainsi que les mesures mises en place pour en garantir la validité et la fiabilité.

3. Méthodologie

Introduction

Notre recherche vise à explorer et à décrire des pratiques de gestion documentaire d'un nouvel objet archivistique, à savoir : les sites Web organisationnels. Plus précisément, le projet de recherche a pour but de documenter la démarche actuelle d'archivage des sites Web dans un contexte organisationnel du point de vue des acteurs impliqués dans ce processus et de la comparer aux recommandations qui émanent de politiques référence en la matière. Pour y parvenir, nous tentons de répondre aux questions suivantes :

1. Que recommandent en général les politiques d'archivage du Web organisationnel?
2. Quelles sont les principales caractéristiques des archives Web?
3. Quelles pratiques d'archivage du Web organisationnel sont mises en place dans des organisations au Québec?
 - a. Quels sont les principaux intervenants impliqués dans le processus d'archivage du Web organisationnel?
 - b. Quelles sont les approches et les techniques d'archivage du Web employées dans un contexte organisationnel?
 - c. Quels facteurs influencent la mise en œuvre d'une politique d'archivage du Web organisationnel?

Afin de répondre à ces questions, nous avons mené une étude de nature qualitative. Ce chapitre présente, tout d'abord, l'approche générale de recherche, ensuite les modes de collecte des données, ainsi que les paramètres d'échantillonnage, et enfin, la méthode d'analyse des données. Nous y décrivons également les moyens qui ont été mis en œuvre pour assurer la qualité et la validité de notre recherche.

3.1 Approche générale

Dans le cadre de cette étude, nous souhaitons, tout en décrivant la démarche d'archivage du Web organisationnel, comprendre la nature des documents archivés et déterminer les

facteurs qui dictent le choix d'une méthode d'archivage. Notre recherche porte sur un phénomène émergent, peu documenté et rarement étudié dans une perspective archivistique (Chebbi, 2007-2008; Zwarich et Parck, 2008). Le choix d'une approche qualitative basée sur des sites d'étude multiples nous permet de produire non seulement des descriptions riches, mais aussi des représentations réelles du phénomène à l'étude (Miles et Huberman, 2003: 27). Nous espérons décrire le processus d'archivage d'un site Web public d'une organisation (le comment) et expliquer les motifs d'une telle action (le pourquoi). Les résultats de l'analyse nous permettront, entre autres, de dégager les particularités et les exigences de l'archivage du Web dans une perspective archivistique, de vérifier si les pratiques sont cohérentes au niveau des différents sites à l'étude (Rispal Hlady, 2002: 64), d'en donner les explications possibles et d'établir un portrait global du phénomène à l'étude (Eisenhardt, 1989; Gagnon, 2005; Miles et Huberman, 2003; Patton, 2002; Yin, 2003).

Nous adoptons également une démarche d'analyse comparative qui nous permet également de confronter les résultats obtenus sur le terrain aux recommandations préconisées dans les politiques d'archivage du Web organisationnel; l'objectif étant d'établir un modèle type d'une politique d'archivage pour le site Web public d'une organisation.

Ce choix méthodologique nous semble donc adéquat pour la conduite d'un projet à visée descriptive et exploratoire comme le nôtre (Patton, 2002: 447; Rispal Hlady, 2002: 39; Yin, 2003: 14).

3.2 Collecte des données

Les trois modes de collecte de données opérationnalisés dans le cadre de cette recherche sont présentés dans la section suivante.

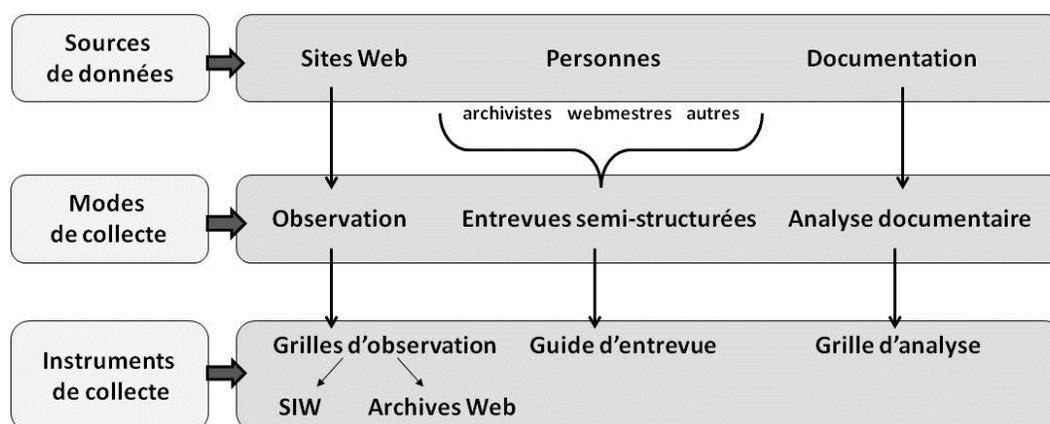
3.2.1 Modes de collecte des données

Différentes techniques de collecte des données sont employées dans le cadre de cette recherche, à savoir : l'observation des caractéristiques techniques des sites Web publics de 11 organisations, l'observation des caractéristiques des archives Web diffusées sur ces sites (n=737), la réalisation d'entrevues semi-structurées avec les acteurs impliqués dans le processus d'archivage dans les milieux choisis (n=21) et l'analyse d'une sélection de politiques

d'archivage du Web organisationnel (n=33) repérées en ligne accompagnées de 22 documents complémentaires ainsi que de la documentation pertinente recueillie lors des entrevues (voir Figure 2, p. 83).

Quatre instruments de collecte ont été développés pour les besoins de cette recherche soit : deux grilles d'observation des caractéristiques des sites Web publics et des documents qui y sont diffusés, un guide d'entrevue semi-structurée et une grille d'analyse des politiques d'archivage et de la documentation recueillies lors des entrevues.

Figure 2 : Sources, modes et instruments de collecte des données.



Nous discutons, ci-après, de ces modes de collecte. Nous exposons, ensuite, le résultat du prétest conduit pour ajuster et valider notre démarche et le contexte dans lequel a été réalisée la collecte des données. Nous établissons, par la suite, le lien entre les questions de recherche, les modes de collecte des données et les instruments de collecte. Nous présentons, enfin, le mode d'échantillonnage privilégié ainsi que les caractéristiques des sites d'étude et des participants à notre recherche.

3.2.1.1 Analyse des politiques d'archivage du Web organisationnel et de la documentation pertinente

3.2.1.1.1 Analyse des politiques d'archivage du Web organisationnel

L'analyse de documentation présente de nombreux avantages (Yin, 2003: 85). Il s'agit d'une source d'information qui offre une image fidèle d'une situation et qui permet

l'explication de certains phénomènes peu investigués. La documentation sert surtout à augmenter la qualité des données recueillies grâce à d'autres méthodes et à étayer plus solidement une proposition, voire la confronter avec d'autres (Tellis, 1997; Yin, 2003).

Dans notre cas, la littérature scientifique et professionnelle en archivistique ne permet pas d'apprendre davantage sur les pratiques organisationnelles en matière d'archivage des SIW, signe de l'état peu avancé des travaux dans la discipline. De même, nous n'avons pas pu recueillir des politiques d'archivage dans tous les milieux d'étude. Nous avons, par contre, noté la diffusion d'un nombre intéressant de politiques d'archivage sur les sites Web provenant d'archives nationales ou de services d'archives, notamment de services d'archives universitaires ayant entrepris des travaux d'archivage de leur site Web institutionnel. Ces politiques d'archivage constituent un corpus documentaire qu'il est nécessaire d'étudier pour deux raisons principales (Voir Annexe 2, p. xvii). Tout d'abord, ces politiques sont notre principale source d'information sur les cadres de référence, la démarche et les techniques recommandées pour l'archivage du Web organisationnel, mais surtout sur la position officielle des institutions archivistiques nationales en matière de conservation des sites Web. Ensuite, les politiques d'archivage, notamment celles établies par des services d'archives universitaires, sont un témoignage unique sur les pratiques réelles d'archivage du Web sur le terrain. En effet, le milieu universitaire figure sur la liste des milieux choisis dans le cadre de notre recherche. L'analyse de la documentation existante, comme le souligne Cellard, nous permet de retracer l'évolution des connaissances et des pratiques sur le phénomène que nous étudions (Cellard, 1997: 254).

L'étude de cette documentation nous engage également dans une démarche comparative plus générale dans l'objectif de vérifier si les recommandations énoncées dans les politiques sont mises en œuvre sur le terrain et d'identifier les facteurs qui entravent leur application. Les résultats de l'analyse de cette documentation contribuent aussi au développement des éléments d'un modèle de politique type pour l'archivage d'un site Web public d'une organisation que nous souhaitons établir.

Une grille d'analyse, inspirée de l'outil validé par Dufour (2003) pour analyser les politiques pangouvernementales liées aux SIW (Canada), a été utilisée pour nous renseigner sur

l'approche archivistique en matière de gestion et conservation des sites Web organisationnels (voir Annexe 3, p. xxii).

La grille comprend deux principales sections. Nous identifions dans la première section les éléments descriptifs de la fiche d'analyse (titre et date de création) et du document analysé (titre, auteur et type). Nous y indiquons également les liens signalés dans la politique analysée avec d'autres documents et nous vérifions systématiquement si ces documents font partie de la liste des documents complémentaires que nous devons analyser. Si c'est le cas, nous intégrons l'analyse de ces documents sous la même fiche d'analyse dans le champ Commentaires.

Nous nous attardons dans la deuxième section à l'analyse de l'objet de la politique analysée (but et objectifs, portée et systèmes Web visés), à l'identification des intervenants clés ayant des rôles et des responsabilités dans l'archivage des sites Web organisationnels, aux unités d'archivage qui sont ciblées de même que les motifs qui soutiennent leur archivage et le statut qui leur a été attribué. Nous examinons également les principes directeurs sur lesquels reposent les recommandations formulées dans une politique, soit des principes généraux, soit des principes liés à la conception des sites Web ou des principes archivistiques. Nous portons une attention particulière au(x) mode(s), approche(s), technique(s), critère(s) de sélection et outil(s) d'archivage recommandé(s) ou mis en œuvre dans le texte de politique à analyser. Nous dégageons également les facteurs à prendre en compte lors de la définition ou la mise en œuvre du processus d'archivage et la stratégie définie pour la préservation à long terme des contenus archivés. Enfin, nous relevons les définitions des concepts à l'étude (ex. : archives Web, site Web). Une description détaillée de ces éléments est fournie dans l'Annexe 3 (p. xxii).

La grille d'analyse provenant d'une grille déjà validée, le prétest n'était pas nécessaire. La grille a été précisée suite à l'analyse des premières politiques.

3.2.1.1.2 *Documentation pertinente*

Lors des entrevues, nous avons pu avoir accès à des documents qui complètent les informations recueillies pendant l'entrevue et valident celles issues de l'observation des sites Web publics et des documents qui y sont publiés (ex. : règles de conservation, procédures techniques de copies de sauvegarde, données sur le volume des fichiers d'un site Web, rapport

d'expérimentation d'un logiciel d'archivage du Web). Ces documents ont été analysés avec la grille d'analyse des politiques d'archivage du Web organisationnel.

3.2.1.2 Observation des sites Web publics

L'observation est une source incontournable de données qui favorise la compilation des informations les plus complètes sur des milieux dont les analyses empiriques sont rares, très fragmentaires ou superficielles (Deslauriers et Mayer, 2000: 136-139; Jaccoud et Mayer, 1997: 212; Laperrière, 2003: 287). Ce mode d'investigation favorise non seulement une compréhension holistique et empiriquement fondée du phénomène à l'étude (Laperrière, 2003: 272), mais aussi la comparaison entre différentes situations observées dans l'objectif de mesurer l'écart entre le discours et la pratique, ou encore, de confirmer des concepts ainsi que des théories déjà formulées (Deslauriers et Mayer, 2000; Jaccoud et Mayer, 1997).

L'observation directe est utilisée dans le cadre de cette recherche afin de recueillir des informations sur les caractéristiques techniques et technologiques des sites Web publics et d'établir une cartographie documentaire de ces systèmes. Cette observation est préalable aux entrevues en personne avec les responsables de gestion de ces systèmes et les gestionnaires des documents. La chercheuse a pu ainsi poser des questions plus précises lors des entrevues et vérifier les conclusions de son observation avec les participants à l'étude.

Nous avons attribué des codes pour identifier les sites Web observés sous la forme suivante : SIW suivi d'un numéro séquentiel du site Web observé (ex. : SIW01).

L'observation des sites Web publics a deux finalités principales : acquérir une meilleure compréhension des contextes technologique et documentaire des archives Web et caractériser ces documents (catégories, familles, types). En effet, les politiques d'archivage du Web définissent des stratégies d'archivage en fonction des niveaux de risques de gestion associés aux sites Web. Le choix d'une stratégie d'archivage dépend, entre autres, de la nature des SIW (ex. : sites statiques vs sites dynamiques) et de l'originalité des documents qui y sont diffusés (document diffusé sur un autre support vs document disponible uniquement sur le site Web sans équivalent sur un autre support).

L'observation des caractéristiques technologiques des sites Web à l'étude et l'identification des stratégies d'archivage appliqués dans les milieux de recherche nous ont

permis d'établir des liens entre le choix d'une stratégie d'archivage et la nature du site Web (type, objectifs et fréquence de mise à jour du site Web). Ces caractéristiques technologiques ont été observées avec une grille d'analyse développée sur le modèle d'une grille élaborée par Dufour (2003) afin de dresser une typologie des SIW dans lesquels interviennent les professionnels de l'information (voir Annexe 4 : Grille d'observation des SIW, p. xxv). Cette grille comporte cinq sections. La première section fournit des informations générales sur le site Web observé telles que : URL du site, catégorie, date de création ou de mise à jour, responsable(s) du site, ses différents usages ou objectifs et la clientèle visée. La deuxième section vise à décrire la technologie de développement et de gestion du site Web observé alors que les troisième et quatrième sections s'attardent sur sa structure et la nature des contenus qu'il véhicule. L'aspect interactif du site Web est présenté dans la cinquième section. Chaque élément d'observation est décrit en détail dans l'annexe 4 (p. xxv).

Ensuite, nous avons observé les caractéristiques documentaires de ces sites Web. Nous avons pu vérifier le caractère original de certains documents parmi l'échantillon des documents étudiés dans le cadre de cette recherche, et ce, grâce à des indices visuels (ex. : forme du document) et en ayant recours à des outils archivistiques (ex. : plan de classification ou calendrier de conservation). Cette démarche conduit, dans une autre étape, à vérifier si les milieux d'étude identifient les mêmes unités d'archivage et y appliquent les mêmes recommandations préconisées dans les politiques d'archivage. Tout comme pour les archives numériques, le respect des règles et standards de conception de même qu'une description exhaustive et pertinente des documents Web facilitent leur traitement et préservation à long terme. L'observation des archives Web nous indiquera si ces deux exigences ont été considérées.

Pour ce faire, nous avons capturé, grâce à l'extension Scrapbook du navigateur Firefox, les sites Web publics des organismes à l'étude. Pour chacun des sites, nous avons capturé une moyenne de 67 documents à des niveaux de profondeur différents. Nous avons assigné un code unique à chaque document capturé qui identifie l'organisme créateur, le niveau de profondeur du document, son ordre séquentiel, et son objet (voir Annexe 5, p. xxxi). Nous avons également veillé à aspirer la carte (ou le plan) des sites Web observés ainsi que le plan de classification et/ou le calendrier de conservation lorsqu'ils sont diffusés sur les sites Web observés.

Finalement, nous avons aspiré la page qui présente l'historique du site Web observé lorsqu'elle existe.

La nature des pages Web observées a été étudiée grâce à une grille d'analyse (voir Annexe 6, p. xxxii). Les éléments observés ont été déterminés à partir de la revue de littérature sur la notion de genre (Beghtol, 2001; Shepherd et Watters, 1998, 1999; Tyrväinen et Päivärinta, 1999; Toms, 2001) et sur la typologie des documents d'archives (Gagnon, 1998; Gagnon, 2000; Gagnon, 2002; Zeller, 2004). Cette grille compte deux principales sections. La première section vise à identifier la page Web à analyser. Elle comporte les champs suivants : le titre de la page Web observée, sa date de création et/ou de mise à jour, son adresse URL, le nom de son créateur, la nature des liens présents sur la page Web et le format de cette dernière. La deuxième section, quant à elle, permet de préciser la nature de la page Web observée. Elle est composée des champs suivants : une description du contenu de la page Web observée, sa catégorie archivistique (document de fonction ou document de gestion), son type (rapport, formulaire, plan stratégique, etc.) et les outils archivistiques utilisés pour valider la catégorie et le type de la page observée (plan de classification, calendrier de conservation). Un champ « commentaires » a été également réservé pour toute autre information utile recueillie lors de l'analyse. Les éléments de la grille sont détaillés dans l'Annexe 6 (p. xxxii).

L'observation doit être utilisée en conjonction avec d'autres sources de collecte de données afin de limiter le risque de biais que peut introduire le chercheur (l'observateur) (Laperrière, 2003: 287). Pour éviter ce genre de biais, nous avons conduit des entrevues en profondeur avec des acteurs organisationnels impliqués dans le processus de gestion et d'archivage des sites Web observés.

3.2.1.3 Entrevues en profondeur

Selon Poupart, deux principaux arguments d'ordre méthodologique légitiment le recours à l'entrevue de recherche (Poupart, 1997: 181-184). Le premier argument porte sur la qualité et la richesse de l'information recueillie. L'interviewé est considéré comme un « informateur clé » susceptible de renseigner non seulement sur ses propres pratiques, mais aussi sur son contexte d'action. Le chercheur tente alors de comprendre ou de reconstituer la réalité telle que perçue par l'interviewé (Poupart, 1997: 181).

Le deuxième argument découle du premier et repose sur les nombreux avantages de l'entrevue de recherche, particulièrement l'entrevue semi-structurée, quand il s'agit de refléter le point de vue des informateurs (Poupart, 1997; Savoie-Zajc, 2003). En effet, l'informateur jouit d'un certain degré de liberté pour s'exprimer dans son propre langage sur l'objet de recherche ce qui diminue le risque de préstructuration de ses réponses en raison d'une forme prédéterminée des questions comme lors de l'usage d'un questionnaire (Poupart, 1997: 182). Par conséquent, le matériel collecté favorise l'émergence de nouvelles dimensions exprimées par les différents sujets interviewés, souvent non pressenties au départ par le chercheur (Poupart, 1997).

Les données collectées lors de l'entrevue de recherche ne reposent pas sur une seule perspective et, par conséquent, offrent la possibilité d'explorer en profondeur les différentes expériences vécues ainsi que le contexte dans lequel elles ont eu lieu (Bergeron, 1997; Poupart, 1997; Savoie-Zajc, 2003). Grâce à une utilisation efficace des techniques de reformulation, l'interviewé est incité à répondre en détail aux questions posées et à établir les liens qui lui semblent utiles. En même temps, le chercheur peut vérifier et valider avec le sujet interviewé sa compréhension de la réalité observée.

Dans le cadre de cette recherche, des entrevues en profondeur ont été menées auprès de deux groupes d'employés, soit les archivistes (n=13) et les webmestres (n=8). Au total, nous avons interviewé 21 personnes dans 11 milieux à l'étude. En premier, nous avons ciblé les archivistes, d'une part, parce que cette catégorie de professionnels était plus accessible à la chercheuse (contacts dans le milieu, réseau professionnel) et, d'autre part, parce que les archivistes sont responsables de la préservation de la mémoire organisationnelle. Un archiviste doit bien connaître l'actif informationnel de son organisation et veiller à sa bonne gestion et conservation. Pour ce faire, il doit également collaborer avec les créateurs de contenus et les responsables des systèmes d'information dans son organisation. Certains archivistes qui ont accepté de participer à notre recherche (n=4) ont invité les responsables de la gestion des sites Web. D'autres sont responsables à la fois de la gestion documentaire et de la gestion du site Web de leur service ou sont impliqués dans le processus de sa gestion et de son archivage (n=3). Un archiviste participant nous a aidée à établir un contact avec le webmestre de son organisation.

Pour s'assurer de la couverture de l'ensemble des thèmes à l'étude lors de la réalisation des entrevues, un guide d'entrevue semi-structuré a été utilisé (voir Annexe 7, p. xxxiv). Le guide d'entrevue est un instrument de collecte des données assez flexible qui peut être ajusté selon les contextes (vocabulaire, ordre des questions, etc.) et qui permet la collecte d'informations riches reflétant le point de vue des différents participants sur le phénomène étudié (Patton, 2002).

Le guide d'entrevue comporte un ensemble de questions ouvertes qui couvrent les sujets suivants : caractéristiques du site Web public de l'organisation, son actif documentaire, la démarche de son archivage, les outils d'archivage utilisés ou développés, les difficultés rencontrées, les travaux futurs, l'archivage des autres SIW de l'organisation, des informations démographiques sur les participants à l'entrevue.

Les entrevues, d'environ une heure trente minutes avec chaque participant, ont toutes été réalisées par la chercheure et enregistrées avec le consentement des participants (voir Annexe 8, p. xxxviii) conformément aux exigences relatives aux aspects éthiques de la recherche à l'Université de Montréal (voir Annexe 9, p. xli). Des notes furent prises durant la réalisation des entrevues en utilisant une feuille de réponses (voir Annexe 10, p. xlii), et ce, afin, de consigner les impressions et réflexions sur le déroulement de l'entrevue, mais aussi pour faciliter ou compléter les transcriptions (au cas où l'enregistrement n'aurait pas une bonne qualité). Un protocole d'entrevue a été également utilisé afin de rappeler l'ensemble de tâches que la chercheure et les participants devront réaliser (voir Annexe 11, p. xlvi). Les entrevues ont été transcrites pour analyse assistée avec le logiciel d'analyse de contenu qualitative NVivo. Un code a été assigné à chaque participant de la manière suivante : R04.1 (R pour répondant, 04 pour le code de l'organisme et 1 pour le numéro du répondant lorsque la chercheure rencontre plus d'un participant dans un même milieu).

3.2.2 Prétest des instruments de collecte des données

Afin de vérifier la validité de nos instruments de collecte des données, nous avons réalisé en janvier 2010 un prétest qui a porté sur la grille d'observation des archives Web, la grille d'analyse des sites Web publics et le guide d'entrevue. La grille d'analyse des politiques ainsi que la grille d'analyse des sites Web n'ont pas été prétestées parce qu'elles découlent de

modèles de grilles déjà validées (Dufour, 2003), mais suite à une première vague d'analyse des politiques, nous avons ajouté le champ Support d'archivage à cette dernière.

Le prétest a été mené sur un site d'étude choisi par convenance. Nous avons, dans un premier temps, observé les caractéristiques du site Web public et des archives Web d'un département au sein d'une institution d'enseignement universitaire. Ensuite, nous avons réalisé une entrevue d'environ deux heures avec le webmestre de ce département. L'entrevue a été enregistrée avec prise de notes lors de l'entrevue. Aucune transcription n'a été faite. Les données recueillies ont été analysées sommairement afin de vérifier l'adéquation de la méthodologie choisie et des résultats obtenus avec les questions de recherche.

Suite à cette première collecte de données, des modifications ont été apportées aux instruments de recherche :

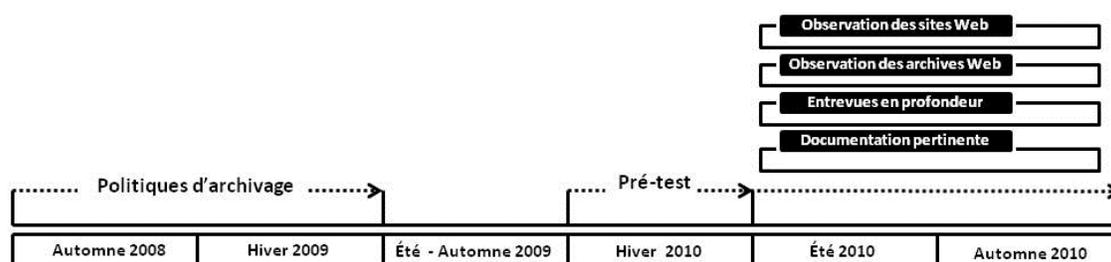
- Guide d'entrevue :
 - o Réorganisation des questions afin de faciliter le déroulement de l'entrevue (ex. : les questions sur l'archivage des autres systèmes Web dans l'organisation ont été regroupées à la fin du guide afin d'éviter une confusion possible entre les systèmes archivés).
 - o Précision de l'énoncé de certaines questions (ex. : des exemples ont été ajoutés afin de clarifier le sens de la question sur les technologies de développement du site Web).
- Grille d'observation des archives Web :
 - o Précision des critères d'évaluation (ex. : nombre de pages ou de documents Web à observer selon le volume du site).
 - o Suppression de trois éléments d'observation peu pertinents : identification, archivage et accès aux archives.
 - o Ajout de trois éléments : Famille du document, Catégorie du document et Contenu du document.
- Guide d'observation des sites Web :
 - o Une redondance entre certains éléments d'observation a été relevée et éliminée.
 - o Précision de la définition de quelques critères d'évaluation.

Les données obtenues de cette première collecte de données ont été intégrées aux données de recherche puisque les ajustements exigés étaient mineurs.

3.2.3 Chronologie de la collecte des données

La collecte de données s'est réalisée en deux étapes (voir Figure 3, p. 92). Lors de la première phase, nous avons collecté un corpus de politiques sur l'archivage des sites Web organisationnels. La deuxième phase s'est étalée de la session d'été 2010 à la session d'automne 2010. Dès l'obtention d'une réponse positive d'un site d'étude, la chercheuse a entamé l'observation des caractéristiques techniques et documentaires du site Web du milieu en question. Suite à l'entrevue, les données recueillies par observation ont été, dans certains cas, revues et la documentation recueillie du milieu a été analysée.

Figure 3 : Chronologie de la collecte des données.



3.2.4 Lien entre les questions de recherche, les modes de collecte et les instruments de collecte

Nous avons développé une liste d'indicateurs en relation avec chacune de nos questions de recherche. Cette liste nous sert à déterminer les éléments d'information à rechercher pour répondre à chaque question tout en précisant le mode de collecte à privilégier et les instruments de collecte à utiliser. Le tableau suivant (Tableau II, p. 93) présente les liens qui existent entre les questions de recherche, les indicateurs, les modes et les instruments de collecte des données. Nous nous sommes référée à ce tableau tout au long de la réalisation de la recherche afin de garder à l'esprit nos questions de recherche et, par conséquent, mieux orienter notre collecte ainsi que l'analyse de données et établir une triangulation des résultats obtenus.

Tableau II : Lien entre les questions de recherche, les modes et les instruments de collecte des données.

QUESTIONS DE RECHERCHE	INDICATEURS	MODES DE COLLECTE	INSTRUMENTS DE COLLECTE
Question 1	<ul style="list-style-type: none"> - Systèmes visés - Intervenants devant être impliqués dans le processus d'archivage, rôles et responsabilités - Motifs d'archivage - Unités d'archivage - Critères de sélection - Mode d'archivage - Approches d'archivage - Techniques d'archivage - Outils d'archivage - Principes directeurs - Stratégies de préservation à long terme 	Analyse des politiques	<ul style="list-style-type: none"> - Grille d'analyse des politiques d'archivage (II.4, II.6, II.7, II.8, II.9, II.11, II.12, II.13, II.14, II.15, II.16, II.17, II.19)
Question 2	<ul style="list-style-type: none"> - Nature du SIW 	Observation du site Web	<ul style="list-style-type: none"> - Grille d'observation du site Web (I.9, II.1, II.2, II.3, IV.4, IV.5, IV.6, V.1, V.2, V.3, V.4, V.5, V.6)
	<ul style="list-style-type: none"> - Éléments d'identification - Composantes (liens et objets documentaires composants la page observée) 	Observation des archives Web	<ul style="list-style-type: none"> - Grille d'observation des archives Web (I.1, I.2, I.3, I.5, I.6, I.8)
	<ul style="list-style-type: none"> - Format des documents - Catégorie du document - Type du document 	Observation des archives Web Entrevue en profondeur	<ul style="list-style-type: none"> - Grille d'observation des archives Web (I.7, I.8, I.9, I.10, I.11) - Guide d'entrevue (3.2, 3.b, 3.d, 3.e)
Question 3.a	<ul style="list-style-type: none"> - Intervenants impliqués dans le processus de gestion 	Observation des SIW	<ul style="list-style-type: none"> - Grille d'observation du site Web
		Entrevue en profondeur	<ul style="list-style-type: none"> - Guide d'entrevue (1.b,

QUESTIONS DE RECHERCHE	INDICATEURS	MODES DE COLLECTE	INSTRUMENTS DE COLLECTE
	d'archivage du site Web de l'organisation - Rôles et responsabilités		2.d)
Question 3.b	<ul style="list-style-type: none"> - Systèmes Web archivés - Motifs d'archivage - Unités d'archivage - Critères de sélection - Mode d'archivage - Approches d'archivage - Techniques d'archivage - Outils d'archivage - Principes directeurs - Stratégies de préservation à long terme 	Entrevue en profondeur	- Guide d'entrevue (4.a, 4.b, 4.c, 5.a, 8.c, 8.d)
		Analyse de la documentation	- Grille d'analyse des politiques (II.4, II.8, II.9, II.11, II.12, II.13, II.14, II.15, II.16, II.17, II.19)
Question 3.c	- Nature du site web	Observation du site Web	- Grille d'observation du site Web (I.9, II.1, II.2, II.3, IV.4, IV.5, IV.6, V.1, V.2, V.3, V.4, V.5, V.6)
	- Ressources disponibles	Entrevue en profondeur	- Guide d'entrevue (4.d, 6.a, 6.b)

3.3 Plan d'échantillonnage

Des critères de sélection ont été alors établis dans le but de choisir un terrain d'étude en adéquation avec nos objectifs de recherche et des répondants « bien informés » qui nous permettent d'apprendre en profondeur sur le phénomène que nous souhaitons étudier (Bergeron, 1997; Denzin et Lincoln, 2003; Miles et Huberman, 2003; Patton, 2002).

Nous présenterons dans la prochaine section le processus de sélection et les caractéristiques des politiques analysées, des sites d'étude, des systèmes Web ainsi que des documents observés. Nous nous attacherons, ensuite, à la description des critères de sélection des sites d'étude et du profil des participants à la recherche.

3.3.1 Politiques d'archivage du Web organisationnel

3.3.1.1 Sélection des politiques

Afin d'étudier le cadre général de gestion du processus d'archivage des sites Web organisationnel, nous avons choisi d'analyser les politiques d'archivage développées par des institutions archivistiques.

Lors d'une première exploration des bases de données bibliographiques au sujet de l'archivage du Web organisationnel, nous nous sommes heurtée au manque de ressources documentaires quant aux pratiques organisationnelles en la matière. Aussi, en raison de leur absence dans les milieux retenus pour cette recherche, nous avons choisi d'effectuer une recherche en ligne de ce type de politiques. Nous avons pu repérer un nombre intéressant de politiques diffusées, entre autres, sur le site Web des archives nationales et de politiques mises en œuvre dans quelques institutions universitaires. Ces politiques constituent ainsi les principales ressources documentaires sur l'archivage du Web dans les organisations. Nous avons repéré au départ 33 politiques qui résultent d'une recherche et d'un processus de veille sur le Web. Les politiques sélectionnées ont été retenues en fonction de la notoriété de l'organisme producteur dans le domaine des archives ou de l'état d'avancement du projet d'archivage dont elles résultent ou pour lequel elles ont été développées (le programme d'archivage doit être bien établi). À la suite d'une première analyse sommaire de ces politiques se sont ajoutés 22 documents complémentaires dont l'analyse complète les informations recueillies du corpus initial. Le Tableau III (p. 96) présente la répartition des textes recueillis selon le type d'organisme qui les a produits.

Tableau III : Répartition des politiques et documents complémentaires analysés selon l'organisme créateur.

TYPE D'INSTITUTION	NOMBRE DE DOCUMENTS ANALYSÉS
Archives (et bibliothèques) nationales	15
Archives (et bibliothèques) provinciales	11
Archives universitaires (université et service d'archives)	17
Association professionnelle	1
Bibliothèques nationales	3
Institutions gouvernementales	8

Les politiques analysées sont principalement des politiques établies par des institutions archivistiques nationales (archives nationales ou archives provinciales) ou par des services d'archives universitaires. Notre échantillon de politiques, de par sa nature, nous offre ainsi la possibilité de comparer des cadres de gestion de nature théorique aux cadres de gestion issus de l'implantation d'une stratégie d'archivage du Web.

Lorsque nous avons entamé l'analyse de ces politiques, il s'est avéré nécessaire d'inclure d'autres documents qui complètent, mettent en contexte ou précisent les informations indiquées dans le corpus initial (n=22). Nous avons analysé au total 55 documents en provenance de 5 pays différents (voir Tableau IV, p. 97 et Annexe 2, p. xvii pour la liste complète des politiques réparties par pays et par type d'organisme créateur). Les références citées dans cette liste sont également intégrées, par ordre alphabétique d'auteur, à la bibliographie.

Tableau IV : Répartition des politiques et documents complémentaires selon le pays.

PAYS	NOMBRE DE DOCUMENTS ANALYSÉS
Australie	34
Canada	7
États-Unis	4
Grande-Bretagne	7
Nouvelle-Zélande	3
Total	55

Nous avons cherché à constituer un corpus de politiques représentatif des milieux étudiés. Les 33 politiques d'archivage du Web organisationnel repérées en ligne sont diffusées par 5 archives nationales (Australie, Canada, États-Unis, Grande-Bretagne et Nouvelle-Zélande), 4 archives provinciales australiennes (New South Wales, Queensland, Tasmania et Victoria), 3 milieux universitaires (University of Melbourne, University of Tasmania et University of Toronto) et une association professionnelle. Nous n'avons pas pu repérer de politiques diffusées par un organisme gouvernemental ou un organisme professionnel. Nous présumons que les politiques recommandées par les archives nationales et les archives provinciales s'adressent également aux organismes publics et gouvernementaux. Donc, nous pouvons postuler que les politiques d'archivage du Web organisationnel repérées en ligne sont représentatives de 3 des 4 types de milieux étudiés dans le cadre de cette recherche, soit le milieu universitaire, le milieu gouvernemental et le milieu archivistique. Ces recommandations restent aussi valables pour le milieu professionnel puisqu'elles portent, de manière générale, sur les documents Web produits dans un contexte organisationnel.

3.3.1.2 Caractéristiques des politiques

La majorité des politiques identifiées (n=30) portent sur l'archivage du Web organisationnel, alors que d'autres (n=15) présentent des informations générales ou décrivent des aspects techniques des projets d'archivage du Web mis en œuvre, à titre d'exemple, par des archives nationales (ex. : BAC, 2007-2008) ou des services d'archives universitaires (ex. : The University of Melbourne, 2008e). Les recommandations relatives à l'archivage des sites Web sont, dans certains cas (n=2), intégrées aux politiques générales sur la gestion des archives

numériques (ex. : University of Tasmania, 2010c). Dans d'autres cas (n=4), la consultation de politiques relatives aux standards d'édition Web était nécessaire, car, à n'en pas douter, le mode de conception des documents affecte et conditionne le processus d'archivage. Nous avons, à titre d'exemple, analysé des politiques qui traitent des règles de redirection des liens vers les pages archivées ou des mesures à prendre en compte par les webmestres avant le retrait en ligne ou la refonte majeure d'un site Web. Dans une proportion égale (n=4), certaines politiques sont émises conjointement par des archives nationales et la bibliothèque nationale d'un pays afin de documenter une approche collaborative en matière d'archivage du Web à l'échelle nationale. Cet échantillon de politiques est particulièrement intéressant à étudier, car il nous permet d'entrevoir la manière dont ces institutions se partagent les rôles et responsabilités dans un territoire où les frontières entre publication et document d'archives ne sont pas évidentes à établir. Les définitions proposées dans ces politiques nous aident justement à discerner les critères d'identification des archives dans un environnement Web. Le Tableau V (p. 98) présente en chiffres la répartition des politiques et documents complémentaires selon les sujets qu'elles couvrent.

Tableau V : Répartition des politiques et documents complémentaires selon les sujets couverts (n=55).

SUJET(S) COUVERT(S)	NOMBRE DE DOCUMENTS ANALYSÉS
Politiques propres à l'archivage des sites Web	30
Politiques sur la gestion des archives numériques uniquement	2
Politiques sur la gestion des archives numériques et publications en ligne	4
Politiques sur les règles d'édition Web qui ont un impact sur le processus d'archivage des sites Web	4
Documents d'information sur les projets d'archivage du Web en lien avec les politiques analysées	15

Quelques-unes des politiques repérées au début de la phase de collecte des données ne sont plus en ligne (BAC, 1999; NAA, 2001a; Public Record Office, 2001). Lorsqu'on s'y réfère dans les nouvelles politiques (ex. : NAA, [s.d.]), nous les conservons aux fins de l'analyse. Dans le cas contraire, nous les remplaçons par la nouvelle version (ex. : TNA, 2009a). Nous

décrivons dans le champ Commentaires les particularités de la politique révolue. D'autres politiques ont été mises à jour, nous avons intégré ces mises à jour au fur et à mesure lors de la phase d'analyse. Ce processus de mise à jour systématique a été arrêté à l'hiver 2012.

Malgré leur avènement sur les sites Web d'archives nationales, nous n'avons pas systématiquement intégré les politiques sur l'archivage du Web 2.0 au corpus documentaire à analyser puisque ce sujet n'est pas l'objet principal de notre recherche. Nous nous sommes limitée à l'analyse de deux politiques sur l'archivage du Web 2.0 dans une perspective archivistique (NARA, [s.d.]a; NSW, 2009b).

3.3.2 Sites d'étude

3.3.2.1 Sélection des sites d'étude

Le phénomène à l'étude se produit dans un contexte organisationnel. Les sites d'observation doivent être minutieusement établis. Un site est défini comme un : « contexte circonscrit dans lequel on étudie des événements, des processus et des résultats. » (Pires, 1997: 140). Les trois critères principaux qui ont déterminé le choix des sites d'études dans le cadre de notre recherche sont les suivants :

- Présence de l'organisation sur le Web : l'organisation doit avoir un site Web public.
- Mise en œuvre de l'archivage du Web dans l'organisation :
 - o Développement/application d'une politique d'archivage des sites Web de l'organisation, ou
 - o Développement/application de règles propres aux sites Web de l'organisation, ou
 - o Réflexion en cours pour établir un programme d'archivage du Web dans l'organisation.
- Situation géographique :
 - o Province de Québec.

Afin de contourner la difficulté rencontrée à identifier des milieux qui répondent à ces critères, nous avons choisi un premier terrain d'étude par convenance. Ce premier terrain d'étude a, entre autres, servi pour le prétest des instruments de collecte. L'effet boule de neige et réseau a permis d'identifier les autres milieux d'études :

- Certains répondants nous ont référée vers d'autres institutions.

- Nous avons aussi eu accès à certains milieux par le biais d'un message adressé aux membres du Groupe d'expertise en gestion documentaire du gouvernement du Québec.
- Finalement, trois organismes ont été identifiés à travers notre réseau de contacts professionnels.

Au bout de ces efforts, nous avons réussi à avoir 11 sites d'étude répartis entre Montréal et la ville de Québec, le nombre de cas compris entre quatre et dix étant généralement souhaitable (Gagnon, 2005: 41; Eisenhardt, 1989: 545).

3.3.2.2 Caractéristiques des sites Web publics analysés

Dans le cadre de cette recherche, les sites Web publics de 11 organismes québécois ont été observés. Nous présentons, dans un premier temps, les caractéristiques générales des sites Web observés. Nous décrivons, ensuite, les caractéristiques techniques de ces derniers.

L'adresse officielle des sites Web observés nous a servi comme point de départ pour délimiter les contours de ces sites et aidé à distinguer le contenu publié sous le contrôle du site à l'étude de ceux qui en sont liés. Lors de l'exploration de ces sites Web, des sous domaines et d'autres noms de domaines ont été identifiés dans quelques cas (ex. : sous-sites, blogue institutionnel, site lié à un événement) que nous avons considéré lors de l'observation et aussi lors de la constitution de l'échantillon des documents à analyser.

Les sites Web observés ont été créés à la fin des années 90 (un seul site Web a été créé en 2004) et ont subi quelques refontes (en moyenne : 2-3 refontes majeures). Ils se répartissent en 4 principales catégories résumées dans le Tableau VI (p. 100). Cette catégorisation est établie selon la nature de l'organisme créateur.

Tableau VI : Catégories des sites Web observés dans les sites d'étude (n=11).

CATÉGORIE DES SITES WEB OBSERVÉS	NOMBRE DE SITES WEB OBSERVÉS
Site Web académique	2
Site Web de services d'archives universitaires	4
Site Web gouvernemental	4
Site Web professionnel	1

Nous avons constaté des pratiques différentes quant à l'identification des responsables sur les sites Web observés. Pour la majorité d'entre eux (n=7), il n'a pas été possible d'identifier avec précision la responsabilité de gestion du site Web observé (nom de la personne/unité responsable de la gestion du site Web). Un seul site Web (SIW01) sur les 11 observés comporte une page Web qui décrit d'une manière exhaustive l'historique du site (date de création, refontes majeures, technologies de développement, etc.). Finalement, trois autres sites nous renseignent sur l'équipe/personne responsable de la gestion du site Web. Tous les sites offrent, sans exception, la possibilité de communiquer avec le webmestre par courrier électronique ou grâce à un formulaire Web.

Les sites Web observés se distinguent aussi au niveau des caractéristiques suivantes :

1. Objectifs du site Web : les sites Web observés visent en général la diffusion de l'information au sujet de l'organisme, les activités et les services qu'il offre (n=11). Il s'agit d'une « vitrine de ce qui se fait » dans l'organisme (R03), une « image de l'institution » (R04.1). Le site Web pourrait être aussi un moyen d'interaction avec la clientèle et les partenaires (n=6). Dans ce cas, le site Web donne accès à des applications et services en ligne (accès sécurisé).
2. Clientèle visée par le site Web : les sites Web observés s'adressent principalement à la clientèle externe de l'organisme, que ce soit le grand public ou une catégorie de clientèle bien ciblée (ex. étudiants, chercheurs, jeunes, personnes âgées, partenaires d'affaires) (n=5). Certains sites répondent à la fois aux besoins informationnels de la clientèle externe et à ceux de la clientèle interne de l'organisme, c'est-à-dire les employés ou encore les membres de la communauté à laquelle appartient l'organisme (n=6). Dans trois cas (SIW06, SIW07 et SIW10), le contenu informationnel peut être personnalisé selon le profil des internautes intéressés par le site (public cible ou grand public), en fonction de l'âge ou selon le statut du client.
3. Fréquence de mise à jour : 5 sites Web affichent des dates de mise à jour au niveau des pages Web observées qui varie selon les pages. Pour les autres sites Web, seule la date de la dernière mise à jour est affichée sur les pages Web observées ou uniquement au niveau de la page d'accueil. L'information concernant la fréquence de mise à jour a été précisée lors des entrevues. La plupart des sites Web observés sont mis à jour

régulièrement, fréquemment et même quotidiennement (n=10). Pour un site, la mise à jour est variable, selon les besoins ponctuels.

4. Typologie : tel qu'illustré dans le Tableau VII (p. 102), les sites Web observés sont de nature variée par rapport à leur technologie de développement, certains sont de nature statique, d'autres sont dynamiques ou à quelques composantes dynamiques.

Tableau VII : Types des sites Web observés selon la technologie de développement (n=11).

TYPE DES SITES WEB OBSERVÉS	NOMBRE DES SITES WEB OBSERVÉS
Site Web statique	5
Site Web à composantes dynamiques	5
Site Web dynamique	1

De plus, on retrouve aussi des différences au point de vue technologique. Le protocole de communication communément utilisé par les sites Web observés est HTTP. Quelques sites (n=6) ont une section à accès sécurisé (accès à une application, accès à un extranet, etc.). Quatre organisations publient en ligne sous un nom de domaine unique. Les sept autres possèdent plusieurs noms de domaines. Ce cas de figure se présente lorsque des sous-sites ou des sites événementiels sont publiés en marge du site principal ou lorsque des sections particulières du site principal sont sécurisées. Ceci se produit aussi lorsque des sections particulières du site sont sous la responsabilité d'autres unités de l'organisme ou développées pour une clientèle donnée ou réservées à un produit/service spécifique.

Les sites Web observés sont produits en HTML, PHP et/ou ASP. Leur structure et mise en forme sont gérées grâce à des feuilles de style. À l'exception du SIW11, l'ensemble des sites Web observés présente une identité visuelle constante. Le multifenêtrage est présent uniquement et partiellement sur un seul site Web observé (SIW03).

Le contenu de ces sites Web est réparti à l'horizontale (niveau de profondeur moyen : N3-N5), le niveau de profondeur le plus bas étant le niveau 5. La navigation se fait généralement grâce aux menus en haut d'une page (ou onglets horizontaux) ou dans la colonne gauche de la fenêtre d'affichage.

L'accès à des bases des données relationnelles (ex. : MySQL) est possible sur 7 sites Web. Les moteurs de recherche sont présents sur 9 sites Web. L'aspect interactionnel des sites Web observés se limite sur 4 sites statiques au courrier électronique. Au niveau des autres sites Web, les moyens d'interaction suivants ont été identifiés :

- Formulaires Web (questionnaire, jeu-questionnaire, rétroaction);
- Applications en ligne (jeux en ligne, réservation, inscription);
- Applications Web 2.0 (Facebook, Twitter, fils RSS).

Les sites Web analysés, de par leur nature, objectifs, clientèle, technologies de développement, structure, contenu et niveau d'interactivité se différencient les uns des autres. Seuls les sites Web gouvernementaux ont une identité visuelle similaire. La variété des sites Web observés nous offre l'occasion de vérifier si les stratégies de leur archivage sont différentes et de déterminer les facteurs liés aux sites Web eux-mêmes qui sont à l'origine de cette variation au niveau des pratiques d'archivage.

3.3.3 Participants à l'étude

Le recrutement de nos répondants dépend de leur implication active dans le processus de gestion et d'archivage du Web de leur organisation, ceci afin de nous assurer de leur disposition à nous fournir des informations riches et exactes (Fortin, 1996; Miles et Huberman, 2003; Patton, 2002).

La littérature professionnelle dans le domaine des archives identifie différents acteurs devant être impliqués dans la gestion et la conservation des archives Web (Khouaja et Couture, 2003-2004: 28; Nicholls et Williams, 2002). Nous avons ainsi rencontré des archivistes (n=13) et des webmestres (n=8). Les archivistes constituent le groupe le plus accessible dans notre cas. Nous leur avons adressé les premiers des lettres d'invitation pour participer à cette recherche (voir Annexe 12, p. 1). Notre intention, lors des entrevues, était de demander à chaque archiviste interviewé de nous mettre en contact avec d'autres catégories d'employés impliquées dans la démarche d'archivage du site Web de son organisation et ainsi ajouter leurs noms à notre plan d'échantillonnage. Sans l'avoir demandé, la plupart des archivistes rencontrés (n=11) ont invité d'autres intervenants impliqués ou devant être impliqués dans l'archivage des SIW de leur organisation. Dans un seul cas (SIW02), l'archiviste nous a mis en contact avec le webmestre

du site Web de son institution. Dans deux autres cas, les archivistes rencontrés n'ont pas invité d'autres intervenants, soit parce que l'archiviste est lui-même le webmestre du site Web de son unité organisationnelle, soit parce que l'archiviste a une bonne connaissance de l'historique de développement du site Web de son institution, de ses opérations de gestion et de ses caractéristiques techniques. Tous les participants à cette recherche ont été remerciés d'avoir accepté de participer à cette recherche (voir Annexe 13, p. li).

Les participants à cette recherche ont alors été sélectionnés par une technique d'échantillonnage empirique non probabiliste, plus précisément un échantillonnage de convenance par « effet boule de neige » (Patton, 2002; Poupart, 1997). Au total, nous avons interviewés 21 individus œuvrant dans 3 types de milieux organisationnels différents (voir Tableau VIII, p. 104).

Tableau VIII : Répartition des participants à la recherche selon les milieux organisationnels.

MILIEUX	CATÉGORIE DU PARTICIPANT		TOTAL
	Webmestres	Archivistes	
Milieu gouvernemental (n=5)	5	7	12
Milieu universitaire (n=5)	3	5	8
Milieu professionnel (n=1)		1	1

Femme = 12
Homme = 9

La catégorie des webmestres regroupe en réalité 5 webmestres et 3 édimestres. Nous avons décidé de les regrouper sous la même catégorie puisque les tâches qu'ils effectuent au niveau des sites Web sont quasi-similaires. Parmi les 13 archivistes rencontrés, 5 sont impliqués dans la gestion et l'archivage de leur site Web institutionnel ou du site Web de leur unité (ex. : développement du contenu, préparation de règles de conservation), les 7 autres n'ont pas de lien avec le site de l'organisation. Ils accomplissent des tâches de gestion documentaire classique. Un seul archiviste parmi l'échantillon des participants est webmestre de son unité.

Les archivistes interviewés ont une maîtrise en bibliothéconomie et/ou sciences de l'information (n=7), ou un certificat en archivistique (n=3). Les webmestres ont généralement une formation collégiale (n=1), un baccalauréat en communication (n=2), une maîtrise en sciences de l'information (n=1) ou une maîtrise en informatique (n=1). Nous n'avons pas pu obtenir les informations relatives au niveau de formation de 7 participants.

Les archivistes participant à cette recherche se distinguent par leur niveau de formation (maîtrise en SI pour la plupart) et le nombre d'années d'expériences dans le domaine de gestion des archives (en moyenne, entre 5 à 10 généralement dans le même lieu de travail). Deux catégories de webmestres se dessinent. Des webmestres avec une expérience de plus que 20 ans dans leur domaine d'expertise (n=3) et des webmestres d'une expérience entre 5 à 7 ans (n=2)¹⁷.

3.4 Analyse des données

Dans le cadre de notre projet de recherche, les données issues de l'observation des sites Web publics, des entrevues et de l'analyse des politiques d'archivage ainsi que la documentation recueillie lors des entrevues ont été soumises à une analyse qualitative basée sur une approche mixte, c'est-à-dire déductive et inductive (Bergeron, 1997; Miles et Huberman, 2003; Patton, 2002; Poupart, 1997; Wildemuth, 2009). Ce type d'analyse consiste à prédéfinir des catégories initiales d'analyse, qui, dans le présent contexte, proviennent du cadre conceptuel et de nos grilles d'observation (approche déductive). Les catégories prédéterminées ont été complétées par celles qui ont émergé des données de recherche (approche inductive). Le logiciel d'analyse NVivo a été utilisé puisque la démarche adoptée se prête bien à une analyse assistée par ordinateur (Miles et Huberman, 2003; Patton, 2002; Tesch, 1990). L'itération entre collecte et analyse de données permet de raffiner notre méthode et catégories d'analyse (Eisenhardt, 1989; Miles et Huberman, 2003; Patton, 2002; Yin, 2003).

Après avoir énoncé la méthode d'analyse appliquée aux données de recherche, nous expliquons dans ce qui suit le processus d'analyse. Nous indiquons, ensuite, les unités d'analyse employées dans le cadre de cette recherche. Enfin, nous décrivons la chronologie de l'analyse des données.

3.4.1 Processus d'analyse

L'analyse des données est le processus qui permet de : « transformer les données en résultats de recherche. » (Patton, 2002: 432 – notre traduction). Ce processus se déroule en deux

¹⁷ Nous n'avons pas obtenu l'information relative au nombre d'années d'expérience auprès de 3 webmestres.

principales phases : organisation des données et interprétation des données (Miles et Huberman, 2003). Nous avons choisi de tenir compte uniquement du contenu manifeste.

La première phase analytique consiste à procéder à un codage descriptif permettant d'obtenir une première vue des données à analyser et un ensemble d'informations contextuelles. Après cette opération de découpage et d'étiquetage (décontextualisation), nous avons effectué des regroupements et une catégorisation de segments significatifs de textes en y associant, de manière déductive, des catégories prédéterminées à partir du cadre conceptuel et des grilles d'observation (Miles et Huberman, 2003; Weldmuth, 2009:310). Un guide de codage a été construit pour encadrer cette opération d'analyse par allers-retours entre les catégories issues de nos questions de recherche, du cadre conceptuel et des instruments de collecte (catégories prédéfinies) et celles empiriquement enracinées (catégories émergent lors du processus d'analyse) (Mukamurera et al., 2006: 115). Le guide a servi pour coder l'ensemble des données recueillies par les différents instruments de collecte (grilles d'analyse et guide d'entrevue). Il propose une définition et un exemple d'application, lorsque pertinent, pour chaque code développé (voir Annexe 15, p. lv).

Nous avons appliqué un codage multiple (Miles et Huberman, 2003: 125-126; Weldmuth, 2009: 310), c'est-à-dire que plusieurs codes ont été attribués à un même segment de texte (ex. : les codes « Approche d'archivage » et « Technique d'archivage » se trouvent souvent superposés au sein du même paragraphe codé). Un même code peut également servir à coder des segments différents de textes (ex. : le code O01 a été appliqué au niveau de l'ensemble des fichiers relatifs aux données recueillies auprès de l'organisme O01).

Nous avons tiré profit du chevauchement entre la collecte des données et leur analyse pour raffiner notre arbre de codage. L'arbre de codage a évolué au fur et à mesure du déroulement de l'étude en intégrant des catégories ayant émergé de façon inductive au cours du processus d'analyse des données du terrain. À cet effet, NVivo offre une grande souplesse dans la manipulation et la catégorisation de données multiples. Pour assurer la stabilité et la constance du codage et de la catégorisation de nos données, nous avons veillé à bien décrire les valeurs des catégories développées afin qu'elles soient précises et mutuellement exclusives :

« The best test of the clarity and consistency of your category definitions is to code a sample of your data. After the sample is coded, the coding consistency

needs to be checked, in most cases through an assessment of intercoder agreement. » (Weldmuth, 2009: 311)

Afin de nous familiariser avec les données et mettre en évidence les caractéristiques de chaque site étudié, nous avons établi des rapports individuels de nature descriptive (analyse intrasite). Dans une autre étape, nous avons regroupé les sites d'étude par type de milieu (milieu gouvernemental, milieu d'enseignement et milieu professionnel) et nous avons procédé à une analyse comparative des résultats afin d'identifier les similarités et les différences entre les milieux étudiés quant à leur démarche d'archivage. Pareillement, nous avons analysé individuellement chaque politique. Nous les avons ensuite regroupées selon le milieu visé (milieu universitaire et milieu gouvernemental) et selon la nature de la politique (politique qui sert de cadre de référence ou politique qui encadre la mise en œuvre un projet d'archivage du Web). Enfin, nous avons confronté les conclusions préliminaires du terrain aux résultats de l'analyse des politiques d'archivage. Des relations ont émergé graduellement entre les variables (ex. : lien entre nature du site Web, méthode d'archivage et technique d'archivage; perception de l'archiviste et stade de développement du projet d'archivage). Cette démarche comparative nous a permis d'aller au-delà de nos premières impressions tout en utilisant des perspectives variées dans l'appréhension des données.

3.4.2 Unité d'analyse

L'unité d'analyse est l'objet ou le phénomène que le chercheur souhaite étudier et sur lequel il veut dire quelque chose à la fin de son étude (Babbie, 2005: 96). L'identification de l'unité d'analyse est cruciale afin de mieux conduire les observations à faire (qui ou quoi sera observé) et de communiquer des interprétations valides sur l'unité d'analyse et non sur l'unité d'observation (Babbie, 2005: 103; Patton, 2002: 229).

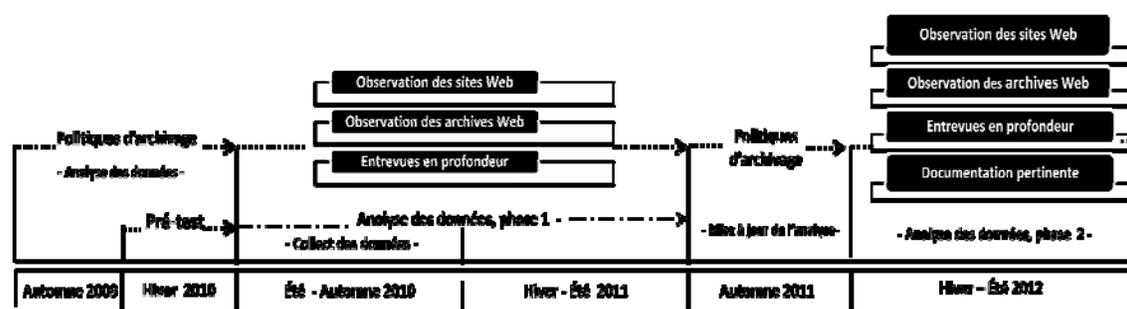
Dans le cadre de cette recherche, quatre unités d'analyse ont été considérées, soit : les sites Web publics observés, les archives Web produites grâce à ces sites Web publics, les stratégies d'archivage de ces systèmes Web et les intervenants impliqués dans leur processus d'archivage.

3.4.3 Chronologie de l'analyse des données

L'analyse des données s'est réalisée en plusieurs étapes et de manière itérative (voir Figure 4, p. 108) :

- Étape 1 : analyse des politiques d'archivage du Web organisationnel avec un exercice d'analyse des données de recherche lors du prétest des instruments de collecte de données (automne 2009 - hiver 2010);
- Étape 2 : codification des données issues des cinq premières collectes de données (été 2010 - été 2011);
- Étape 3 : analyse des politiques mises à jour (automne 2011);
- Étape 4 : deuxième vague d'analyse des données collectées sur les sites d'étude (hiver-été 2012).

Figure 4 : Chronologie de l'analyse des données.



L'analyse du matériel collecté offre des réponses à notre questionnement du départ qui porte sur la démarche actuelle d'archivage des sites Web dans un contexte organisationnel. Elle valide l'assertion révélée par notre analyse de la littérature concernant l'émergence de nouveaux types de documents suite à l'implantation et l'usage de nouvelles technologies. Elle permet également de préciser les contours de l'intervention des personnes impliquées dans le processus d'archivage des sites Web publics d'une organisation et apporte des éléments complémentaires au sujet des stratégies d'archivage.

3.5 Qualité de la recherche

Afin d'assurer la rigueur méthodologique d'une recherche de nature qualitative, différentes mesures doivent être appliquées lors de la préparation du plan d'étude et la

réalisation de la recherche (Gagnon, 2005; Miles et Huberman, 2003; Patton, 2002; Rispał Hlady, 2002; Yin, 2003). Les mesures concourant à la crédibilité, à la fiabilité, à la constance et à transférabilité des résultats de notre étude sont explicitées ci-après.

3.5.1 Préparation de la recherche

Afin de nous préparer à la réalisation de notre recherche, nous avons pris en considération trois principaux moyens, soit : l'acquisition des habiletés de recherche, la conduite d'un prétest et l'élaboration d'un protocole de recherche (Patton, 2002; Yin, 2003).

3.5.1.1 Habiletés de la chercheure

Nous avons eu l'occasion de participer à quelques projets de recherche au cours de notre parcours au programme du doctorat où nous avons appris, entre autres, à concevoir des instruments de collecte de données et à traiter des données de recherche tout en assurant leur anonymat et respectant les règles de confidentialité. Ceci nous a permis de faire preuve de rigueur, de maîtriser différents modes de collecte de données et d'être sensible non seulement aux stratégies et moyens qui garantissent la qualité d'une recherche, mais aussi aux sources de biais.

Nous étions initiée aux techniques de l'observation et de l'entrevue ce qui nous a facilité la conception des grilles d'observation et du guide d'entrevue ainsi que la conduite des entrevues. Pour être efficaces, les instruments de collecte doivent répondre aux questions de recherche et être appliqués de manière uniforme. Pour ce faire, nous nous sommes assurée de bien établir le lien entre les questions de recherche et les instruments de collecte.

Pour augmenter la crédibilité et la fiabilité de notre recherche, nous avons développé des outils de contrôle de la cohérence de notre démarche de collecte et d'analyse de données (ex. : protocole de recherche, guide de codage). Nous avons documenté systématiquement les décisions prises au cours de la réalisation de cette recherche. En complément à ces mesures, nous avons également accordé une attention particulière à la triangulation des données et à la validation de nos résultats auprès des pairs.

3.5.1.2 Conduite d'un prétest

La conduite d'un prétest aide à mieux évaluer l'adéquation entre ces instruments de collecte de données et les questions de recherche. La conduite du prétest nous a offert l'occasion de valider l'adéquation des outils de collecte aux questions de recherche, de vérifier la stabilité de ces outils et leur capacité à collecter des données de qualité. C'était un moment crucial de la recherche qui nous a permis de maîtriser nos outils, mais surtout de préciser nos questions de recherche.

3.5.1.3 Élaboration d'un protocole de recherche

L'élaboration d'un protocole de recherche a desservi différentes fins. Tout d'abord, la chercheure ne risque pas de perdre de vue les objectifs de la recherche. Ensuite, le recours à cet outil permet de préserver une certaine conformité dans l'application des règles et procédures d'utilisation des instruments de collecte et des procédures d'analyse.

Le protocole développé rappelle, entre autres, les questions de recherche. Il comprend également le calendrier de réalisation de la recherche, les modes et les instruments et de collecte et d'analyse ainsi que le lien entre ces derniers et les questions de recherche.

Cet outil nous a guidée dans la réalisation de notre étude et contribue à rendre transparent notre processus de recherche (Miles et Huberman, 2003; Yin, 2003).

3.5.2 Réalisation de la recherche

Nous discutons dans les points suivants des mesures de qualité mise en œuvre lors de la conduite de cette recherche.

3.5.2.1 Crédibilité de la recherche

La crédibilité consiste à démontrer la concordance entre les propos des répondants et les interprétations formulées par le chercheur (Rispał Hlady, 2002: 102). Pour ce faire, nous avons établi une stratégie de triangulation des modes de collecte de même que des sources de preuves (Denzen et Lincoln, 2003; Lessard-Hébert et al., 1995; Patton, 2002; Yin, 2003). Nous nous basons non seulement sur l'observation de sites Web publics et des documents qu'ils diffusent,

mais aussi sur l'expérience des individus telle que décrite dans les entrevues et dans la littérature professionnelle.

Au cours de la réalisation de cette étude, nous sommes retournée souvent aux questions de recherche afin d'établir une chaîne d'évidence entre ces dernières et les données collectées (Yin, 2003). L'établissement du plan d'échantillonnage nous a orienté vers le choix de sites adéquats à l'investigation de nos questions de recherche tout en faisant appel à divers répondants bien informés et impliqués à des niveaux différents dans le processus d'archivage de sites Web publics. Des données riches et pertinentes ont été obtenues. En nous basant sur une approche d'analyse comparative, nous avons été attentive à rechercher constamment des explications ou des propositions contradictoires et, de cette manière, nous avons pu démontrer avec confiance la convergence au niveau des conclusions (Rispał Hlady, 2002).

En complémentarité de l'itération entre la chercheuse et ses données, nous avons tenu à corroborer certaines observations réalisées avant les entrevues avec les participants à l'étude (triangulation des acteurs). Un échantillon du matériel collecté et analysé a été également soumis à une validation par les pairs (triangulation des chercheurs) (Miles et Huberman, 2003; Mukamurera et al., 2006). De plus, nous avons publié les résultats préliminaires de notre recherche dans quelques revues avec comité de lecture. Ces résultats ont été également exposés dans le cadre de quelques manifestations scientifiques.

3.5.2.2 Stabilité des données

Il s'agit de prouver la justesse, l'authenticité et la plausibilité des résultats par rapport à la réalité du terrain ou à une théorie (antérieure ou émergente). Le chercheur doit prouver qu'il a bien compris et saisi la perception des informateurs de même que la dynamique du phénomène à l'étude (Rispał Hlady, 2002).

Pour renforcer la confiance dans nos résultats de recherche, nous avons, tout d'abord, défini avec soin les concepts à l'étude et établi les liens qui existent entre les questions de recherche et les instruments de collecte des données (voir Tableau II, p. 93). De cette façon, la chercheuse s'est assurée de colliger des informations qui répondent aux objectifs de recherche.

Nous avons également vérifié la constance des résultats en invitant des collègues à réaliser le même processus d'observation et d'analyse. Nous nous sommes assurée de la

pertinence de notre stratégie d'échantillonnage et discutée de nos résultats de recherche avec des collègues et, surtout, avec notre comité de recherche (triangulation des chercheurs) (Gagnon, 2005; Patton, 2003).

3.5.2.3 Transférabilité des résultats

Nous avons adopté une approche d'étude de nature qualitative, par conséquent, nous ne visons pas à vérifier la généralisation de nos résultats de recherche, mais plutôt atteindre une « *représentativité théorique* » (Miles et Huberman, 2003: 59). Nous souhaitons contrôler la transférabilité de ces résultats à d'autres contextes (Gagnon, 2005: 35). Pour atteindre cet objectif, nous avons veillé à produire une description étoffée de l'échantillon étudié (sites d'étude, systèmes observés et participants à la recherche) (voir section 3.3, p. 94). Ceci permet d'établir des similitudes avec d'autres contextes (Miles et Huberman, 2003).

Nous avons également pris soin de sélectionner des sites et des participants en fonction de la pertinence de leurs caractéristiques aux objectifs de notre recherche. L'analyse intersites nous a assuré de la pertinence théorique de notre échantillon puisqu'elle nous révèle que nos sites incluaient aussi bien des sites Web de nature statique que dynamique, des projets d'archivage du Web bien aboutis et d'autres au stade de développement (Yin, 2003: 133). Quelques participants sont non seulement responsables de l'archivage des sites Web publics étudiés, mais aussi impliqués dans leur gestion et maintenance. Des informations riches et complètes ont pu être collectées sur la gestion de l'ensemble du cycle de vie d'un site Web. La diversité des perspectives quant au sujet étudié était assurée. Lors du processus d'analyse, nous avons noté une redondance d'information dans les dernières entrevues et aucune nouvelle catégorie n'a été ajoutée au bout de la neuvième entrevue. Nous avons atteint la saturation théorique, notre échantillon était bien représentatif (Mukamurera et al., 2006).

Pour augmenter la transférabilité des résultats, Miles et Huberman (2003) recommandent la comparaison des résultats de recherche avec ceux des études antérieures ou la consultation d'experts dans le domaine (Miles et Huberman, 2003). À ce niveau, nous avons comparé les nôtres aux résultats de différentes enquêtes menées au sujet de l'archivage du Web (Day, 2003b; Gomes, 2011; Gregory, 2009; Internet Memory Foundation, 2010; NDSA, 2012). De même, nous avons consulté les responsables du dépôt légal numérique à BANQ pour les

questionner sur l'état actuel de la réflexion au sein de leur institution sur le sujet étudié (voir Annexe 14, p. lii). Nous avons également communiqué avec la direction des centres d'archives à BAnQ au sujet des règles de conservation établies pour gérer les sites Web.

3.5.2.4 Fiabilité de la recherche

La fiabilité de la recherche ne porte pas directement sur les données, mais vise plutôt les instruments et les procédures de collecte des données (Lessard-Hébert et al., 1995). Elle se définit en termes d'utilité, de précision et de stabilité des outils utilisés de même que dans la régularité dans leur emploi (Rispal Hlady, 2002). Elle consiste, tout simplement, à vérifier si un autre chercheur est capable de reproduire la même expérience vécue sur le terrain et d'aboutir aux mêmes conclusions, en suivant les procédures utilisées par le premier chercheur (Rispal Hlady, 2002; Yin, 2003). Afin de pouvoir répondre au critère de fiabilité, nous avons décidé de :

- expliciter et documenter systématiquement toutes les décisions prises tout au long de la réalisation de la recherche dans des mémos ou notes personnelles;
- élaborer un protocole de recherche afin d'uniformiser la démarche de recherche;
- prétester les instruments de collecte afin de vérifier leur stabilité et capacité à collecter des données riches et de qualité qui répondent aux questions de recherche;
- vérifier certaines données issues de l'observation auprès des participants à l'étude (Lessard-Hébert et al., 1995; Gagnon, 2005; Rispal Hlady, 2002).

Grâce à ces mécanismes, une chaîne d'évidences a pu être établie documentant ainsi le déroulement de la recherche et les décisions qui ont été prises.

Conclusion

L'approche méthodologique a été décrite dans ce chapitre. Nous avons choisi de mener une recherche qualitative de nature descriptive et exploratoire. Nous avons démontré la pertinence de notre choix méthodologique pour étudier un sujet peu documenté de manière empirique, soit l'archivage des sites Web organisationnels. L'objectif de notre recherche est d'analyser les recommandations de politiques référence sur l'archivage des sites Web organisationnels, d'observer les caractéristiques technologiques de 11 des sites Web publics ainsi que les caractéristiques des archives Web qui s'y retrouvent et de décrire les pratiques d'archivage du Web dans un contexte organisationnel.

Trois principaux modes de collecte de données ont été utilisés : l'analyse des politiques d'archivage et de la documentation recueillie, l'observation de sites Web publics et des documents qui y sont diffusés et, enfin, des entrevues avec des intervenants impliqués dans le processus d'archivage de ces systèmes dans 11 organisations au Québec. La pertinence des grilles d'observation et du guide d'entrevue développés pour les besoins de cette recherche a été démontrée. Les données recueillies ont été organisées et analysées selon une approche d'analyse de contenu qualitative de manière déductive et inductive. Les codes et catégories d'analyse ont été consignés dans un guide de codage.

Enfin, des mesures de contrôle de la qualité de la recherche ont été appliquées tant à l'étape de la préparation que de la réalisation de la recherche. Nous avons, à cet effet, mis en œuvre une stratégie de triangulation des données et des sources de preuves ainsi qu'une triangulation des données recueillies auprès des acteurs et des chercheurs. Aussi, nous avons produit une description détaillée de notre échantillon à l'étude et nous avons confirmé sa saturation théorique. Enfin, nous avons développé des règles et des procédures de collecte et d'analyse de données dans un souci d'uniformisation et de transparence de notre processus de recherche.

Les résultats de la collecte et de l'analyse des données sont exposés dans le chapitre 4.

4. Résultats

Introduction

Afin d'apporter des réponses aux questions de recherche qui animent cette étude, nous exposons, dans le chapitre 4, les résultats de l'analyse des données. Rappelons que notre recherche vise à : (1) analyser les recommandations de politiques d'archivage du Web organisationnel, (2) dresser un portrait des types d'archives Web retrouvés sur les sites Web publics de milieux organisationnels au Québec, et (3) décrire les pratiques d'archivage du Web dans ces milieux afin d'évaluer l'adéquation des pratiques actuelles aux caractéristiques des archives Web et aux recommandations des institutions archivistiques en matière d'archivage du Web dans les organisations.

L'analyse des politiques d'archivage du Web organisationnel (33 politiques et 22 documents connexes), l'observation des caractéristiques technologiques et documentaires des sites Web publics des milieux à l'étude (11 sites Web publics et 737 pages Web) ainsi que des entrevues avec les acteurs impliqués dans le processus d'archivage des sites Web dans ces milieux (21 répondants) nous ont permis d'apporter des réponses à nos questions de recherche.

Les résultats de l'analyse seront présentés selon l'ordre des questions de recherche.

4.1 Analyse des politiques d'archivage du Web organisationnel

Afin d'étudier le cadre général de gestion du processus d'archivage des sites Web organisationnel, nous avons choisi d'analyser les politiques d'archivage du Web de quelques milieux organisationnels. En raison de leur absence dans les milieux à l'étude, nous avons recherché des politiques diffusées sur le Web. C'est ainsi que

nous avons repéré et analysé 33 politiques d'archivage du Web organisationnel et une vingtaine de textes connexes (politiques, guides, procédures, etc.).

Puisque nous n'avons pas créé de fiches d'analyse propres aux documents complémentaires et que les résultats de l'analyse de ces derniers ont été intégrés aux fiches descriptives des politiques, nous considérons qu'au total nous avons 33 politiques analysées. C'est sur ce chiffre que nous nous basons lorsqu'il s'agit de caractériser de manière quantitative nos politiques. L'annexe 1 présente les textes étudiés par pays et par catégorie d'organisme (voir Annexe 1, p. i). Les références citées dans cette liste sont également intégrées, par ordre alphabétique d'auteur, à la bibliographie.

Nous présentons dans ce qui suit les résultats de cette analyse. En nous appuyant sur les éléments descriptifs et analytiques de la grille d'analyse, nous décrivons, tout d'abord, la portée et la finalité des politiques analysées. Ensuite, nous identifions les catégories des SIW visés par les politiques retenues de même que les intervenants devant être impliqués dans le processus d'archivage du Web organisationnel. Par la suite, nous dégagons les critères de sélection des archives Web et des unités à archiver. Nous précisons, enfin, les stratégies d'archivage et de préservation de même que les principes sur lesquels s'appuie toute action d'archivage du Web organisationnel.

4.1.1 Portée et finalité des politiques analysées

Les politiques d'archivage retenues (n=33) s'appliquent aux documents produits ou diffusés par une organisation par l'entremise de technologie Web. Ces politiques visent, comme l'indique la citation suivante, à assister les créateurs de contenus, les gestionnaires des sites Web et les archivistes à implanter des stratégies appropriées pour la gestion et la conservation des archives Web :

« This Advice will assist agencies in their efforts to determine what records should be kept, how long they should be kept for, and how they should be treated once captured. » (Public records office of Victoria, 2007: 3)

Ces politiques sont établies dans le respect du cadre réglementaire, des normes et standards de gestion documentaire. Tel que démontré dans le Tableau IX (p. 117), le lecteur y est renvoyé :

- aux textes législatifs et réglementaires en matière de gestion documentaire,
- aux normes de gestion documentaire, ou
- aux politiques et procédures de gestion et d'archivage des sites Web.

Tableau IX : Cadres de référence cités dans les politiques d'archivage du Web organisationnel.

TYPES DE CADRES DE RÉFÉRENCE	CATÉGORIES DE CADRES DE RÉFÉRENCE
Textes législatifs et réglementaires en matière de gestion des documents et des archives	<ul style="list-style-type: none"> – Législations nationales (fédérales ou provinciales) – Lois et règlements relatifs à un domaine d'activité (ex. : secteur gouvernemental) – Politiques internes (ex. : Administrative Functions Disposal Authority, Australie) – Politiques nationales (ex. : Designing and Implementing Recordkeeping Systems, Australie) <p><i>Exemple</i> : « The Public Records Act 2005 requires public offices and local authorities to create full and accurate records of their affairs in accordance with normal prudent business practice. Records associated with websites, intranets and extranets form part of these records. » (Archives New Zealand, 2009: 5)</p>
Normes de gestion documentaire	<ul style="list-style-type: none"> – Normes et standards nationaux (ex. : Australian Standard AS 4390–1996 sur le Records Management) – Standards sur les métadonnées – Standard sur la gestion des risques – Normes régionales (ex. : Recordkeeping Metadata Standard for Commonwealth Agencies) – Normes internationales (ex. : ISO 15489) <p><i>Exemple</i> : « Ensuring the accessibility of web-based records over time raises the same issues that apply to other electronic or paper-based records. Australian Standard AS 4390–1996, <i>Records Management</i> emphasises the need to maintain the functionality of a record over time for as long as required. » (NAA, 2001a: 33)</p>
Politiques, procédures de gestion et d'archivage des sites Web	<ul style="list-style-type: none"> – Guide pour le développement de sites Web – Liste de contrôle pour la vérification des mesures de gestion documentaire à appliquer dans le cas d'un site Web <p><i>Exemple</i> : « The most important collection of requirements can be found in the Victorian Government Website</p>

TYPES DE CADRES DE RÉFÉRENCE	CATÉGORIES DE CADRES DE RÉFÉRENCE
	Management Framework. The Framework contains a number of standards for different aspects of website management. Applying these standards, which represent good website practice, will certainly make the task of creating good records from websites easier.» (Public Record Office Victoria, 2007: 11)
Autres politiques référence établies par des institutions archivistiques	<ul style="list-style-type: none"> – Politique d’archives nationales – Politiques d’archives provinciales – Politiques d’archives universitaires – Politiques d’organismes gouvernementaux <p><i>Exemple</i> : « Please refer to the Archives Office of Tasmania State Records Guideline No 15 on Recordkeeping for Websites and Web Pages and, The University Web Services – Policies and Standards web page for further information. » (University of Tasmania, 2010: 11)</p>

Dans leur ensemble, les politiques analysées, exemples à l’appui, visent à :

- présenter les stratégies possibles d’archivage et de préservation des sites Web,
 - « There are currently three viable options for capturing records from a website. These are: harvesting the website; capturing the web transactions; and capturing the records from the back end systems. » (Public Record Office Victoria, 2008: 8)
- référer le lecteur (webmestre, archiviste ou créateur de contenus) aux principes généraux de gestion d’un site Web,
 - « Sub-domains should also have associated XML Sitemaps. Further guidance on XML Sitemaps is available in ‘Exposing your website to search engines’ (TG122). Guidance on the use of sub-domains in ‘Naming and registering websites’ (TG101). » (Central Office of Information, 2009)
- rappeler les principes archivistiques sur lesquels doit être fondée toute intervention de capture, de gestion et de préservation des archives Web.
 - « For web records the following requirements from the *Electronic Recordkeeping Metadata Standard* and the accompanying *Technical Specifications* are particularly relevant. » (Archives New Zealand, 2009: 17)

Ces indications d’ordre général sont, dans certains textes, agrémentées par des exemples de mise en œuvre et des explications d’ordre technique. Ainsi, certaines politiques indiquent, de manière générale, la liste des documents Web à archiver, les

techniques de leur capture et les outils pour les mettre en œuvre (ex. : NAA, 2001b; Public Record Office Victoria, 2008; Queensland State Archives, 2006a; TNA, 2009a). Les politiques du milieu universitaire sont plus précises puisqu'elles justifient le choix d'une stratégie d'archivage déjà en place et démontrent sa pertinence pour le contexte organisationnel dans lequel elles sont appliquées. À titre d'exemple, dans le plan de gestion des collections de l'University of Melbourne, on indique la fréquence des captures du site Web de même que les outils utilisés et ceux développés pour identifier les documents à capturer :

« The 'whole of domain' quarterly captures aim to provide a record of the University's 'web presence' as at the 1 Jan, 1 April, 1 July and 1 October each calendar year.

Other web pages are captured because they contain University corporate web records, as identified by the Enterprise Classification Scheme (ECS). Records Services currently maintains a spreadsheet that matches up various corporate University urls against specific functions/activities as listed from the ECS. » (University of Melbourne, 2008a)

4.1.2 Systèmes Web visés par les politiques

Les politiques d'archivage analysées visent l'ensemble des types de SIW, qu'ils soient statiques ou dynamiques. Ces politiques recommandent donc l'archivage des types de SIW suivants :

- Sites Web publics (n=33 politiques);
- Intranets (n=7 politiques);
- Extranets (n=5 politiques);
- Réseaux privés virtuels (n=1 politique);
- Sites de commerce électronique (n=1 politique).

Quelques politiques indiquent également le sort réservé aux sites Web inactifs, aux microsites et aux sites Web événementiels (ex. : Public Record Office Victoria, 2007). Dans une politique, les sites Web relatifs aux enquêtes publiques ont été exclus de l'archivage : « Websites generated by public inquiries will be accessioned separately and not as part of this project. » (TNA, 2006: 2). Puis, ils ont été intégrés au programme d'archivage dans une version plus récente de la politique : « The National

Archives may also wish to capture and preserve websites with a limited lifespan such as those of public inquiries or campaigns.» (TNA, 2011: 4). Les sites Web d'enquêtes publiques (ex. : site Web de la Commission Bouchard-Talylor) et ceux relatifs aux campagnes électorales doivent être préservés à cause de leur nature éphémère et parce qu'ils documentent des événements et des décisions qui ont un impact majeur sur les citoyens et le gouvernement. Ces sites sont des sources d'information et de témoignage de grande valeur.

Quelques politiques analysées (n=9) établissent l'analyse des risques comme cadre méthodologique pour évaluer la vulnérabilité redditionnelle d'un site Web et déterminer, en conséquence, la stratégie d'archivage la plus appropriée à chaque type de SIW (ex. : NARA, 2005; Queensland State Archives, 2006a). Lorsqu'un site Web devient la première source d'information pour les citoyens sur les activités et les services de l'organisme, ce dernier devrait mettre en place les mesures nécessaires pour s'assurer que l'information diffusée est vraie, mise à jour, bien gérée et archivée.

« It is recommended public authorities conduct a risk assessment to help determine the recordkeeping approach that will best meet their needs. Using a risk management framework will help public authorities to identify, assess and report on risks relating to their online resources and services, and to plan appropriate risk mitigation strategies, which may include improving recordkeeping activities. » (Queensland State Archives, 2006a: 38)

Rappelons que le modèle de gestion des risques vise, dans ce contexte, l'établissement des niveaux de risques associés au site Web d'une organisation et l'identification des mesures de gestion documentaire les plus appropriées pour chacun des niveaux de risque. Pour identifier et évaluer ces niveaux de risque, les politiques étudiées indiquent les facteurs devant être analysés, dont les plus importants sont les suivants :

- La nature du SIW (site Web statique, site Web dynamique);
- La nature des transactions/actions possibles sur le site Web (niveau d'interactivité);
- La fréquence de mise à jour du site Web;
- la visibilité publique du site Web.

Les politiques analysées ne proposent pas de solutions d'archivage particulières pour chacun de ces types de SIW, on y porte, toutefois, une attention particulière aux sites Web publics.

4.1.3 Intervenants impliqués dans le processus d'archivage du Web organisationnel

Les politiques analysées proviennent de milieux divers et de cultures archivistiques différentes de celle de l'archivistique québécoise qui prône une gestion intégrée des documents et une continuité des actions entre les gestionnaires des documents et les archivistes. La distinction entre ces deux professions n'est pas retenue dans le contexte de la présente recherche. Aussi, nous avons choisi de ne pas mentionner les différentes dénominations des fonctions assumées par les intervenants identifiés dans les politiques analysées (ex. : archiviste, gestionnaire des documents, webmestre, édimestre). Les intervenants sont identifiés selon leur champ d'intervention (ex. : archiviste, webmestre).

Cinq catégories d'intervenants clés susceptibles d'être impliqués dans le processus de gestion et d'archivage d'un site Web sont identifiées dans les politiques analysées, soit :

1. Les gestionnaires du site Web
 - a. Les webmestres de l'organisation (n=11 politiques)
 - b. Les contractants (ex. hébergeur externe du site Web organisationnel) (n=1 politique)
2. Les archivistes (n=11 politiques)
3. Les gestionnaires de contenus
 - a. Les employés de l'organisation (n=6 politiques)
 - b. Les utilisateurs du site Web (cas des sites Web interactifs et des documents transactionnels) (n=2 politiques)
4. Le personnel responsable de la gestion de l'infrastructure technologique et communicationnelle de l'organisation (n=3 politiques)
5. Les services juridiques de l'organisation (n=2 politique)

6. Les archives nationales ou autres institutions archivistiques (archives provinciales, service d'archives) (n=3 politiques)

Même si elles identifient les intervenants devant prendre en considération les mesures qui y sont indiquées, les politiques analysées ne précisent que rarement les rôles et responsabilités associés à chacune de ces catégories d'intervenants. En effet, seules cinq politiques précisent les rôles et responsabilités des différents intervenants au niveau d'un site Web. Nous les présentons dans le Tableau X (p. 122). Il est évident que les rôles et responsabilités en lien avec la gestion et l'archivage des sites Web d'une organisation se définissent aussi en fonction des besoins, des ressources et des moyens en place.

Tableau X : Rôles et responsabilités en matière de gestion et d'archivage des sites Web.

INTERVENANTS	RÔLES ET RESPONSABILITÉS
Les gestionnaires de sites Web	<ul style="list-style-type: none"> – Signaler au service des archives les pages Web à capturer – Appliquer les directives de la politique d'archivage du Web organisationnel les concernant (Melbourne University, 2010) – Gérer les aspects informatiques des sites Web – Entretenir les sites Web – Administrer les droits d'affichage des gestionnaires de contenus et les aider à déterminer le format le mieux adapté aux documents affichés – Éditer, préparer et afficher les documents – S'assurer que les métadonnées pertinentes soient saisies (BAC, 1999)
Les archivistes	<ul style="list-style-type: none"> – Développer et implanter les politiques, procédures et outils archivistiques pour l'archivage du Web organisationnel (Melbourne University, 2010; NARA, 2005) – Surveiller et coordonner les activités de gestion documentaire du site Web (BAC, 1999) – S'impliquer dans les activités de conception, développement et migration du site Web vers d'autres plateformes – Préparer et veiller à l'application des délais de conservation des pages Web (Public Record Office, 2001; NAA, 2001)
Les gestionnaires de contenus	<ul style="list-style-type: none"> – Créer et gérer le contenu informationnel affiché sur le site Web (BAC, 1999; NARA, 2005) – S'assurer que le contenu des pages affichées et de leurs descriptions (métadonnées) soit exact et complet (BAC, 1999)

INTERVENANTS	RÔLES ET RESPONSABILITÉS
Les gestionnaires de l'infrastructure technologique et communicationnelle	<ul style="list-style-type: none"> – Gérer le cheminement de l'information vers les sites Web et assurer le contrôle de la qualité – Décider de l'affichage et du retrait des informations/documents sur le site Web (BAC, 1999)

D'après ce tableau, la responsabilité d'archivage d'un site devrait être partagée entre les différents acteurs organisationnels identifiés :

« Managing web records correctly is a shared responsibility extending to not only CEOs and records managers, but those stakeholders who create the web content and maintain the structure of the website. There is also the further responsibility of ensuring that external contractor provision of website management or additional website work including content and graphics creation, technical and other web support services are appropriately recorded. » (NSW, 2009)

Le mode de gestion et l'archivage du site Web doivent être définis de manière collaborative en réunissant tous les intervenants concernés. Les services juridiques de l'organisation doivent être consultés afin de s'assurer que les mesures qui découlent de l'application de la politique d'archivage sont en conformité avec les lois et règlements et protègent les intérêts de l'organisation (ARMA International, 2009: 11; Public Record Office Victoria, 2008: 15).

Les politiques rappellent ainsi l'importance de prévoir et de planifier une stratégie d'archivage au tout début du cycle de vie du site Web afin d'éviter une perte irréversible de contenus à valeur permanente et, par la même occasion, contrôler le mode de création et de gestion des documents Web afin de faciliter le processus d'archivage et de garantir la création d'archives Web complètes, authentiques et exploitables sur le long terme. À titre d'exemple, on met souvent l'accent dans ces textes sur l'intérêt d'associer des métadonnées pertinentes et complètes aux documents créés et d'adopter des formats de diffusion normalisés.

Le corpus analysé comprend un nombre important de politiques d'archivage du Web organisationnel développées par des archives nationales et des archives provinciales (n=26 textes analysés). Selon ces documents, les archives nationales ont pour responsabilité le développement de politique nationale pour l'archivage et la

préservation des archives Web (NAA, 2001; NARA, 2005). C'est à ces institutions que revient également la responsabilité d'identifier les frontières entre document d'archives et publication dans un environnement Web (Archives New Zealand, 2009) et d'orienter les organisations vers les stratégies les plus pertinentes qui soient en cohérence avec celles préconisées au plan national. À cet effet, nous avons noté, lors de l'analyse des principales initiatives d'archivage du Web (voir Annexe 1, p. i), que certaines archives nationales ont entrepris des projets d'archivage du Web gouvernemental (ex. : BAC, 2007-2008; TNA, [s.d.]a). Ces projets se basent principalement sur la capture de portraits annuels de sites Web gouvernementaux. Il serait ainsi important de coordonner, par le biais d'une politique d'archivage, les actions et de spécifier les rôles et responsabilités que partagent les archives nationales et les organisations à ce sujet.

4.1.4 Unités d'archivage et critères de sélection

Malgré son importance, la distinction entre publication et document d'archives dans un environnement Web n'a pas été considérée dans toutes les politiques analysées. Les seules qui le font (n=2) ne réussissent pas à établir de manière catégorique les frontières entre ces deux familles de documents. Il y a toujours une zone d'ombre où la différence entre les deux n'est pas bien définie, ce qui n'aide pas à trancher lorsqu'il s'agit d'établir les rôles et responsabilités conséquents (Archives New Zealand, 2009; Queensland State Archives, 2006b). Une autre politique s'attarde sur cet aspect et offre, à travers des exemples précis et des définitions fondées sur des textes réglementaires, une distinction claire entre publication et document d'archives (NAA et NLA, 2002). Les National Archives of Australia et la National Library of Australia sont les deux seules institutions qui disposent d'un cadre de référence quant au partage des responsabilités liées à l'archivage du Web (NAA et NLA, 2002: 5).

Deux autres politiques présentent une liste de questions à se poser pour décider de la nature archivistique d'un document Web (ARMA International, 2009: 2;

Queensland State Archives, 2008: 1). Nous citons celle des Queensland State Archives qui est plus directe et concise :

« The following five questions provide a quick and simple checklist to identify whether something is a public record:

- Was it made, sent or received in the course of business?
- Does it document a decision taken?
- Does it document advice given?
- Does it document the process of arriving at a decision?
- Is it required to be kept by legislation? » (Queensland State Archives, 2008: 1)

D'autres institutions archivistiques considèrent la provenance du site Web comme critère qui permet de délimiter leur champ d'intervention sur le Web :

« Websites generated by non-government organisations may be selected and preserved by other bodies, such as the British Library, the Wellcome Trust or the Higher Education Funding Council for England. » (TNA, 2006: 1)

The National Archives (Royaume-Uni) a choisi de retenir uniquement les sites Web gouvernementaux, plus précisément les sites Web du gouvernement central dans son champ d'intervention :

« The National Archives has now taken the decision to archive all websites of the central government web estate in order to provide continued access to key government documents through links persistence and also because of the huge increase in the significance of the web in enabling government to carry out its business. » (TNA, 2011: 3)

Les institutions ayant implanté un programme d'archivage de leur site Web public ont dû établir les frontières de leur site Web afin de s'assurer de la collecte de l'ensemble des documents relatifs aux activités de l'organisation et de pouvoir paramétrer le robot de collecte (ex. : University of Toronto) :

« Any page, whether on a University server or not, which contains information published on behalf of the University of Toronto and is linked into an official university home page collection is a University of Toronto page and is considered to be part of the overall University of Toronto web site and its considered to be a University of Toronto web page. » (University of Toronto, 2006)

Selon les politiques analysées, les sites Web, comme tout autre document d'archives, doivent être gérés afin d'assurer la continuité des affaires de l'organisation, documenter ses activités et respecter les exigences légales :

« Records of websites and business records created and published on agency websites must be saved into a recordkeeping system to provide evidence of business activity and meet legal requirements. » (Archives Office of Tasmania, 2009)

La définition des archives Web dans les politiques d'archivage analysées se présente en une liste de composantes Web à archiver. Les deux définitions¹⁸ qui suivent nous permettent d'établir une liste complète de ce qui constitue les archives Web :

« Web records can be defined as records generated, captured or received using web technologies. Web records on an organisation's website can generally fall into three categories 1) web content records, including introductory or descriptive information about the organisation, its activities or particular areas it manages, but also publications that it makes available on the web, 2) inputs/outputs of transactions performed via the web e.g. data submitted via online forms and 3) website administration records, such as statistics, diagnostics etc, which provide evidence of website operations and management. » (NSW, 2009a)

« Web site-related records include (1) web content records, which represent information presented on a web site, and (2) web site administrative records, which provide evidence of the management and operations of the web site. » (NARA, 2005: 5)

Les archives Web sont donc formées de la liste suivante de documents :

1. Documents diffusés sur le site Web (le site Web au complet, certaines sections du site Web ou une sélection de documents Web) (n=33 politiques);
2. Documents relatifs aux transactions réalisées grâce au site Web (n=24 politiques);
3. Documents administratifs relatifs à la gestion du site Web (n=19 politiques).

¹⁸ Il s'agit des deux définitions les plus exhaustives que nous avons pu repérer lors de l'analyse des politiques d'archivage.

Nous pouvons, dès lors, distinguer trois niveaux d'archivage, soit : un macro-archivage qui vise la collecte de l'ensemble du site Web, un micro-archivage où l'on vise la collecte d'une section particulière ou des documents spécifiques sur le Web et un méta-archivage qui cible plutôt les documents de gestion du Web. Ces niveaux d'archivage sont présentés dans le Tableau XI (p. 127) tout en indiquant les composantes Web qui y sont associées.

Tableau XI : Unités d'archivage d'un site Web organisationnel par niveau d'archivage.

NIVEAU D'ARCHIVAGE	UNITÉS D'ARCHIVAGE
Macro-archivage	– Site Web entier
Micro-archivage	<ul style="list-style-type: none"> – Sections du site Web Documents diffusés sur le site Web, – Documents relatifs aux transactions Web, – Traces de toute transaction réalisée en ligne, – Journaux historiques des activités (date et heure de l'événement, information sur l'utilisateur, ressources ou services utilisés, ressources ou services retournés à l'utilisateur).
Méta-archivage	<ul style="list-style-type: none"> – Documents relatifs à la gestion du site Web, – Instructions/politiques relatives à la publication et à la gestion des documents Web, – Documents relatifs à l'ajout, à la modification et au retrait des ressources et services du site Web, – Documents relatifs aux éléments visuels du site Web, – Contrats de propriété intellectuelle et documents relatifs aux conditions d'accès et d'utilisation des informations disponibles sur le site Web, – Documents relatifs au fonctionnement du site Web, – Documents relatifs à la technologie utilisée pour le développement du SIW, – Informations relatives aux procédures d'authentification, – Informations relatives à l'utilisation des outils de recherche sur le site (outil consulté, fonctionnalités de recherche, date d'utilisation, requêtes, résultats de la recherche).

Chaque site Web est particulier de par sa technologie de développement et ses finalités. Différents types de documents peuvent en découler :

« Websites serve a variety of purposes for public offices ranging from delivery of basic information to the ability to download forms and

complete transactions online seamlessly. For every purpose the web is used for, a different record may be generated. » NSW, 2009

Les politiques analysées n'indiquent pas avec précision quels sont les unités d'archivage et les niveaux d'archivage à considérer. Elles recommandent à cet effet de procéder à une analyse des risques associés à l'usage du site Web. Les résultats de cette démarche vont préciser les documents devant être capturés :

« A web record may be any information in whole or part that appears on a website and provides evidence of business activity. In some situations, an entire website may be managed as a record. However in most instances, a record will be an element. For example a page, content item, image, document, submitted form, or log entry recording an action taken. » (Archives New Zealand, 2009: 7)

Les documents générés doivent être évalués et intégrés au calendrier de conservation afin de gérer leur cycle de vie (quand les capturer, dans quel format/support, à quelle fréquence, pour combien de temps et à quelles fins). À ce niveau, ils sont soumis aux mêmes règles de gestion que les autres documents d'archives :

« Records of websites, whether they are publications or transactional records of online business processes, are subject to the same disposal rules as other official records of your organisation. » (NSW, 2009a)

4.1.5 Stratégies d'archivage et principes de base

Les stratégies d'archivage varient selon le type d'organisme qui les établit. Dans les milieux où la politique vient encadrer un projet en cours, les stratégies sont précises et justifiées en fonction du contexte d'usage, des ressources et des besoins de l'organisation. Par exemple, The University of Melbourne a choisi d'archiver de manière fréquente la page d'accueil de son site Web et de produire quatre portraits annuels de ce dernier. Elle n'a pas développé une application maison pour réaliser ces captures. Elle a choisi d'utiliser les outils développés par Internet Archive. Lorsqu'il s'agit de politiques d'archivage du Web proposées par des archives nationales (n=15), ces dernières présentent l'éventail de stratégies existantes. Elles décrivent leurs avantages et inconvénients et les scénarios possibles menant aux choix d'une

combinaison d'approches et de techniques d'archivage en fonction de la nature du site Web à archiver, des risques qu'il présente et des exigences auxquelles est soumise l'organisation (ex. : NAA, 2001b; Archives New Zealand, 2006a; Public Record Office Victoria, 2008). Les politiques suggèrent ainsi d'adopter une combinaison des approches suivantes :

1. Archivage en ligne : mise en ligne d'une copie du site archivé à une date antérieure avec une copie de sauvegarde du site sur le serveur (ou dépositaire local du site Web). Cette approche n'a pas été identifiée dans la revue de littérature. Elle a émergé des données de recherche;
2. Approche orientée objet : capture périodique du site Web ou des documents sélectionnés sur le site Web ainsi que les métadonnées qui y sont assignées avec suivi des changements apportés au site Web et tenue d'un journal historique des activités ou des transactions;
3. Approche orientée événement : capture des événements ou des transactions ayant lieu entre le site Web et un utilisateur. Il s'agit de documenter la date et l'heure de l'événement, l'adresse IP depuis laquelle la transaction a été réalisée, le profil de l'utilisateur, la requête ou l'action réalisée, les ressources utilisées et ayant servi à l'utilisateur ainsi que les métadonnées associées à la transaction et aux documents produits.

Les politiques d'archivage proposées par des archives nationales rappellent aussi les principes archivistiques devant être respectés, sans quoi l'intégrité des archives capturées et leur complétude ne seraient pas garanties. Pour opérationnaliser ces principes, des mesures d'archivage peuvent porter sur l'ensemble du site Web (macro-archivage), viser des sections ou des pages particulières du site Web (micro-archivage), ou cibler des documents de gestion du site Web (méta-archivage). Les actions suivantes ont été recommandées.

Macro-archivage des sites Web :

- Capture de l'ensemble du site Web à des intervalles prédéterminés;
- Conservation des anciennes versions du site (si elles existent) et des portraits réalisés;

- Sauvegarde d'une copie en ligne du site archivé (archives en ligne du site).

Micro-archivage des sites Web :

- Sauvegarde périodique de la carte du site;
- Capture des documents « signifiants » au fur et à mesure de leur création dans un système d'archivage;
- Conservation d'une copie des formulaires en ligne;
- Capture des transactions depuis le serveur.

Méta-archivage des sites Web :

- Production périodique et archivage du rapport consignait les changements apportés au site Web incluant le nom des auteurs du changement et la date;
- Conservation des courriels communiqués à un utilisateur suite à l'utilisation d'une application;
- Sauvegarde du rapport produit par une base de données suite à l'entrée de données en ligne par un utilisateur;
- Périodiquement, documentation des technologies de développement du site, des procédures d'édition et de gestion des contenus, des guides et procédures de gestion de l'information diffusée sur le site Web.

Selon les politiques analysées, ces trois niveaux d'actions sont complémentaires. Nous n'avons pas pu confirmer cette assertion sur le terrain. Dans les milieux universitaires étudiés, par exemple, les organismes cumulent non seulement des captures périodiques de leur site Web, auxquelles nous n'avons pas eu accès, mais procèdent aussi à un archivage sélectif de pages Web particulières. Nous n'avons pas pu vérifier si les webmestres produisent une documentation Web qui, elle aussi, est archivée. Les webmestres interviewés ne documentent pas leurs tâches essentiellement par manque de temps. Nous n'avons pas pu confirmer, par conséquent, si les portraits des sites Web sont suffisants ou s'il faut archiver d'autres types de documents pour constituer des archives Web complètes et de qualité.

Pour opérationnaliser ces choix, l'ensemble des politiques analysées suggère une démarche systématique :

- Évaluer la vulnérabilité redditionnelle du site Web;
- Décrire et documenter les composantes d'un site Web grâce aux métadonnées;

- Documenter tout changement qui survient sur le site Web et ses différentes composantes (documentaires et technologiques);
- Utiliser des liens permanents;
- Assigner des délais de conservation aux archives Web;
- Intégrer les documents créés dans un système d'archivage.

Les politiques analysées mettent l'accent sur les critères de sélection des archives Web. En plus des besoins administratifs et des exigences légales, les archives Web sont également conservées pour leur valeur sociétale. L'approche d'évaluation des archives Web ne diffère donc pas de celle utilisée pour les archives en général :

« [...] records of websites can be appraised in the same way as any other record. PROV's Appraisal Policy suggests that the following factors be considered in determining a retention period:

- Business operation and administration needs
- Legislative and regulatory requirements
- Public administration, organisational and social accountability
- Social obligations of the Victorian public sector
- Social interests of the people and groups of Victoria
- Fundamental rights and entitlements (including human and cultural rights)
- Significance to Victoria. » (Public Record Office Victoria, 2007: 16)

L'évaluation archivistique d'un site Web ne se limite pas au contenu informationnel. Toutes les composantes d'un site Web doivent être considérées :

« The appraisal of web records requires all elements of a website be considered, including pages, content items, files, background processes, look and feel, automated or manual transactions, and evidence of business activity in creating and maintaining websites. » (Archives New Zealand, 2009: 11)

Au niveau des outils d'archivage, les politiques étudiées indiquent que la capture d'un site Web pourrait être réalisée à distance grâce à un logiciel robot (*crawler*) ou en faisant appel à un service d'archivage. À titre d'exemple, The Melbourne University et University of Toronto sont abonnées au service *Archive-It* d'Internet Archive pour la capture périodique de leur site Web public. L'archivage d'un site pourrait également être possible en exploitant les fonctionnalités d'un

système de gestion de contenus ou en intégrant les documents Web à un système d'archivage électronique, et ce, depuis leur création. Ces deux approches sont résumées dans la citation suivante :

« Approaches to choosing a system to manage web records include:
 - using a web publishing system, such as a content management system (CMS), as a recordkeeping system [...],
 - using an electronic recordkeeping system, such as an Electronic Document and Records Management System (EDRMS), to manage content published on the web. [...] » (New Zealand Archive, 2009: 9)

Les politiques analysées n'élaborent pas beaucoup sur les stratégies de préservation des sites Web archivés. On y rappelle seulement les problèmes d'obsolescence des technologies, la nécessité de planifier la migration périodique des documents vers d'autres supports, l'intérêt de choisir des formats normalisés pour la conception des sites Web organisationnels ou l'utilisation d'identifiants uniques et la documentation des procédures. Les solutions de pérennisation des sites Web restent très théoriques, comme l'indique cette citation :

« Normalisation (or format migration) means accessing a web record in a different format from that in which it is currently held. A normalisation or migration strategy ensures archived material remains available independent of the system that created it, yet important elements of the record, such as look and feel and context are retained. » (Archives New Zealand, 2009: 25)

En aucun cas les politiques analysées font référence à des normes telles que la norme ISO 28500 : 2009 qui définit le format de fichier WARC pour l'archivage de l'internet ou la norme ISO 14721 relative au modèle OAIS. Pareillement, les politiques d'archivage ne précisent pas toujours les principes d'édition des sites Web. Elles renvoient toutefois les personnes concernées aux guides des sites Web qui présentent des standards et des règles pour la création et la gestion des sites Web.

Au moment de la constitution du corpus de politiques à analyser, le phénomène du Web 2.0 n'était pas assez développé pour que l'on retrouve des politiques lui étant propres. Cet aspect n'est ainsi pas présent dans notre analyse, mais il est à noter que des politiques plus récentes abordent les problématiques liées à

l'archivage du Web 2.0 et en proposent des stratégies d'archivage (ARMA International, 2011; NARA, [s.d.]a; NSW, 2009b).

Les politiques analysées se distinguent les unes des autres sous plusieurs aspects. Alors que certaines portent sur des aspects théoriques liés aux archives Web, d'autres sont de nature plus technique. Quelques-unes rappellent les principes archivistiques pour la gestion des archives Web, plusieurs s'attardent sur les avantages et inconvénients des approches et techniques d'archivage propres aux sites Web. Le lecteur est souvent renvoyé à d'autres politiques et documents complémentaires pour pouvoir dresser un portrait complet de la démarche de gestion, de conservation et de préservation des sites Web.

Somme toute, les politiques analysées se complètent, mais elles n'offrent pas toutes les informations nécessaires sur l'ensemble du processus d'archivage d'un site Web. La dispersion des recommandations et des outils à utiliser ou à mettre en place ainsi que la focalisation sur les problématiques techniques n'offrent pas une vision globale du processus de gestion et d'archivage d'un site Web organisationnel.

En observant les forces et les faiblesses de chacune des politiques étudiées, et suite à l'observation des milieux, nous proposons dans le chapitre Discussion un modèle de politique type pour l'archivage des sites Web organisationnel qui présente les principes directeurs à suivre et la méthodologie d'analyse des risques à mettre en place afin d'identifier les documents à conserver ainsi que la méthode, la technique et la fréquence d'archivage avec des exemples d'outils à développer (ex. : règle de conservation spécifique, politique de gestion des URL).

En résumé, depuis les débuts du Web, les institutions archivistiques ont reconnu l'intérêt de bien gérer les documents issus des technologies, et ce, pour les besoins des organisations ou de la recherche. Plusieurs politiques d'archivage du Web sont diffusées en ligne. Nous les avons analysées dans l'objectif d'identifier les systèmes Web visés, les responsables de l'archivage des sites Web dans une organisation, l'unité d'archivage ainsi que les choix méthodologiques et techniques devant être établis.

L'archivage du Web organisationnel est une tâche partagée entre différents intervenants : webmestres, archivistes et tout autre acteur impliqué sur le plan technique ou de gestion d'un site Web organisationnel. Ces politiques rappellent la valeur archivistique d'un site Web organisationnel (valeur primaire et valeur secondaire ou sociétale) et identifient les documents et les composantes d'un site Web à archiver. L'archivage d'un site Web peut se réaliser à trois niveaux complémentaires : macro-archivage (l'ensemble du site Web), micro-archivage (certaines sections ou éléments particuliers du site Web) et méta-archivage (documents relatifs à la gestion et à la maintenance du site Web). Les unités d'archivage sont déterminées grâce à une démarche d'analyse des risques. Les documents devant être archivés afin de témoigner des activités en ligne d'une organisation doivent être intégrés dans un calendrier de conservation afin de gérer leur cycle de vie.

4.2 Observation des archives Web

La deuxième question de recherche s'intéresse aux caractéristiques des archives Web. Nous avons observé les sites Web publics de 11 organismes québécois dans l'objectif de dresser un portrait de la typologie documentaire d'un site Web public et de comprendre la nature des archives dans un tel environnement.

La nature des pages Web observées a été étudiée grâce à une grille d'analyse (voir Annexe 4, p. xxv). Nous avons analysé au total 737 documents. Le nombre de documents analysés varie d'un site Web à l'autre puisqu'il dépend de la taille et du niveau de profondeur du site Web observé de même que de la nature des documents aspirés. Nous avons également étudié la carte (ou le plan) des sites Web observés, car ce document permet d'avoir un aperçu plus complet du site et de sa structure telle que définie par l'organisme créateur. L'arborescence créée par l'aspirateur Web utilisé (Scrapbook) a été également analysée puisqu'elle nous permet d'identifier visuellement et d'analyser les formats et les types des documents aspirés par niveau de profondeur (voir Annexe 5, p. xxxi).

Afin de déterminer et de valider les catégories et les types des pages Web capturées, nous avons consulté les plans de classification institutionnels des organismes à l'étude, lorsqu'ils sont diffusés en ligne (n=10). En effet, les subdivisions d'un plan de classification sont déterminées en se basant sur les activités de gestion communes à tous les organismes (ex. : gestion des finances) et sur les activités qui soutiennent la mission propre d'un organisme (ex. : enseignement). Ces subdivisions nous aident à identifier les documents de gestion (documents relatifs aux activités de gestion) et les documents de fonction (documents liés à la mission propre de l'organisme). En vertu du *Règlement sur la diffusion de l'information et sur la protection des renseignements personnels*, les organismes publics au Québec visés par ce règlement doivent diffuser sur leur site Web 15 classes de documents/renseignements dont le plan de classification institutionnel (c. A-2.1, r. 2, art. 4)¹⁹. De même, nous avons eu recours aux calendriers de conservation de ces organismes afin de déterminer la valeur archivistique des documents aspirés ou du site Web, si indiquée dans ces outils. Pour l'organisme n'ayant pas inclus un plan de classification ou un calendrier de conservation sur son site Web, un plan de classification (ou un calendrier de conservation) d'un autre organisme aux mandats similaires a été utilisé (Chebbi et Cherif, 2010).

Les caractéristiques techniques des documents diffusés sur les sites Web organisationnels, les valeurs archivistiques qu'ils représentent ainsi que les conditions de validité qu'ils rencontrent sont présentées dans la première section. Les sections qui suivront font état des différents modes de création et de la typologie documentaire d'un site Web organisationnel.

4.2.1 Caractéristiques des archives Web

Rappelons tout d'abord que les archives Web sont l'ensemble des documents issus ou relatifs à un SIW, produits dans le cadre des activités d'un organisme, y

¹⁹ Ce règlement ne s'applique pas à un organisme professionnel.

compris le site Web dans sa globalité, les documents diffusés sur ce site ainsi que la documentation relative à sa gestion.

Les politiques d'archivage du Web organisationnel analysées dans la section 4.1 de ce chapitre affirment la valeur archivistique du Web organisationnel, établissent les unités d'archivage et proposent des approches et techniques d'archivage pour le Web organisationnel. Toutefois, ces politiques ne mentionnent pas l'intérêt de considérer la nature propre des archives Web comme un facteur déterminant lors du choix d'une stratégie d'archivage. De même, il ressort des entrevues réalisées dans des milieux organisationnels au Québec que la copie de sauvegarde ou la réalisation périodique d'un portrait du site Web public de l'organisation (selon la technique d'archivage à la source ou la technique d'archivage à distance) sont les deux principales approches d'archivage du Web organisationnel, approches souvent choisies uniquement en fonction du cycle de vie du site Web (mises à jour importantes ou refonte). Or, il nous paraît important de considérer la nature propre des documents diffusés ou produits sur ces sites Web publics ce qui complète, de notre point de vue, la liste des facteurs à prendre en considération lors de l'analyse des risques associés à un site Web et facilite, par conséquent, la définition de l'approche et de la technique d'archivage les plus appropriées.

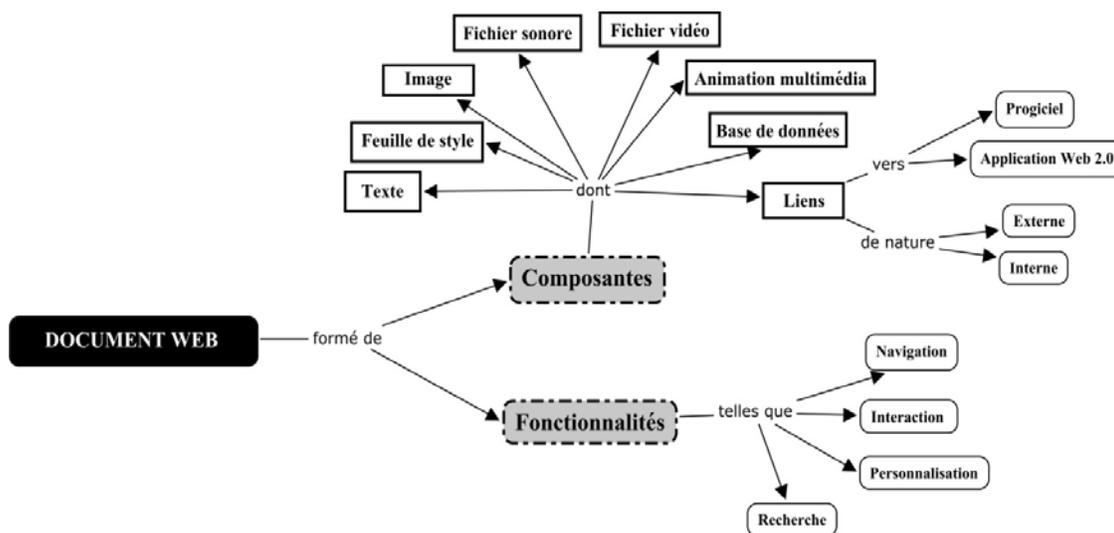
L'observation des caractéristiques des sites Web publics de 11 organismes au Québec et des types de documents qui les composent constitue une démarche dans laquelle nous nous sommes engagée afin de mieux comprendre les caractéristiques des archives Web et d'apporter des précisions au modèle de politique pour l'archivage du Web organisationnel.

4.2.1.1 Composantes et fonctionnalités d'une page Web

N'ayant pas accès aux serveurs des sites Web observés, l'analyse des caractéristiques techniques des documents a été faite uniquement à partir des pages Web affichées du côté client. Ainsi, pour les sites dynamiques où des contenus sont produits par la compilation, sur le serveur, de pages dynamiques, seul le résultat de la compilation est affiché et accessible du côté client.

Suite à l'analyse des pages Web sélectionnées, nous remarquons donc qu'un document natif Web (page HTML, fichier multimédia) se distingue d'un document traditionnel par ses composantes et ses fonctionnalités (ex. : un document numérisé ou produit en format PDF diffusé en ligne). En effet, un document Web se distingue des autres documents par un contenu informationnel protéiforme et de nombreuses fonctionnalités Web (voir Figure 5, p. 137).

Figure 5 : Composantes et fonctionnalités d'un document Web.



Le contenu informationnel des documents Web observés peut être composé de texte, d'images, de sons, de vidéos, d'autres objets multimédias et de liens hypertextuels. Ce contenu se présente sous différents formats, les principaux étant HTML, XML, DOC, PDF. Les images sont en format JPEG, GIF, PNG, etc. Nous avons également identifié des fichiers sonores en format MP3 et des animations multimédias en format SWF. Une même page Web en format HTML, par exemple, peut donc intégrer ces différents formats.

Un document Web peut renvoyer l'internaute, par le biais de liens URL qu'il intègre, vers des sections dans le corps même du document ou vers un autre document ou section du site Web. Nous avons noté que certains liens conduisent au téléchargement de plugiciels (ex. : Acrobat Reader) ou renvoient également vers des applications Web 2.0 telles que :

- Facebook (n=4 sites)
- Twitter (n=4 sites)
- YouTube (n=3 sites)
- Blogue (n=2 sites)
- Delicious (n=2 sites)
- Flickr (n=1 site)
- Fichier balado (n=1 site)

Une portion de la production documentaire des organismes étudiés dans le cadre de cette recherche se retrouve donc sur le Web 2.0. Les applications adoptées n'offrent pas toujours à l'organisme créateur un contrôle total sur les contenus qui y sont créés ou déposés (ex. Flickr ou YouTube).

Les sites Web analysés ont une identité visuelle et une structure stable. Ces sites sont produits dans le respect des normes d'édition Web proposées dans les politiques d'édition Web de l'organisme concerné ou celles recommandées dans les standards internationaux (aspect visuel et ergonomique uniforme, règles de nommage des fichiers et adresses URL, choix des formats de fichier, etc.). La mise en forme du contenu informationnel, plus précisément celui généré en format HTML, est gérée indépendamment du contenu informationnel grâce à l'usage de feuilles de style.

Tel que schématisé dans la Figure 5 (p. 137), des fonctionnalités (ex. : navigation, personnalisation) se greffent également aux pages Web observées. Il est possible de naviguer dans un document Web ou à partir d'un document Web vers un autre, grâce à un ensemble organisé de moyens pour se déplacer dans un site (ex. : menus, icônes de navigation). Les documents Web étudiés comportent un moteur de recherche qui permet d'effectuer une recherche dans les pages d'un site ou dans une base de données connectée au site. Des demandes d'inscription ou d'abonnement, des formulaires, des jeux-questionnaires peuvent être remplis en ligne. L'internaute peut aussi interagir avec le webmestre ou le responsable d'une activité. Finalement, nous avons repéré quelques fonctionnalités de personnalisation du site Web telles que : un changement de la taille de la police, le choix d'un style de présentation.

L'accès aux composantes et fonctionnalités d'un document Web dépend d'une myriade de programmes et de l'infrastructure technologique en place. L'exécution de ces programmes et la gestion de l'infrastructure technologique génèrent des documents qui préservent la trace des opérations de gestion et des actions/transactions ayant lieu grâce à l'utilisation du site Web (ex. : journal historique, cookies, journal des transactions, etc.). Ces documents de gestion ainsi que les composantes logiques d'un site Web (ex. programmes, scripts) doivent être archivés. Ces documents sont nécessaires non seulement pour comprendre l'évolution du site Web lui-même, mais aussi pour contextualiser les documents qui en découlent, garantir leurs fiabilité, intégrité, authenticité et exploitabilité.

En somme, un site Web organisationnel se caractérise non seulement par la variété des composantes et fonctionnalités des pages Web, mais aussi par la richesse du tissu de liens qui le forme.

4.2.1.2 Valeurs archivistiques et critères de validité

Nous examinerons dans cette section les valeurs que pourraient avoir les archives Web, si ces dernières satisfont aux conditions de validité d'un document d'archives, et les moyens à utiliser pour déterminer et maintenir ces valeurs.

4.2.1.2.1 Valeurs archivistiques

Suite à la recension des écrits et à l'analyse des politiques d'archivage du Web organisationnel, nous avons noté trois niveaux d'archivage d'un site Web organisationnel, soit le site Web lui-même (macro-archivage), les documents qui composent le site Web (micro-archivage) ou ceux qui servent à sa gestion (méta-archivage). Dans quelques milieux étudiés, les règles de conservation que nous avons pu recueillir se rapportent à un niveau ou à l'autre. Un seul cas tient compte des deux premiers niveaux d'archivage recommandés dans plusieurs politiques d'archivage analysées dans la section 4.1 (p. 115).

Dans cette section, nous analysons, dans un premier temps, les valeurs archivistiques que peut présenter le site Web lui-même et les documents produits ou

diffusés grâce à un site Web organisationnel (les valeurs archivistiques d'un document ont été définies dans la section 2.2.2.3 de la revue de littérature, p. 39). Dans un deuxième temps, nous nous interrogeons si ces documents satisfont les conditions d'authenticité, de fiabilité, d'intégrité et d'exploitabilité indiquées dans la norme ISO 15489-1 (ISO, 2001a : 7-8). Cette analyse nous permet, par l'étude du respect des critères de validité, de mieux comprendre la valeur archivistique d'un site Web organisationnel.

Un document produit ou reçu dans le cadre d'une activité possède, dès sa création, une valeur archivistique. Cette valeur se décline en valeur primaire et valeur secondaire. La valeur primaire est la raison d'être du document. C'est une qualité de chaque document produit ou reçu à des fins administratives, légales, financières ou probatoires. La valeur secondaire est une qualité que peut avoir un document d'archives en fonction des informations de portée scientifique et historique qu'il contient. Ces valeurs évoluent tout au long du cycle de vie du document et sont gérées grâce au calendrier de conservation. Cet outil détermine, pour chaque période du cycle de vie du document, sa valeur et son sort final. Nous souhaitons donc comprendre comment on peut évaluer la valeur archivistique d'un objet documentaire aussi composite et complexe qu'un site Web organisationnel.

Afin de déterminer la valeur archivistique d'un site Web, nous avons procédé à deux niveaux d'analyse :

- Macro-analyse : analyse des sites Web observés dans leur globalité à la lumière des règles de conservation recueillies lors des entrevues et des règles de conservation recommandées dans les politiques d'archivage du Web analysées dans le cadre de cette recherche, mais également celles publiées par la CREPUQ (2002) et BAnQ (2011).
- Micro-analyse : analyse des types de documents et des différents éléments qui composent le site Web en ayant recours aux calendriers de conservation des milieux à l'étude. Lorsque ces calendriers ne sont pas accessibles en ligne, nous examinons les calendriers d'autres milieux similaires.

Dans le cadre de cette étude, nous n'avons pas pu analyser les documents relatifs à la gestion des opérations Web (méta-archivage) pour deux principales raisons. D'abord, l'observation a eu lieu du côté client des systèmes étudiés. Ensuite,

nous avons obtenu que très peu d'exemples de ces documents lors des entrevues et, selon les webmestres rencontrés, les pratiques de gestion d'un site Web ne semblent pas toujours bien documentées.

Nous avons appliqué cette démarche aux sites Web des institutions d'enseignement et des services d'archives universitaires (n=5). Nous nous sommes limitée à ces deux milieux afin de respecter les conditions de confidentialité et d'anonymat relatives à cette recherche puisque les types de documents propres aux autres milieux permettraient d'identifier l'organisme participant à cette recherche.

Les règles de conservation recueillies dans les milieux étudiés ne couvrent pas tout le cycle de vie du site Web. Par exemple, certaines règles de conservation étudiées s'apparentent à une procédure de copie de sauvegarde (exemple du SIW02). En effet, ces règles spécifient le type des documents à capturer, le moment de leur capture, le support de conservation, mais n'indiquent pas le sort final des documents capturés. En outre, elles ne couvrent pas toujours les documents relatifs à la gestion du site Web ou ses composantes technologiques ou encore ne prennent pas en considération le site Web lui-même. La date de réalisation des captures, de même que leur fréquence, sont fixées par convenance ou en se basant sur les dates de réalisation de certaines activités administratives (début et fin de l'année universitaire pour les sites académiques, début et fin de l'année administrative pour les autres sites). Nous nous interrogeons si ces choix respectent toujours les exigences légales et les besoins administratifs.

Quant aux délais de conservation, ils sont variables d'un contexte à un autre et diffèrent selon l'unité d'archivage considérée. Le site Web demeure actif jusqu'à son remplacement par une autre version. On recommande, en parallèle, de réaliser des portraits annuels du site Web alors qu'il est encore actif ou lorsqu'il couvre un événement ponctuel. Ces portraits ne font pas l'objet de règle spécifique, ce qui explique l'accumulation d'un nombre de plus en plus grand de ces documents dans les organisations qui procèdent à l'archivage de leur site Web public selon cette méthode.

L'exemplaire principal des documents de gestion du site Web est conservé à l'état actif aussi longtemps que le site Web existe. Ensuite, les documents les plus représentatifs de l'évolution du site Web sont versés aux archives définitives. Les documents de support sont éliminés (ex. : les anciennes versions des différentes sections d'information du site Web).

La plupart des règles étudiées découlent d'une gestion centralisée du processus d'archivage du site Web public de l'organisation même si différents services ou unités sont responsables de la diffusion de leurs propres contenus. Une seule règle est appliquée dans un contexte de gestion décentralisée du processus d'archivage. Les unités organisationnelles ont à ce niveau la responsabilité de définir la période et la fréquence de capture de leur propre site Web. Le support papier est indiqué pour les documents de gestion et le support numérique est suggéré pour les portraits des sites Web.

L'analyse démontre aussi qu'on ne prend pas en considération la nature de certains types de documents, notamment ceux propres à un environnement Web ou de nature nominative, devant être gérés par des règles d'édition et de conservation spécifiques (ex. : page d'accueil du site Web, pages faisant l'objet de mises à jour fréquentes, pages présentant des données de nature sensible qui exposent l'organisme à des poursuites judiciaires). Les politiques analysées recommandent d'inventorier ce type de documents et de leur appliquer des mesures de gestion et d'archivage bien particulières (ex. : interdire la collecte de ces documents, les exclure du champ de collecte des robots moissonneurs, les capturer de manière fréquente).

Afin d'étudier l'adéquation des règles de conservation à la nature et aux composantes d'un site Web organisationnel, nous présentons, dans le Tableau XII (p. 143), la valeur archivistique de quelques types de documents que nous retrouvons sur un site Web académique et celui d'un service d'archives universitaire auxquels s'appliquent une même règle de conservation. Ces deux sites sont sous le contrôle du même organisme (O03), toutefois, leur mode de gestion et d'archivage est décentralisé. Chacune des unités organisationnelles est responsable de la définition d'une stratégie d'archivage de son site Web qui soit cohérente et qui découle de la

règle générale suggérée à toutes les unités. Il est à noter qu'il existe, pour les deux cas, une même règle pour les documents relatifs à la gestion du site Web.

Des règles spécifiques aux documents en ligne n'existent pas pour ces deux milieux. L'appartenance à une famille de document est déterminée par analogie avec des documents répondant à une même fonction ou activité (ex. : un blogue étudiant est produit dans le cadre des activités étudiantes). Quant aux types de documents en particulier, nous avons appliqué les règles préparées pour des documents sur support papier à la version Web de ces derniers. Ces règles sont appliquées dans les milieux étudiés aux documents sur support papier et aux documents numériques bureautiques.

Tableau XII : Exemples de valeurs archivistiques d'un site Web public et de ses composantes.

TYPE DE DOCUMENT	FAMILLE DE DOCUMENT	VALEUR PRIMAIRE	SORT FINAL/VALEUR SECONDAIRE
Règle pour un site Web universitaire (n=2 sites)			
Site Web	Ressources informationnelles	X ²⁰	X ²¹ Précision : Élimination des anciennes versions. Conservation d'une liste des sites Web de l'organisme et des documents représentatifs de la création, de l'évolution de la structure et des éléments majeurs du site Web.
Règles pour des documents sur support papier (n=8 documents)			
Présentation d'un espace de travail	Moyens et outils d'enseignement	X	X Précision : Certains documents sont conservés à des fins de conservation permanente
Affichage des charges de cours	Gestion des ressources humaines	X	Élimination

²⁰ Document ayant une valeur primaire.

²¹ Document ayant une valeur secondaire.

TYPE DE DOCUMENT	FAMILLE DE DOCUMENT	VALEUR PRIMAIRE	SORT FINAL/VALEUR SECONDAIRE
Affiches, dépliants	Communication	X	X Précision : Certains documents sont destinés à des fins de conservation permanente
Programmes d'études	Programme d'enseignement	X Précision : Document en vigueur ou jusqu'à son remplacement par une nouvelle version	Élimination
Règles en vigueur dans un espace de travail commun	Règles et politiques		X Conservation
État général des fonds	Ressources informationnelles	X Précision : il existe une règle pour l'équivalent papier, mais elle n'est pas adaptée à la version Web	X Précision : il existe une règle pour l'équivalent papier, mais elle n'est pas adaptée à la version Web
Expositions virtuelles	Ressources informationnelles	- Document à évaluer – (pas de règle pour ce type de document) Il n'existe pas de règle pour l'équivalent papier. Il existe une règle qui concerne les documents de gestion des activités et produits liés à la diffusion et exploitation des ressources informationnelles.	- Document à évaluer – (pas de règle pour ce type de document) Il n'existe pas de règle pour l'équivalent papier. Il existe une règle qui concerne les documents de gestion des activités et produits liés à la diffusion et exploitation des ressources informationnelles.
Blogue étudiant	Activités et associations étudiantes	X Précision : il existe une règle pour l'équivalent papier, mais elle n'est pas adaptée à la version Web	X Précision : il existe une règle pour l'équivalent papier, mais elle n'est pas adaptée à la version Web (élimination ou conservation dans les fonds d'archives privées des associations étudiantes)

TYPE DE DOCUMENT	FAMILLE DE DOCUMENT	VALEUR PRIMAIRE	SORT FINAL/VALEUR SECONDAIRE
			qui les ont produits)

Ainsi, malgré la présence d'une règle générale pour le site Web et de règles spécifiques pour certains types de documents qui ont été publiés sur le site Web, de nouveaux types de documents propres au site Web n'ont pas leur propre règle, ce qui nous amène à nous demander si la règle générale du site Web est pertinente pour ces documents Web ou s'il est nécessaire de prévoir des règles spécifiques pour ces documents. Un participant à cette recherche (R03) nous confirme qu'un des documents diffusés sur le site Web de son unité n'est plus produit en version papier et que la version Web est l'unique document complet et officiel dont dispose son service. À ce niveau, la règle de conservation établie pour l'équivalent de ce document dans un format papier ne nous semble pas encore pertinente pour la version Web (fréquence d'archivage, support de conservation, etc.). De même, les expositions virtuelles, même si elles sont produites à partir d'un matériel existant (photographies numérisées), sont également un produit documentaire original et unique. Nous n'avons pas retrouvé de règles spécifiques à ce type de document. Cet exemple démontre l'importance de procéder à une analyse fine des composantes d'un site Web, notamment des nouvelles formes documentaires qu'il génère, avant d'établir une stratégie d'archivage.

La capture du site Web à elle seule (macro-archivage) n'est pas suffisante. Dans certains cas (ex. : site Web dynamique), il faudrait également procéder à la capture d'éléments particuliers du site Web et/ou la documentation de ses composantes technologiques (technologies de développement, journaux historiques, plugiciels, scripts, etc.) à une fréquence et à des moments précis du cycle de vie du site Web. Nous pouvons donner l'exemple d'un site Web géré par une base de données (SIW02) où seul le contenu de la base de données est archivé. Les archives ainsi constituées ne sont ni lisibles ni exploitables.

Les copies de sauvegarde et les portraits périodiques du site Web constituent une image du site à un moment particulier de son cycle de vie, des « archives provoquées » selon Chabin (2002). Ces archives constituées doivent être également gérées par les règles de conservation afin d'en déterminer le sort final (tri, élimination ou conservation permanente). Les règles de conservation recueillies lors des entrevues ne couvrent pas cet aspect alors que dans quelques textes de politiques d'archivage du Web organisationnel analysées dans la section 4.1 (p. 115), on recommande de préciser le sort final et le support d'archivage des portraits du site Web. De même, on recommande d'archiver les documents ayant servi à la gestion et à la maintenance du site Web. Or, dans les milieux étudiés, les pratiques de gestion du site Web ne sont pas toujours documentées.

Il ressort donc que les règles de conservation utilisées dans les sites à l'étude sont soit incomplètes, soit inadéquates. Les documents de gestion d'un site Web doivent faire partie de la liste des documents à archiver et doivent être intégrés aux règles de conservation relatives au site Web. Ces règles devraient assurer la gestion des documents archivés tout au long de leur cycle de vie en précisant le sort final des portraits réalisés du site Web et de toute autre composante archivée.

4.2.1.2.2 Critères et conditions de validité

Nous nous sommes ensuite interrogée sur la validité des documents Web à capturer. En effet, pour être valide un document d'archives doit être authentique, fiable, intègre et exploitable (ISO, 2001) afin qu'il puisse servir de preuve ou de trace de cette activité dont il est le produit (voir section 2.2.2.4 de la revue de littérature, p. 40). Puisque nous n'avons pas accès au répertoire des sites Web observés (observation faite du côté client des documents seulement), nous nous sommes limitée à des constats généraux quant aux critères de validité. Nous n'avons pas pu analyser d'une manière systématique et exhaustive l'ensemble des critères de validité d'un document d'archives, notamment les critères d'intégrité et d'exploitabilité des archives Web, puisque nous n'avons pas accès aux archives constituées suite à l'archivage des sites Web observés. Nous avons pu noter les problèmes suivants :

- **Date de création ou de mise à jour** : la date de création ou mise à jour de la page Web n'est pas toujours indiquée. On indique dans quelques cas une année à côté de la mention du droit d'auteur, mais on ne peut pas vérifier s'il s'agit de la date de mise en ligne du site ou de sa mise à jour.
- **Identification de l'unité/personne responsable du contenu affiché** : on n'indique pas l'unité ou la personne responsable du contenu publié sur le site Web. On reconnaît le propriétaire du site à partir du bandeau en haut d'une page Web ou de la mention relative aux droits d'auteur (©) en bas de la page. Les documents en format PDF ou dans d'autres formats bureautiques sont bien identifiés (entête du document avec mention relative au droit d'auteur). Généralement, la métadonnée auteur indique le nom de l'organisme, mais ne spécifie pas l'unité ou la personne responsable du contenu diffusé en ligne.
- **Signature du document** : Les documents ne sont pas toujours signés. Nous avons noté que plusieurs documents en format PDF (ex. : rapport annuel, plan stratégique, communiqué, mot du directeur), qu'ils soient numérisés ou natifs numériques, ne comportent pas de signature.
- **Version en ligne ou version archivée du site Web** : sur certains sites Web, il est possible d'accéder à une version archivée du site. Toutefois, après quelques clics, on ne sait plus dans quelle version du site nous naviguons. Les politiques d'archivage proposent à cet effet d'ajouter un entête au document indiquant qu'il s'agit d'une version archivée, la date de son archivage ou de retrait du site, etc. ou de rediriger l'internaute vers un autre site où le contenu archivé est encore diffusé (Secrétariat du Conseil du Trésor du Canada, 2012).

Nous pouvons remarquer que malgré l'existence de règles éditoriales pour le site Web, les organismes étudiés ne disposent pas encore de politiques et procédures pour préserver les caractéristiques essentielles d'un document d'archives quand il est produit dans un environnement numérique réseauté.

Les informations qui permettent de bien identifier un document Web et de vérifier sa validité ne sont pas toujours indiquées dans le champ « métadonnées » du code source d'une page Web. On se limite souvent à indiquer le titre du document, le responsable moral du site (nom de l'organisme) et quelques mots clés décrivant son contenu. L'absence d'une description complète et riche d'une page Web peut occasionner des problèmes lors de l'archivage du site ou lorsqu'il s'agit d'établir une stratégie de préservation des sites et documents Web archivés telle que mentionnée dans quelques politiques analysées.

En résumé, le site Web lui-même ainsi que les documents qui le constituent présentent une valeur archivistique. L'original des documents diffusés sur un site Web organisationnel est produit dans un autre système documentaire et est géré par des règles de conservation. Ceci explique cette tendance générale à conserver le document original et non sa version Web dans les milieux étudiés. D'où l'intérêt de procéder à une macro-analyse et à une micro-analyse d'un site Web organisationnel afin d'adapter les règles de conservation aux nouveaux types de documents.

Nous avons également constaté que les documents, notamment ceux en format Web (ex. : page en format HTML), ne respectent pas toutes les conditions de validité d'un document d'archives, ce qui nous renvoie au besoin d'établir des politiques et procédures qui permettent de gérer le cycle de vie d'un document dès sa création afin d'être en mesure ensuite de préserver ses qualités archivistiques.

4.2.2 Modes de création

Les documents diffusés sur les sites Web observés sont produits dans différents environnements technologiques. Pour les distinguer, nous nous basons sur l'observation directe des caractéristiques de forme et de structure des documents analysés. Nous nous appuyons également sur des données fournies par l'aspirateur Web Scrapbook. Ainsi, nous avons pu identifier cinq principaux modes de création que nous décrivons dans les paragraphes suivants.

4.2.2.1 Numérisation et mise en ligne des documents

Parmi les documents diffusés sur les sites Web observés, nous notons la présence de documents initialement créés sur support papier qui ont été numérisés et convertis en format PDF, JPEG, etc. (ex. : procédure de transfert des archives courantes – SIW03; photographie d'une personnalité (fonds d'archives) – SIW04; rapports annuels numérisés et convertis en PDF – SIW02). Ces documents se présentent en faible proportion toujours à des niveaux de profondeur importants dans les sites.

Les fichiers d'origine sont gérés et conservés dans un répertoire spécifique sur le serveur institutionnel. Les photographies et documents diffusés sur le Web diffèrent souvent des fichiers originaux (ex. : édition des photos numérisées, format de diffusion différent du format original). Ils sont gérés dans le répertoire du site Web. Ils se présentent dans un format statique et sont dépourvus de fonctionnalités Web.

Ces documents maintiennent les caractéristiques conventionnelles de forme et de structure de leurs équivalents analogiques (ex. : pagination, identification de l'auteur, signature, numéro d'approbation, cote du document).

4.2.2.2 Création de documents par des suites bureautiques

Nous identifions également des documents produits par des suites bureautiques telles que Microsoft Office. Voici quelques exemples :

- Modèle d'un travail académique en format DOT – SIW01
- Bordereau de transfert de documents en format DOC – SIW03
- Tableau en format XLS présentant des taux de participation – SIW06
- Ateliers et présentations en format PPT – SIW07

Ces documents sont donc diffusés dans leur format natif (DOT, DOC, XLS, PPT) ou ont subi une transformation dans un format plus approprié pour le Web. Nous en avons identifié quelques-uns créés à l'aide d'une suite bureautique et convertis en format PDF pour diffusion sur le site Web grâce à des logiciels de la suite Adobe tels qu'Adobe Acrobat et Adobe Illustrator (ex. : présentation PPT d'une formation convertie en format PDF – SIW04). Ces documents ne sont pas publiés en surface du site Web, mais plutôt accessibles aux niveaux de profondeur suivants : N3, N4 et N5. Ils commencent à perdre certains de leurs critères de validité (ex. : nom d'auteur non indiqué, pagination absente, date de publication ou de mise à jour non indiquée).

4.2.2.3 Création de documents par des suites multimédias

Lors de l'analyse, nous avons également distingué des documents qui sont produits par des suites multimédias (ex. : Flash Player). Ces documents sont produits pour le site Web, mais ne se présentent pas en format HTML. Généralement, il s'agit de documents photographiques, de vidéos ou d'animations multimédias. À titre d'exemple, les expositions virtuelles sont un produit documentaire destiné à être diffusé en ligne. Elles sont constituées généralement de documents numérisés (ex. : archives textuelles, archives photographiques) présentés selon un scénario particulier grâce à des logiciels d'animation multimédia (ex. Flash Player). Le niveau d'interactivité varie et dépend de la technologie utilisée pour produire l'exposition. Ces documents, comparés à leur forme conventionnelle, sont enrichis de fonctionnalités Web : navigation, recherche, personnalisation.

À l'examen des plans des sites observés et de l'arborescence des fichiers Web capturés par Scrapbook, nous remarquons que ces documents sont localisés aux premiers niveaux de profondeur d'un site Web, c'est-à-dire les niveaux N0, N1 et N2.

4.2.2.4 Création des documents par des éditeurs Web

Le recours à des éditeurs Web favorise la création de documents propres au site Web. Nous avons identifié des documents développés en HTML ou ASP. Ces documents sont gérés par des feuilles de style en CSS. Ces documents se retrouvent non seulement en surface des sites Web, mais aussi à des niveaux de profondeur plus importants, c'est-à-dire situés à plusieurs clics de la page d'accueil. Ces langages de balisage permettent d'encapsuler différents objets documentaires et de faciliter l'exécution de plusieurs scripts (ou commandes). Une page Web en format HTML peut alors contenir des images, des animations multimédias et des appels de différentes fonctionnalités Web (ex. : saisie d'un formulaire Web).

4.2.2.5 Création des documents par des systèmes de gestion de contenu

Des documents propres au Web sont générés de manière dynamique lorsque des systèmes de gestion de contenu sont utilisés pour gérer le site Web. Ces

documents sont uniques et n'ont pas d'équivalents dans un autre environnement de production documentaire. Il s'agit des billets publiés sur des blogues organisationnels, des archives en ligne du site Web lui-même, des documents produits suite à des transactions en ligne.

Nous pouvons également citer l'exemple des agendas en ligne dont la forme reste similaire aux agendas papier, mais auxquels s'ajoutent néanmoins des fonctionnalités Web. L'internaute pourrait effectuer par exemple une recherche dans l'agenda afin de repérer un événement passé ou à venir, afficher la liste des événements par catégorie d'activité, etc. (ex. : agenda en ligne sur les sites SIW06 et SIW09).

Somme toute, un site Web organisationnel est alimenté par différents systèmes et applications ce qui entraîne la diversification de l'origine et des formats des documents qui le composent.

4.2.3 Typologie des archives Web

À cette étape de la recherche, nous souhaitons examiner les composantes des sites à l'étude afin de dresser une cartographie des catégories et types de documents Web qu'on y retrouve et de déceler l'existence possible de nouveaux types de documents propres au Web organisationnel. Dans cette section, nous présentons les résultats de l'observation, sous un angle archivistique, d'une sélection de documents diffusés sur les sites Web retenus dans le cadre de cette étude.

Afin d'identifier les types d'archives propres au Web ainsi que les nouveaux types d'archives dont il est à l'origine, nous nous inspirons des travaux récents sur la notion de genre (voir section 2.2.1.3 de la revue de littérature, p. 33). Les résultats de ces travaux indiquent que les attributs de but/fonction, de contenu et de forme permettent d'identifier et de catégoriser les documents qui présentent des attributs similaires. Ces travaux démontrent que l'usage de la technologie Web et son appropriation par une communauté provoquent non seulement la transformation des documents existants, mais aussi la création de nouveaux documents qui se distinguent

par leurs attributs de but ou de fonction (lien avec une action, une activité), de contenu, de forme et de fonctionnalités (Shepherd et Watters, 1998, 1999).

Nous avons étudié un corpus de 737 documents Web afin d'identifier les catégories et types de documents d'archives existants et de vérifier la présence de nouveaux types de documents propres aux sites observés.

Les catégories et types de documents Web identifiés sont présentés dans un premier temps. Ensuite, nous discutons de l'identification de nouvelles archives propres aux sites Web.

4.2.3.1 Types et catégories des documents d'un site Web organisationnel

Il a été possible d'identifier le type de plusieurs documents analysés en les comparant à des documents similaires produits dans d'autres environnements. À titre d'exemple, les rapports annuels, les plans de classification, les calendriers de conservation, les politiques de gestion documentaire, les affiches et dépliants, les guides d'étudiants, que nous avons repérés sur les sites Web sont diffusés généralement en format PDF. Ils restent similaires à leurs équivalents sur support papier. Ces documents conservent leurs conditions de validité, c'est-à-dire, la dénomination sociale de l'organisme créateur (en-tête du document), la structure du document, la pagination du document, la signature, la date de création ou de mise à jour (pied de page), etc.

Il en va de même pour la forme et le contenu de certains documents. Par exemple, à partir d'une page Web dont le contenu est relatif aux mandats ou à l'historique de l'institution, nous parvenons à identifier le type de ce document en se basant principalement sur le titre du document et son contenu. Les éléments qui permettent d'identifier son producteur, sa date de publication sur le site Web ou de sa mise à jour se retrouvent dans d'autres fichiers (ex. : les métadonnées permettant d'identifier l'auteur du document sont consignées dans la feuille de style).

Certains documents (ex. : communiqués) sont diffusés dans leur format numérique conventionnel (document en format PDF). Lorsqu'ils sont diffusés en format Web (document en format HTML), ils perdent leur forme conventionnelle qui

permet d'identifier leur type de document de même que certaines de leurs conditions de validité (ex. : omission de la date de création, pagination). Grâce à leur attribut de contenu (objet du document), il est possible de reconnaître les documents ayant subi des modifications importantes au niveau de leur forme ou ceux dont l'appellation n'est plus la même (ex. Salle de presse, Nouvelles, Archives).

En nous basant sur des critères de forme, de contenu et de fonction, nous avons pu classer les documents identifiés (documents conventionnels, documents évolutifs et nouveaux types de documents) sous les deux principales catégories de documents d'archives : les documents de gestion et les documents de fonction.

DOCUMENTS DE GESTION

Nous retrouvons les documents communs à tous les organismes, les documents de gestion, en surface des sites Web observés, soit les niveaux N0, N1 et N2.

Les documents de gestion forment environ 30% des documents analysés. Ces documents intègrent de plus en plus de fonctionnalités Web tout en conservant les propriétés de leurs équivalents sur support papier. À titre d'exemple, la liste du personnel se présente en format HTML. Ce format de présentation permet d'identifier le nom de l'employé, le titre de son poste, ses coordonnées avec la possibilité de le contacter en cliquant tout simplement sur son adresse courriel.

Les documents de gestion que nous avons pu identifier sur les 11 sites analysés appartiennent aux familles suivantes :

- Direction générale : plan stratégique, plan annuel, mission, mandats, historique, etc.
- Gestion des ressources humaines : liste du personnel, membres des comités, etc.
- Gestion des ressources financières : formulaire de don.
- Communication et relation avec le client/citoyen : dépliants, publications administratives, plan d'accès au bâtiment, services en ligne, service d'abonnement, foire aux questions, communiqués, etc.

Certains types de documents se retrouvent uniquement en surface du site Web (page d'accueil, page d'orientation, page présentant des informations mises à jour), d'autres peuvent se retrouver à des niveaux différents. Nous n'avons pas été à même d'établir un modèle constant de la répartition des documents, chaque site étant particulier.

Le Tableau XIII (p. 154) expose quelques types de documents de gestion retrouvés à deux niveaux de profondeur, soit le niveau 1 et le niveau 2.

Tableau XIII : Exemples de documents de gestion par niveau de profondeur et par fréquence d'apparition.

DOCUMENTS DE GESTION			
Niveau 1		Niveau 2	
Page d'accueil du site Web	n=11 sites		
Coordonnées, Pour nous joindre	n=11 sites		
Mission et histoire	n=6 sites	Mission et mandats	n=5 sites
		Historique	n=1 site
Communiqués	n=5 sites	Communiqués	n=1 site
Publications	n=4 sites	Publications	n=3 sites
Politiques de diffusion et d'accès à l'information	n=4 sites	Politiques de diffusion et d'accès à l'information	n=1 site

Nous n'avons pas repéré un grand nombre de documents relatifs à la gestion des processus administratifs, ce qui est tout à fait normal vu la nature de ces documents et le risque qu'ils contiennent des informations sensibles au sujet de l'organisation. Les documents de gestion diffusés sont à caractère public et informatif, car en vertu du *Règlement sur la diffusion de l'information et sur la protection des renseignements personnels* (Chap. A-2.1, r. 2, art. 4), les organismes publics au Québec visés par ce règlement doivent diffuser sur leur site Web différents types de documents/renseignements relatifs à leurs activités de gestion (ex. : rapports annuels, plans stratégiques, plan de classification).

DOCUMENTS DE FONCTION

Une proportion importante des pages Web des sites étudiés est reliée aux fonctions et aux activités propres à l'organisme. Les sites Web observés comportent environ 70% de documents de fonction. Ceci est un indice extrêmement important sur le lien qui existe entre le site Web et les activités propres à l'organisation. Le site Web découle de ces activités et les documente. En même temps, c'est un moyen qui contribue à la réalisation de ces activités (ex. : promotion des programmes d'études), d'où l'attention que doit accorder une organisation à son site Web.

Voici quelques documents de fonction propres au milieu d'enseignement (n=2 sites)²² :

- Banque d'emplois
- Liste de sujets d'étude
- Cours (matériel didactique)
- Demande d'admission
- Demande d'admission – Aide mémoire
- Démonstration d'une formation (capsule vidéo)
- Description d'un programme d'étude
- Guide de l'étudiant
- Guide de stage
- Horaire des cours
- Liste des diplômés
- Liste des travaux de recherche réalisés
- Profils des professeurs
- Répertoire des cours

Ces documents se retrouvent à différents niveaux de profondeur d'un site Web et se présentent sous différents formats : page HTML, animation en format SWF ou document en format PDF.

4.2.3.2 Nouvelle typologie des archives Web

Suite à l'observation des sites Web des organismes à l'étude, de même que des documents publiés sur ces sites, nous avons pu identifier différents types de documents. Certains documents ont été faciles à identifier puisque leur intitulé et leur

²² Malgré le nombre important des documents de fonction analysés, nous ne pouvons pas donner d'exemples des autres milieux étudiés au risque d'identifier ces derniers puisque les documents de fonction sont liés directement à la mission d'une organisation.

forme restent similaires aux documents en format analogique. Nous pouvons conclure que même s'ils ont migré sur le Web, ces documents, produits pour le site Web, n'ont pas subi de modifications majeures au niveau de la forme et de la structure.

Quant aux documents produits sur le site Web, il s'agit des documents générés grâce à une base de données ou à un système de gestion de contenus suite à une action particulière de la part de l'internaute. Ainsi, lorsqu'un internaute souhaite s'abonner à une liste d'envoi (SIW07), il reçoit une confirmation de son inscription au service. Ce document n'existe pas auparavant sur le site Web observé, il est généré suite à l'action de l'abonnement. Ces systèmes Web peuvent générer de manière dynamique des types existants (ex. : bulletin de notes) de même que de nouveaux types de documents (ex. : archives en ligne).

Le type de certains documents n'a pas été facile à identifier et, pour ce faire, il fallait procéder par analogie, c'est-à-dire, rechercher des éléments de forme ou de contenu similaires ou qui se rapprochent de la production documentaire plus conventionnelle. Ces documents se différencient graduellement des autres formes documentaires au niveau de :

- leur forme (ex. : agenda en ligne);
- leurs fonctionnalités (ex. : possibilité de faire une recherche dans un état général des fonds).

Finalement, nous avons identifié de nouveaux types d'archives Web qui se différencient complètement des formes documentaires traditionnelles au niveau de la forme et des fonctionnalités. Ce sont, entre autres les types d'archives suivants :

- Archives en ligne d'un document Web (ex. : communiqués);
- Expositions virtuelles;
- Foire aux questions;
- Page d'accueil, « Page d'orientation » (page ne présentant qu'une liste de liens Web);
- Blogue institutionnel.

Certaines organisations ont également développé leur page Facebook, diffusent des vidéos sur YouTube (ou autres services similaires), des photos sur

Flickr et détiennent des comptes Twitter. Des liens vers ces applications du Web 2.0 sont repérés sur plusieurs sites observés. L'abonnement à ces services est, à n'en pas douter, à l'origine de la création de contenus enrichis des fonctionnalités propres au Web social (ex. : partage, étiquetage, commentaires). L'archivage des médias sociaux est d'actualité. Lors du congrès 2012 du Conseil international des archives, l'archiviste national des États-Unis a invité la communauté internationale des archivistes à réfléchir sur les enjeux et les moyens d'archivage du Web 2.0 (ICA, 2012), sachant que les contenus diffusés sur le Web social ne sont pas sous le contrôle de l'organisation.

Ainsi, quatre catégories de documents coexistent sur un même site Web : les documents numérisés, les documents numériques bureautiques diffusés sur le site, les documents produits pour le site (ex. : animations FLASH) et les documents produits par le site (ex. : pages Web dynamiques). En nous basant sur les attributs de forme et de structure, nous avons pu identifier des types de documents existants, mais ayant subi quelques transformations sur le Web, tout comme de nouveaux types de documents qui se distinguent par leur forme, structure et fonctionnalités.

En résumé, nous avons présenté dans cette section le résultat de l'observation des caractéristiques techniques et documentaires de 11 sites Web publics afin de comprendre la nature et la typologie des archives lorsqu'elles sont produites ou qu'elles se retrouvent dans un environnement Web. Un site Web public s'est révélé comme un objet documentaire complexe et protéiforme, une forme documentaire à part entière dont les composantes documentaires et technologiques ainsi que les fonctionnalités sont aussi variées que différentes et issues de divers modes de création. Ces documents qu'ils soient propres au Web ou produits dans d'autres environnements, mais diffusés sur le site Web ont une valeur archivistique (primaire et secondaire) sauf que les règles de conservation élaborées pour leur gestion ne couvrent pas à l'heure actuelle l'ensemble de leur cycle de vie. Une grande proportion des documents diffusés sur les sites observés sont intimement liés à la mission propre de l'organisme créateur, ce qui confirme l'intérêt de préserver un site Web à des fins administratives et historiques ainsi que le besoin de créer des archives

représentatives de leur évolution et des activités qu'il documente. Les portraits du site Web peuvent s'avérer, dans ce cas, insuffisants notamment lorsque les organismes les réalisent dans une logique d'accumulation sans les soumettre à une évaluation archivistique qui permettrait, avec le temps, d'identifier les portraits les plus représentatifs du site Web et de faciliter leur traitement à des fins d'exploitation et de diffusion. Les documents de gestion d'un site Web ne figurent pas toujours sur la liste des documents à conserver. Nous avons aussi remarqué que les outils et méthodes d'édition des sites Web observés ne permettent pas de créer des documents d'archives qui respectent toutes les conditions de validité.

Même si certains de ces documents se retrouvent dans d'autres systèmes documentaires, leur édition/publication sur le site Web leur procure des propriétés de forme et des fonctionnalités qui les distinguent des documents originaux. En effet, l'observation des archives Web a révélé l'existence de pages Web dont la forme et les fonctionnalités dépendent de la technologie de production. Cette dernière leur confère des caractéristiques uniques qui les différencient des autres types de documents que l'on retrouve dans un environnement traditionnel (ex. : environnement papier ou environnement numérique bureautique).

L'examen du contenu (objet, sujet) du document et de l'activité/action qui lui donne naissance devient essentiel pour catégoriser le document et identifier son type. Grâce à cette logique, nous avons pu identifier de nouveaux types de documents dont la forme, le contenu et les fonctionnalités sont propres au Web. Ces documents devraient être intégrés au calendrier de conservation.

4.3 Analyse des stratégies d'archivage du Web organisationnel

Afin de connaître les stratégies d'archivage du Web organisationnel, nous avons mené des entrevues avec des webmestres (n=8) et des archivistes (n=13) impliqués dans le processus d'archivage des sites Web publics de 11 milieux organisationnels au Québec. La réalisation de ces entrevues nous a également permis

de compléter et de valider notre observation des 11 sites Web publics à l'étude. Vu le temps alloué pour cette recherche et la difficulté à trouver des milieux qui répondent à nos critères d'échantillonnage, nous avons décidé de ne pas étendre les entrevues à d'autres intervenants (ex. : personnel responsable de la gestion de l'infrastructure technologique et communicationnelle de l'organisation).

Les entrevues viennent compléter, dans un premier temps, les résultats de nos observations des caractéristiques technologiques et documentaires des sites Web publics des organisations dans lesquels œuvrent nos participants. Dans un deuxième temps, elles permettent de comparer les pratiques organisationnelles aux recommandations énoncées dans les politiques d'archivage analysées dans la section 4.1 (p. 115) et leur adéquation aux caractéristiques des documents Web observés.

Dans le cadre des entrevues réalisées, les participants nous ont renseignée sur les caractéristiques et l'historique du site Web public de leur institution, sur leurs rôles et responsabilités quant à la gestion et l'archivage du site Web public de l'organisation. Les résultats des entrevues nous ont aidée à mieux comprendre le processus de publication de l'information sur le site Web public des sites à l'étude, les stratégies d'archivage mises en œuvre pour le conserver, les unités d'archivage ciblées et les finalités d'un tel processus. Les données recueillies lors des entrevues nous ont permis de vérifier si les contenus archivés sont exploités de quelque manière et si une stratégie d'archivage est développée pour garantir l'accessibilité et la pérennité des documents archivés.

Nous exposons, ci-après, les résultats d'analyse des entrevues de recherche.

4.3.1 Intervenants impliqués dans le processus d'archivage

Dans le cadre de cette recherche, nous avons interviewé deux principaux groupes d'intervenants impliqués dans le processus d'archivage du Web organisationnel, soit les archivistes (n=13) et les webmestres (n=8). Au total, nous avons donc interviewé 21 personnes dans les 11 milieux à l'étude.

Les informations recueillies auprès de ces deux groupes d'intervenants portent sur la gestion de l'ensemble du cycle de vie du site Web et des documents qui y sont diffusés. Ainsi, nous avons pu avoir une description précise des différents modes de création d'un site Web, des modalités d'affichage et de mise à jour du contenu en ligne de même que des procédures de sauvegarde et d'archivage du site. Le Tableau XIV (p. 160) dresse un portrait des rôles et responsabilités des répondants en matière de gestion et d'archivage de leur site Web organisationnel. Ces rôles et responsabilités seront détaillés dans les paragraphes qui suivent.

Tableau XIV : Rôles et responsabilités des répondants quant à la gestion et l'archivage du site Web public de leur organisation.

FONCTION	MILIEUX DE RECHERCHE	RÔLES/RESPONSABILITÉS
Webmestre	Milieu de l'enseignement (4 milieux, 4 participants)	<ul style="list-style-type: none"> – Diffusion/publication de l'information sur le site Web – Respect de l'identité visuelle du site Web – Mise à jour et révision linguistique de l'information publiée sur le site Web – Refonte du site Web – Gestion d'autres systèmes d'information Web (sous-sites, intranet, extranet) – Réalisation de sauvegardes (<i>back-up</i>) du site Web – Développement ou adaptation d'applications/scripts pour l'archivage du site Web – Application et contrôle de l'application de politiques relatives au site Web (politique de confidentialité, politique linguistique, politique de diffusion de l'information)
	Milieu gouvernemental (3 milieux, 4 participants)	
Archiviste	Milieu de l'enseignement (5 milieux, 5 participants)	<ul style="list-style-type: none"> – Développement de politiques/règles de conservation du site Web – Participation active au processus d'archivage du Web – Conservation des archives du site Web – Lancement d'une réflexion pour la mise en œuvre d'une stratégie d'archivage du site Web de l'organisme devenu dynamique
	Milieu gouvernemental (4 milieux, 6 participants)	
	Milieu professionnel (1 milieu, 1 participant)	

FONCTION	MILIEUX DE RECHERCHE	RÔLES/RESPONSABILITÉS
Archiviste Webmestre	Milieu universitaire (1 milieu, 1 participant)	– Gestion et archivage du site Web (publication/production de l'information et conservation du site Web)

4.3.1.1 Rôles et responsabilités des webmestres

Les webmestres (n=8) assurent principalement des tâches techniques liées à la gestion du site Web dont le respect de l'identité visuelle du site Web, l'édition, la mise en ligne et la mise à jour des contenus ainsi que la réalisation des copies de sauvegarde. Ils sont responsables de la centralisation du processus de diffusion de l'information. Les gestionnaires des contenus produisent le contenu informationnel qui alimente les différentes sections du site Web. Ce contenu est alors acheminé aux webmestres qui l'éditent en format Web. Certains documents sont diffusés dans leur format d'origine (ex. : formulaire en format DOC, tableau statistique en format XLS) ou convertis en format PDF (ex. : rapport annuel, document de formation).

Les webmestres du milieu gouvernemental (n=3) indiquent qu'ils procèdent également à une révision linguistique avant de publier un contenu en ligne, et ce, en respect de la *Politique gouvernementale relative à l'emploi et à la qualité de la langue française dans l'administration* (Gouvernement du Québec, 2011).

Les webmestres sont généralement supportés dans leurs tâches par des gestionnaires de l'infrastructure technologique et communicationnelle, et ce, au niveau de la gestion de l'infrastructure technologique du site Web. La gestion et la maintenance du site Web organisationnel sont une responsabilité partagée entre la direction des technologies de l'information (gestion de l'infrastructure technologique) et la direction des communications (édition et suivi du contenu), les webmestres étant souvent associés à la direction des communications (n=8) et les spécialistes en technologie à la direction des technologies de l'information.

4.3.1.2 Rôles et responsabilités des archivistes

Comme présenté dans le Tableau XIV (p. 160), les archivistes (n=13) ne sont pas très impliqués dans la gestion et l'archivage des sites Web de leur organisation. Ce manque d'implication des archivistes dans l'archivage du Web organisationnel relève, selon ce que nous avons observé dans les sites d'étude, de quatre facteurs fréquemment remarqués dans les milieux à l'étude :

1. Manque de maîtrise des technologies Web;
2. Absence d'une connaissance approfondie de l'actif documentaire du site Web ainsi que des fonctions et objectifs de ce dernier;
3. Manque de collaboration entre archivistes et webmestres;
4. Inégalité au niveau du partage des responsabilités entre les unités d'une même organisation.

En effet, les archivistes rencontrés n'ont pas eu de formation sur l'archivage des sites Web et ne semblent pas tous maîtriser les technologies Web. Ils interviennent plus sur des documents papier ou des documents numériques bureautiques. Plusieurs d'entre eux ne reconnaissent pas la valeur archivistique des documents Web. Les pratiques d'archivage du Web organisationnel ne sont pas bien documentées dans la littérature professionnelle en archivistique. Un cadre de référence dans ce champ d'intervention n'existe pas tant au niveau institutionnel que national. De plus, la conception et la gestion des sites Web relèvent de la responsabilité des webmestres. Un seul archiviste parmi les 13 rencontrés a une connaissance pratique des techniques d'édition d'un site Web, puisqu'il est responsable de la gestion du site Web de son unité. Même s'ils ne sont pas directement impliqués dans la gestion du site Web, d'autres archivistes (n=4) connaissent les principes théoriques d'édition et de fonctionnement des sites Web (ex : niveau de profondeur à respecter, lien avec une base de données, etc.). Ils reconnaissent également l'importance d'être impliqués dans leur gestion et leur archivage. À titre d'exemple, un des archivistes participant à cette étude a été responsable d'un projet visant à analyser la qualité ergonomique et informationnelle du site Web de son unité. Le projet a mené à une refonte du site Web actuel afin de

correspondre aux besoins du public cible et rendre le site plus dynamique et ergonomique. À l'exception de ces cinq participants, les autres archivistes n'ont pas de connaissances en technologies Web.

Nous avons aussi remarqué lors des entrevues que seulement 50% des archivistes rencontrés connaissent l'actif documentaire du site Web de leur organisation et de ses principales sections tandis que 30% des archivistes ne connaissent ni la nature du site Web, ni le processus de publication, de gestion et d'archivage de ses contenus. Néanmoins, la majorité des archivistes sont d'avis qu'un site Web institutionnel devrait être archivé. Pour certains (n=8), le site Web fait partie du patrimoine documentaire de l'organisation. Tel que le confirme un archiviste du milieu universitaire (R03) participant à cette recherche, le site Web est « une trace de la présence en ligne de l'organisation et un outil qui supporte l'organisation dans la réalisation de sa mission, il devrait être archivé ». Un site Web doit être considéré comme un document « multiformat » qui a une « valeur administrative et historique » (R03).

Pour d'autres (n=2), les documents diffusés sur le site Web sont pris en charge par d'autres systèmes documentaires internes à l'organisation (ex. : système de gestion électronique de documents, système de gestion des contenus). Même si ce n'est le cas, il est même « inutile d'en produire une autre copie puisqu'on peut la retrouver sur Internet Archive » (R05). Quelques archivistes rencontrés lors de cette étude (n=3) ne sentent pas l'obligation de l'archivage du Web organisationnel « tant que des institutions telles que Bibliothèque et Archives Canada ou Bibliothèque et Archives nationales du Québec n'ont pas encore statué sur le sujet » (R10). Il n'est pas donc nécessaire de « se lancer dans un projet d'envergure sans cadre de référence ni ressources suffisantes et compétentes pour le gérer » (R10).

Par contre, les archivistes affirment, quand on leur demande leur point de vue, que le processus d'archivage d'un site Web ne devrait pas différer de celui des documents d'archives en général. Partant ainsi de leur vécu, les archivistes déduisent que la gestion documentaire de ce nouveau type de document ne devrait pas différer

notamment dans ses principes de celle qu'ils appliquent déjà aux documents numériques bureautiques ou multimédias.

Le manque de connaissances des technologies Web de la part des archivistes et leur perception négative de la valeur informationnelle des sites Web sont à l'origine d'un partage inégal au niveau des responsabilités de gestion et d'archivage des sites Web dans les organismes étudiés. Dans les sites à l'étude, les webmestres sont responsables, dans la majorité des cas, de l'ensemble des tâches relatives à la gestion et à l'archivage des systèmes Web de leur organisation. Par contre, le webmestre n'est pas toujours à l'affût des exigences juridiques et archivistiques en matière de gestion des documents administratifs. Nous avons constaté, lors des entrevues, que lorsque l'archiviste adopte une démarche proactive, le webmestre devient plus attentif et sensible aux exigences de gestion documentaire.

Alors que les bibliothécaires ont aussitôt reconnu la valeur scientifique, historique ou culturelle que pourrait présenter le Web public, les archivistes ne sont pas toujours arrivés à un consensus autour de cette question. La perception de l'archiviste de cet objet documentaire qu'est un site Web est également importante. Deux archivistes parmi ceux interviewés dans le cadre de cette recherche ne semblent pas être préoccupés par l'archivage du Web organisationnel puisqu'ils conçoivent « le site Web en tant que médium de plus de diffusion de l'information administrative » (R05). Ces archivistes semblent avoir une conception similaire à celle des webmestres : le site Web est outil technologique. Les copies de sauvegarde sont suffisantes pour son fonctionnement.

4.3.2 Stratégies d'archivage

Dans cette section, nous présentons, tout d'abord, les approches et techniques d'archivage observées dans les 11 sites d'étude. Ensuite, nous préciserons les outils et les supports d'archivage du Web organisationnel. Par la suite, nous étudierons les méthodes de pérennisation mises en œuvre dans les milieux observés. Il est à noter que, pour trois milieux, les données recueillies manquent de détails techniques du fait

de n'avoir pu rencontrer les webmestres. Les données obtenues sur le site Web et son mode de gestion proviennent en ce cas des archivistes.

4.3.2.1 Approches d'archivage

Nous avons rencontré une difficulté majeure lorsque nous avons voulu recueillir des informations à propos des approches d'archivage. Le terme archivage ne signifie pas la même chose chez les webmestres et les archivistes. Nous souhaitons donc clarifier ces nuances avant de présenter les approches d'archivage utilisées dans les milieux à l'étude. Du point de vue du webmestre, archiver un site signifie en produire à une fréquence particulière une copie de sauvegarde (*back-up*). Pour l'archiviste, l'archivage du site Web consiste à capturer le site Web à un moment bien particulier de son cycle de vie dans un format qui garantit sa pérennité à long terme, autrement dit, en produire un portrait à un moment donné.

Notons que les webmestres des milieux étudiés (n=11 milieux) veillent à la réalisation de copies de sauvegarde du site Web public de l'organisation. Cette approche consiste plus précisément à créer régulièrement et de manière automatisée des copies de sauvegarde (*back-up*) grâce à des scripts du site entier par l'organisme producteur du site ou par la société qui héberge le site Web. Il s'agit d'une mesure de sécurité qui garantit une mise en ligne rapide du site Web en cas de problèmes. Cette approche ne permet pas de constituer une collection des différentes versions du site Web, car une nouvelle sauvegarde écrase la précédente. La sauvegarde du site Web est souvent la seule approche qui est opérationnalisée lorsque l'archiviste n'est pas impliqué dans les travaux de l'équipe des webmestres ou lorsqu'il ne voit pas l'intérêt de conserver le site Web. Elle a été identifiée comme étant une mesure nécessaire à inclure dans une stratégie d'archivage (Archives Office of Tasmania, 2009; University of Toronto, 2006).

Suite à l'analyse des entrevues, nous notons que quatre cas de figure se présentent, les trois premiers correspondant à des milieux où le site Web public de l'organisation n'a pas été archivé pour diverses raisons, tandis que le quatrième cas

de figure, le plus fréquent, regroupe des milieux ayant mené de projets d'archivage plus au moins aboutis.

Dans un premier cas (n=1 milieu), la décision a été prise de ne pas archiver le site Web de l'organisation parce que : (1) le site Web ne présente pas de valeur primaire ou secondaire et (2) qu'en cas de besoin, un autre organisme (Internet Archive) le fait déjà.

Pour un deuxième cas (n=1 milieu), l'archivage en est au choix d'une stratégie. L'archiviste en collaboration avec le webmestre de son unité mène une réflexion et plusieurs expérimentations pour déterminer la stratégie d'archivage la plus appropriée.

Dans un troisième cas (n=1 milieu), l'organisme cherche à reconstituer les archives de son site Web et réfléchit sur la question de l'archivage. Les archivistes de l'organisme souhaitent en effet récupérer les versions antérieures du site Web de leur institution. Faute de moyens financiers, ils n'ont pas pu faire une acquisition rétrospective des archives de leur site auprès d'Internet Archive. Des réunions seront organisées avec la nouvelle équipe des webmestres pour discuter de l'archivage du site Web de l'organisation.

Quant au quatrième cas de figure, le plus fréquent, les organismes (n=8 milieux) procèdent à l'archivage de leurs sites Web selon des approches différentes qui, n'étant pas mutuellement exclusives, sont les suivantes :

- Archivage intégral du site Web (n=5 milieux) : cette approche vise à réaliser un portrait complet du site Web selon une périodicité prédéfinie. Trois organismes appliquent cette méthode dans le respect d'une règle de conservation (un de ces trois organismes est en attente de l'approbation de son projet de règle de conservation par BAnQ). Dans un autre cas, le webmestre procède à l'archivage du site Web à des moments particuliers de son évolution. Ses choix étaient cohérents avec la règle de conservation de son institution de tutelle même s'il n'était pas au courant de son existence. Dans un autre organisme, il n'existe pas de règle, mais la procédure d'archivage est officieusement appliquée.
- Approche mixte (n=5 milieux) : cette approche permet à la fois de réaliser des copies de sauvegarde du site Web entier ou de la capture périodique d'un portrait du site en plus d'archiver de manière sélective des

contenus/documents bien particuliers. Dans un milieu, un portrait du site Web est réalisé annuellement tout en appliquant des règles spécifiques aux rapports annuels publiés en format PDF sur le site Web (SIW06). Dans un autre, une copie des cours diffusés sur le site Web est effectuée en marge des portraits périodiques du site Web (SIW01).

Le choix de l'une de ces deux approches d'archivage dépend de la nature du site Web, des usages qui en sont faits et du mode de gestion des documents qui y sont publiés. Un des archivistes participant à cette recherche nous a expliqué qu'« il y a peut-être des choses à archiver, mais pas tout » (R05). Un autre participant nous a confirmé que l'archivage de la base de données liée au site Web public de son organisation ne l'intéresse pas (R06). Selon cet archiviste, il est plus important de garder la trace d'une information ou d'un service offert en ligne dans son contexte, c'est-à-dire, le site Web. La réalisation d'un portrait annuel du site Web, dans ces deux cas, suffit. Un autre participant démontre l'importance d'archiver l'état général des fonds diffusé sur le site Web du service des archives puisqu'« il n'y a plus d'état général en papier » (R03) confirmant ainsi que certains documents sont propres au site Web et, comme ils n'ont pas d'équivalents dans un autre environnement documentaire, ils doivent être archivés. Une approche mixte est dans ce cas privilégiée.

4.3.2.2 Techniques d'archivage

Les techniques d'archivage des sites Web publics varient d'un organisme à un autre. Nous avons identifié trois principales techniques non mutuellement exclusives²³, soit :

- La capture à distance grâce à un robot moissonneur (*crawler*) (n=2 milieux) : cette technique est effectivement utilisée dans un milieu professionnel où le logiciel HTTrack est employé pour collecter les pages de l'ensemble du site Web de l'organisme. Le service d'archives d'une institution d'enseignement universitaire expérimente différentes techniques, dont celle du moissonnage à distance en testant différents robots moissonneurs (SIW04).

²³ Ces trois techniques sont expliquées au point 2.4 de la revue de littérature (p. 65).

- L'archivage à la source (n=9 milieux) : cette technique est utilisée dans la plupart des organismes à l'étude. C'est la méthode la plus efficace pour réaliser des copies de sauvegarde du site Web ou pour capturer de manière sélective des documents bien particuliers.

Des scripts et des procédures d'archivage sont développés pour réaliser des captures périodiques du site Web public de quelques organismes (n=6). Les webmasters des organismes consultés sont responsables de ces développements techniques.

La fréquence des copies de sauvegarde (*back-up*) est quotidienne alors que la fréquence des captures du site Web à la source (copie du répertoire du site Web) varie d'un milieu à un autre. Généralement, dans les milieux d'enseignement, l'archivage du site Web se fait à la fin de chaque session d'études (n=2). Dans le milieu gouvernemental (n=1), le milieu professionnel (n=1) et les services d'archives universitaires (n=1), deux captures par année sont réalisées. Selon les personnes interviewées, cette fréquence permet de révéler les changements qu'a subis un site Web au cours d'une année.

- Le tiers-archivage (n=3) : cette technique est adoptée par un seul organisme (O02). Cette technique n'a pas été identifiée lors de la recension des écrits. Elle a émergé des données de recherche.

En plus de réaliser des copies de sauvegarde à l'interne, l'organisation exige de l'hébergeur du site de fournir une copie du contenu de la base de données qui alimente le site. La conservation de l'aspect visuel et de la structure du site Web est secondaire. Un autre organisme (O11) parmi les 11 consultés voulait acquérir des archives du site Web public de son institution sur Internet Archive afin d'avoir des versions antérieures d'un site ayant subi plusieurs refontes avec des équipes différentes de webmasters. Faute de moyens financiers, l'organisme n'a pas pu le faire. Un troisième n'archive pas son site Web sachant que les archives de ce dernier se retrouvent dans Internet Archive.

Aucun des milieux interviewés n'a développé une stratégie pour pérenniser les archives de son site Web. Dans un milieu, l'archiviste nous confirme qu'il n'a pas accès aux archives Web puisqu'elles sont conservées par le webmaster de l'organisme. Lorsque les rôles et responsabilités ne sont pas clairement définis, le risque d'engendrer des problèmes d'exploitabilité et de préservation des archives devient élevé (ex. : dans ce cas, l'archiviste ne peut déterminer sur quel support ces archives ont été conservées, si elles doivent être migrées sur un autre support, ou si, étant destinées à un archivage à long terme, elles pourraient servir dans la production d'autres documents, la préparation ou l'évaluation de certaines actions et décisions,

etc.). Dans un autre milieu, l'archiviste est conscient de cette problématique, mais il n'a pas encore envisagé une solution. Le webmestre d'un autre milieu ne se soucie pas des problèmes d'accessibilité aux archives du site Web gravées sur un CD-ROM il y a plus de 5 ans déjà. Les webmestres ne portent pas une attention particulière à cette problématique : « On n'a pas faite [sic] l'exercice. On ne me l'avait pas demandé. » (R09.2)

Les sites Web sont archivés sur des CD-ROM (n=5), alors que les copies de sauvegarde sont réalisées directement sur les serveurs (n=7). Ces choix actuels des supports de conservation ne garantissent toutefois pas l'accessibilité à ces documents sur le long terme, nous confirment quelques participants à cette étude. Cet état de fait démontre à quel point il serait judicieux d'établir une politique d'archivage afin de recommander un support d'archivage ou une solution de stockage plus pérenne et de définir une stratégie de préservation des documents archivés à long terme.

4.3.2.3 Outils de gestion et d'archivage du Web organisationnel

Deux outils archivistiques sont à considérer lors de la mise en œuvre et la gestion d'un projet d'archivage d'un site Web organisationnel, à savoir : le calendrier de conservation et la politique d'archivage.

Il importe de rappeler, dans un premier temps, que le calendrier de conservation est un outil archivistique qui permet de définir grâce à une règle de conservation²⁴ le sort d'un document d'archives à chacune des étapes de son cycle de vie. Il est établi à l'interne par l'équipe des archivistes de l'organisation et il doit être approuvé au Québec par les archives nationales lorsqu'il s'agit d'un organisme public. Dans notre cas, 6 des 11 sites d'étude disposent de règles de conservation pour le site Web public de leur organisation.

Le milieu universitaire au Québec, par le biais de la CREPUQ, a proposé deux règles de conservation. Une règle porte sur le contenu diffusé sur un site Web universitaire et la deuxième concerne plutôt les documents relatifs à la gestion du site

²⁴ Une règle de conservation précise le délai de conservation d'un document tout au long de son cycle de vie et son sort final ainsi que le support de conservation au besoin.

Web. Ces deux règles ont inspiré uniquement l'archiviste de l'organisme O3 (O1, O4 et O11 ne l'appliquent pas). L'organisme O2 applique, en effet, une règle informelle puisque cette dernière n'a pas encore été approuvée par BAnQ. Quant à l'organisme O6 qui a développé une règle de conservation formelle, il songe à la mettre à jour. En effet, la règle est trop détaillée (la règle vise la capture de différentes composantes du site Web) et difficile à appliquer dans le contexte actuel. L'archiviste voudrait mettre à jour la règle pour qu'elle soit plus opérationnelle. Les autres organismes (n=5) n'ont pas de règles spécifiques au site, mais ils appliquent des règles de conservation existantes à la version originale des documents diffusés sur le Web.

Les règles que nous avons pu recueillir (n=3) sont différentes dans leur portée et niveau de précision. La première règle porte uniquement sur le site Web en tant que contenu informationnel à archiver périodiquement. La deuxième règle s'intéresse aux documents relatifs à la gestion du site Web (création, mise à jour et développement du site Web) alors que la troisième règle couvre à la fois le contenu informationnel, la structure et les documents de gestion du site Web. Il s'agit d'une règle générale qui permet une gestion décentralisée de l'archivage des sites Web des différentes unités. Cette dernière autorise la destruction de tous les documents relatifs à la structure et à la gestion du site Web et permet la conservation, par procédure d'échantillonnage, des portraits périodiques du site Web.

Malgré l'absence d'une politique formelle pour l'archivage de leurs sites Web, nous avons remarqué que certaines recommandations préconisées dans les politiques d'archivage des sites Web sont appliquées de manière informelle dans les sites d'étude, telles que la périodicité des captures, l'application des délais de conservation, l'attribution des rôles et responsabilités et le choix de supports de stockage pérenne. Nous avons pu consulter des exemples de procédures d'archivage dans deux milieux à l'étude. La première se présente sous la forme d'une entente d'archivage avec l'hébergeur du site Web de l'organisation. Elle est très générique, mais précise l'approche à utiliser, la périodicité de son application et la personne responsable de son application. La deuxième est une procédure technique présentant les modalités et la fréquence d'exécution d'un script d'archivage. Aussi, nous avons

pu examiner les résultats d'une première expérimentation d'archivage du site Web d'un service d'archives universitaire (O4). L'archiviste responsable du projet a veillé à communiquer les résultats de ce projet à la direction des communications non seulement pour sensibiliser les responsables de cette direction aux enjeux liés à l'archivage du Web organisationnel et les risques auxquels s'expose l'organisme, mais aussi pour montrer qu'il existe des solutions techniques à évaluer par cette direction avant de les implanter.

4.3.2.4 Exploitation des archives Web

Suite à l'analyse des entrevues avec les personnes impliquées dans le processus de gestion et d'archivage des sites Web organisationnels à l'étude et de l'observation de ces sites, nous constatons que l'exploitation des sites Web archivés n'a pas été envisagée dans les milieux d'étude. Les participants à cette recherche entrevoient difficilement comment le matériel capturé pourrait :

1. répondre à des besoins scientifiques ou historiques. Le Web est perçu comme un média d'information, un des canaux que l'organisme utilise pour diffuser plus rapidement une information à jour sur ses activités et services auprès de ses partenaires et public cible.

Un des archivistes participant à cette recherche (R05) n'est pas convaincu de la valeur scientifique ou historique que pourrait présenter un site Web organisationnel. Selon lui, la mission de conserver le Web incombe aux bibliothécaires. Nous pourrions nous questionner si cette assertion ne découle pas du manque de connaissance de ce nouvel objet documentaire et de l'aura technologique qui l'enveloppe.

2. être exploité alors que les méthodes utilisées actuellement ne garantissent ni la lisibilité ni l'accessibilité du contenu archivé. La technologie évolue à un rythme effarant, les usages qui en sont faits ne cessent de se transformer. L'archiviste n'a pas eu le temps nécessaire de bien connaître et de s'adapter à ces évolutions. L'analyse des projets d'archivage du Web (voir Annexe 1, p. i) montre, qu'actuellement, la priorité des archivistes est de pouvoir capturer un objet documentaire aussi dynamique, instable et complexe. Les archives

nationales mènent une réflexion sur les besoins que pourrait satisfaire l'archivage des sites Web organisationnels et expérimentent quelques solutions techniques. L'exploitation des archives d'un site Web est une autre dimension du projet encore difficile à mettre en œuvre ou à évaluer.

Un archiviste (R03) reconnaît que cette dimension devrait être prise en considération depuis même la capture du site Web et son archivage. Selon ce répondant, même si l'exploitation des archives Web n'est pas envisagée à court terme, il faudrait s'assurer que la collection constituée pourrait être exploitable dans le long terme. Il est important, toujours selon ce répondant, de garantir la pérennité des documents Web conservés. Or, les solutions existantes ne répondent pas toujours à cette condition. Malgré cette contrainte, un archiviste (R02) conserve précieusement, de peur de perdre un fragment du patrimoine documentaire de son institution, les CD-ROM et DVD sur lesquels sont gravées des versions archivées du site Web de son organisation même si l'information qui y est inscrite n'est pas intelligible.

Cette préoccupation d'exploitation à long terme ne semble pas être la priorité des autres répondants puisqu'à cette étape de leur projet d'archivage, ils sont à la recherche des méthodes les plus appropriées pour capturer le contenu et remédier à une gestion incomplète du cycle de vie des archives Web.

D'après les résultats de l'analyse, l'archivage du Web dans les organisations ou services d'archives au Québec en est à ses débuts. Le Web organisationnel est encore perçu comme un média de diffusion d'informations sur l'organisation et un moyen d'interaction avec les usagers. Les usages actuels de ce média ne justifient pas l'intérêt qu'il pourrait présenter pour la recherche et l'histoire. De même, les techniques actuelles d'exploitation des collections Web sont très élémentaires (voir Revue de littérature, Section 2.6, p. 69).

4.3.3 Facteurs à prendre en considération lors de la mise en œuvre d'un projet d'archivage

Nous analysons, ci-après, les facteurs mentionnés par les répondants qui déterminent les choix relatifs au processus d'archivage du Web organisationnel.

Lorsqu'il s'agit d'identifier les difficultés rencontrées lors de la mise en œuvre d'un projet d'archivage du site Web organisationnel, nos répondants ne sont pas toujours capables de les déterminer. L'observation des caractéristiques des sites Web étudiés nous révèle des aspects assez complexes de ces sites qui ne sont pas bien maîtrisés par les archivistes rencontrés (ex. : technologies de développement, langages de développement, etc.). Les projets d'archivage menés actuellement dans les milieux de la recherche sont à leur premier stade de développement. Les rôles des différents intervenants dans le processus d'archivage des sites Web publics des organismes étudiés ne sont pas clairement définis.

Le manque de connaissance du processus de gestion d'un site Web par les archivistes ne favorise pas leur intégration aux autres équipes responsables de la gestion des opérations ou des technologies qui soutiennent le site Web public de leur organisation. Par conséquent, les stratégies d'archivage et les choix mis en œuvre par ces équipes ne respectent pas nécessairement la perspective archivistique.

L'archivage du Web n'impose pas toutefois la mise à contribution de plusieurs employés. Selon un des répondants (R01), une seule personne compétente et disponible pourrait se charger de l'archivage du site Web de l'organisation. Cette personne pourrait être le webmestre ou le responsable de la gestion de l'infrastructure technologique du site Web. Selon R01, il est même possible d'automatiser l'opération de sauvegarde du site Web grâce à des scripts et de prendre à des intervalles réguliers une copie sur CD ou DVD du site Web. La difficulté réside plutôt dans la disponibilité de cette personne et de son niveau de maîtrise des techniques de programmation. Dans un autre milieu (O06), la responsabilité de l'archivage est plus importante à cause de la complexité et de la taille du site Web, ce qui exige l'assignation d'une personne pour cette tâche : « Il me faut une personne qui se consacre à ça. » (R1.06). En ce qui concerne les ressources matérielles, l'archivage d'un site Web organisationnel ne requiert pas un investissement important en matériel de stockage ou en applications technologiques. C'est à même le serveur Web qu'il est possible de conserver ses versions archivées. L'essentiel est que l'organisme dispose d'un serveur qui répond à ses besoins en termes de capacité de stockage des

documents produits, insiste notre répondant R01. Les supports de stockage tels CD-ROM, DVD ne sont pas très dispendieux. Certes, ces supports sont pertinents pour un stockage à court terme, ils peuvent toutefois engendrer des problèmes de pérennité lorsqu'une stratégie de migration n'est pas définie. Plusieurs de nos répondants ont confirmé ce risque (n=3).

Malgré le peu d'informations que nous avons recueillies à propos des difficultés rencontrées lors de la mise en œuvre ou de la gestion de l'archivage des sites Web organisationnels dans chacun des milieux, la mise en commun des données nous a permis de dresser un portrait des facteurs à prendre en compte :

- Complexité du site Web : ce facteur dépend non seulement de la nature des technologies Web utilisées pour créer et gérer l'information diffusée sur le site Web, mais aussi de la nature des activités et des services offerts sur le site Web;
- Fréquence de mise à jour du site Web : le choix des périodes et des intervalles d'archivage du site Web ou de la capture de certaines de ces composantes a été bien justifié par nos répondants et dépend surtout de la nature des activités de l'organisation;
- Niveau de communication et de collaboration entre webmestre et archiviste : facteur très déterminant, le manque de communication entre ces deux acteurs organisationnels pourrait conduire à des solutions techniques qui ne garantissent pas la pérennité et par conséquent l'exploitabilité des archives Web. Souvent, les webmestres sont responsables de la gestion des sites Web. Ils se contentent d'une stratégie d'archivage basée sur la production de copies de sauvegarde. Un des webmestres rencontrés n'était pas au courant que son organisme a une règle de conservation pour les sites Web (SIW01). L'implication de l'archiviste permet, dans un premier temps, de conscientiser les webmestres à la valeur archivistique que pourrait présenter le site Web et, dans un deuxième temps, d'adopter une approche d'archivage plus appropriée;
- Hébergement du site Web : nous avons également remarqué que lorsque l'archivage du site Web est réalisé par une tierce personne, soit l'hébergeur du site Web et, notamment en absence d'une politique d'archivage, la stratégie d'archivage arrêtée n'est pas fiable et ne respecte pas les caractéristiques propres des archives. À titre d'exemple, l'organisme O2 reçoit aux six mois une version archivée de son site Web sur CD-ROM. L'archiviste (R02) ne peut pas avoir accès aux documents archivés. En effet, l'hébergeur du site Web s'est contenté de faire une copie à plat de la base de données (BDD) qui

alimente le site Web. Pour pouvoir visualiser le contenu du CD-ROM, il faut que le webmestre installe des programmes spécifiques. Donc, ici les archives du site Web ne sont pas exploitables. Elles ne répondent pas aux caractéristiques de complétude (une partie du site Web est archivée, dépourvue de contexte de création et d'usage) et d'intégrité (perte de la structure et de la forme initiale de la BDD).

- Ressources matérielles : quoiqu'elle ne constitue pas une difficulté majeure, le manque de ressources matérielles pourrait aussi pousser les gestionnaires des sites Web ou les archivistes à faire des choix qui ne garantissent pas l'exploitabilité des archives Web ou la préservation de leurs caractéristiques essentielles (authenticité, fiabilité, intégrité, exploitabilité).

En résumé, nous avons présenté dans cette section les résultats d'analyse des entrevues réalisées dans 11 sites d'étude. En premier lieu, nous nous sommes penchée sur les rôles et responsabilités associés à la gestion d'un site Web dans un milieu organisationnel. Nous avons noté le manque d'implication des archivistes dans ce processus. Ensuite, nous avons étudié les approches, techniques et outils d'archivage des sites Web observés. À ce niveau, nous n'avons pas pu avoir un modèle d'archivage, notamment propre aux sites Web gouvernementaux qui, dans notre cas, sont encadrés par les mêmes règles de gestion. Les travaux d'archivage dans les milieux d'étude sont à leurs débuts. Les solutions technologiques ne manquent pas, mais il reste à réfléchir sur le statut du site Web et à entrevoir les raisons pour lesquelles il devrait être archivé. Pour l'instant, aucun des organismes ne possède une stratégie de préservation à long terme des documents qu'il archive ni ne réfléchit aux usages futurs de ces archives. Plusieurs facteurs entrent en jeu et peuvent expliquer cette situation, les plus importants sont : la perception des archivistes de ce nouvel objet documentaire qu'est le site Web et l'absence d'une approche collaborative de son archivage.

Les stratégies adoptées dans les milieux d'étude ne sont pas toujours adaptées à la nature du contenu diffusé en ligne ou créé pour la gestion du site Web. Les copies de sauvegarde, considérées jusqu'alors comme approche d'archivage dans plusieurs milieux, ne peuvent être, à elles seules, suffisantes pour la conservation des documents archivés à long terme.

Nous remarquons que la réflexion menée par l'archiviste au sujet de l'archivage du Web n'est pas avancée dans tous les milieux. Les projets d'archivage n'ont pas encore atteint une maturité qui permet d'évaluer les résultats et de revoir les approches choisies. Nous notons également un problème au niveau de la perception de la nature d'un site Web et de la valeur qu'il peut présenter dans la continuité des affaires de l'organisation et pour l'histoire ou la recherche. Même si les politiques d'archivage considèrent un site Web comme un document d'archives devant être géré tout au long de son cycle de vie, plusieurs archivistes n'en sont pas encore convaincus. De même, la monopolisation de la gestion du site Web par les webmasters constitue un autre obstacle. Les ressources disponibles qu'elles soient matérielles, financières ou humaines ne forment pas un obstacle majeur.

Conclusion

Notre recherche vise à développer une meilleure compréhension de la nature des archives Web et de décrire la démarche de leur archivage dans un contexte organisationnel du point de vue des acteurs impliqués dans ce processus. Les résultats exposés dans ce chapitre apportent des réponses à nos questions de recherche : (1) que recommandent les politiques d'archivage du Web organisationnel, (2) quels sont les types d'archives Web retrouvés sur les sites Web publics de quelques milieux organisationnels au Québec, et (3) quelles sont les pratiques d'archivage du Web dans ces milieux.

Les résultats révèlent que :

- il existe des exemples de politiques d'archivage recommandées par des institutions archivistiques. Ces exemples de politiques témoignent de l'état de réflexion et de l'avancement de la recherche dans ce champ émergent de la discipline archivistique. Les politiques étudiées ne sont pas faciles à mettre en œuvre et, dans certains cas, elles restent limitées dans leurs propositions. Le modèle de gestion des risques associés à la diffusion de l'information sur un site Web organisationnel, prôné comme un cadre méthodologique pour le choix d'une stratégie d'archivage appropriée au Web organisationnel, ne prend pas en considération les caractéristiques propres aux archives Web. En outre, ces politiques doivent être accompagnées par d'autres outils conçus

pour les besoins d'archivage d'un site Web organisationnel (règles de conservation, liste de sites Web à capturer, liste des objets à capturer, etc.),

- les archives Web ont des caractéristiques qui leurs sont propres. Ces caractéristiques découlent non seulement de l'usage des technologies Web, mais de l'appropriation de ces technologies. Plus que jamais, le contenu d'un document Web est indépendant de sa mise en forme, sa restitution dépend d'une myriade d'applications, programmes et logiciels. Les archives Web ne satisfont pas toujours aux conditions de validité d'un document d'archives produit dans un environnement traditionnel. Même si les règles d'édition des documents Web sont souvent respectées, elles ne sont pas toujours suffisantes pour produire des documents faciles à capturer et à préserver. À cet effet, les politiques d'archivage recommandent d'appliquer non seulement différents niveaux d'archivage (archivage de portraits du site Web, archivage des unités particulières du site Web et archivage des documents ayant servi à la gestion du site Web), mais aussi d'associer des métadonnées permettant de maîtriser le cycle de vie du site Web,
- dans les milieux étudiés, l'archivage du Web est encore à un stade expérimental. L'archivage du Web n'est pas toujours une priorité dans ces milieux. En effet, les archivistes ne sont pas tous convaincus de la valeur archivistique que peut présenter un site Web organisationnel ou bien ils ne sont pas impliqués dans son processus d'archivage. À date, les organismes n'ont pas été encore confrontés à des litiges ou problèmes liés à l'information diffusée sur le site Web ni sensibilisés à son intérêt scientifique ou historique d'autant plus que cette information est consignée dans d'autres systèmes documentaires. Ce constat explique pourquoi, dans certains cas, l'archivage du Web organisationnel se limite à la production de copies de sauvegarde du site Web public d'une organisation. Le manque de ressources et de compétences a été également identifié comme un autre facteur qui explique le manque de retour d'expériences à ce sujet.

Ces résultats seront discutés dans le chapitre suivant.

5. Discussion

Introduction

Nous avons présenté au chapitre précédent les résultats détaillés de l'analyse des données relatives aux recommandations des politiques d'archivage du Web organisationnel repérées en ligne, aux caractéristiques des archives Web ainsi que des pratiques d'archivage observées dans 11 milieux organisationnels au Québec. Cette démarche nous a permis de constater des écarts dans les perceptions de la valeur archivistique du Web et dans les pratiques de son archivage. Dans le présent chapitre, nous tentons d'en comprendre les raisons et de proposer un modèle de politique qui tienne compte des caractéristiques des archives Web et de la réalité des organismes étudiés.

La discussion des résultats est organisée en trois sections. Dans un premier temps, nous proposons une réflexion sur le site Web en tant que nouvel objet archivistique. Nous souhaitons ainsi mieux comprendre la perception des archivistes par rapport aux nouveaux types d'archives qui découlent de l'usage des technologies Web dans un contexte organisationnel. Par la suite, nous voulons évaluer les écarts constatés entre les pratiques établies dans les milieux étudiés par rapport aux pratiques recommandées dans les politiques d'archivage repérées en ligne. La discussion de ces points nous permettra de dégager, dans une troisième section, un modèle de politique type pour l'archivage de sites Web organisationnels.

5.1 Le site Web en tant que nouvel objet archivistique

La valeur archivistique d'un site Web n'est pas toujours admise et l'archivage du Web organisationnel n'est pas encore au stade d'une pratique généralisée. Ceci nous amène à examiner, dans la présente section, le statut d'un site Web organisationnel en tant qu'objet documentaire ayant une possible valeur archivistique et à discuter des pratiques actuelles de son archivage.

5.1.1 Le site Web en tant qu'objet documentaire : les particularités

Les données obtenues de l'observation des sites Web permettent de souligner le caractère complexe et éphémère d'un site Web (Bachimont et al., 2005; Ghitalla et Boullier, 2004; Koehler, 1999; Leleu-Merveil, 2004; Thelwall, 2004). En effet, les données de recherche démontrent que les composantes d'une page Web peuvent être réparties dans différents fichiers et sa mise en forme gérée indépendamment du contenu grâce à une feuille de style. Les résultats révèlent aussi qu'un site Web est un mélange de formats documentaires issus de divers modes de création. Les pages Web constituant un site Web, même celles qui sont statiques, offrent de multiples fonctionnalités d'affichage et de navigation ce qui explique le recours à une myriade de programmes et d'applications afin de visualiser le contenu informationnel ou exécuter les fonctionnalités du Web. Un site Web n'est pas seulement un objet documentaire²⁵ investi de sens, un objet conceptuel dans l'optique de Thibodeau (2002). C'est aussi une inscription de signes (*bits*) sur un médium (objet physique) reconnue par une application logicielle (objet logique). Seule la coexistence de ces deux éléments permet l'intelligibilité de l'objet conceptuel (Thibodeau, 2002).

Les sites Web observés, comme confirmés par les participants à notre recherche, font aussi l'objet de mises à jour fréquentes, souvent irrégulières, qui touchent, dans certains cas, l'ensemble du site, dans d'autres, des contenus bien particuliers (ex. : mise à jour des règles de conservation disponibles sur le site d'un service d'archives, mise à jour de la section Communiqués disponible sur un site Web gouvernemental). De même, les pages Web observées se présentent dans une structure tentaculaire formée d'un tissu riche de liens internes et externes. Ces deux caractéristiques font du site Web un objet documentaire dont les frontières sont difficiles à établir dans le temps et dans l'espace (Brown, 2006; Masanès, 2005). Le Web impose sa propre logique éditoriale et communicationnelle qui affecte profondément les attributs conventionnels des documents (ex. : fixité) et complique l'identification de leurs frontières spatio-temporelles. Les sites Web sont donc de

²⁵ Nous utilisons cette expression, car le site Web représente un ensemble interrelié de documents de nature diverse. Nous le considérons comme un « document conteneur ».

nature évolutive (changement constant dans le temps) et distribuée (composantes réparties dans différents systèmes et espaces de conservation). Le caractère dynamique des sites Web leur confère un aspect inachevé. De même, l'interaction entre ses documents (liens internes et externes) ne facilite pas leur délimitation en tant qu'unité documentaire. L'unité d'archivage est, par conséquent, difficile à établir et surtout à fixer à l'avance (Bachimont et al., 2005).

Dans un contexte organisationnel, ces caractéristiques ont des conséquences très importantes. Le caractère éphémère des documents et l'effacement graduel du lien entre l'action et le document qui en découle ainsi que du lien entre le document et l'organisme qui le produit (ex. : documents diffusés sur le Web 2.0) ne facilitent pas non plus la reconnaissance de la valeur archivistique que pourrait présenter le site Web public d'une organisation (Thibodeau, 1999-2000).

D'un point de vue archivistique, la nature complexe d'un document Web réside, en fait, dans (1) son caractère dynamique et la dispersion de ses composantes et (2) l'incapacité éventuelle sur le long terme à le reconstituer dans l'état où il était au moment de sa capture (Duranti, 2004; ISO, 2001a). La *Loi sur concernant le cadre juridique des technologies de l'information* impose, à cet effet, la capture de toutes les composantes logiques afin de maintenir l'accessibilité à ses composantes physiques et la conservation de sa mise en forme et fonctionnalités :

« Un document technologique, dont l'information est fragmentée et répartie sur un ou plusieurs supports situés en un ou plusieurs emplacements, doit être considéré comme formant un tout, lorsque des éléments logiques structurants permettent d'en relier les fragments, directement ou par référence, et que ces éléments assurent à la fois l'intégrité de chacun des fragments d'information et l'intégrité de la reconstitution du document antérieur à la fragmentation et à la répartition. » (*Loi concernant le cadre juridique des technologies de l'information*, chap. C-1.1, art. 4)

En ce sens, l'archivage du site Web en tant qu'unité documentaire, c'est-à-dire la collection des pages Web qui le forment, ne suffit pas pour le reconstituer tel qu'il était au moment de son archivage ou garantir sa lisibilité et son accessibilité dans le temps (Barry, 2004; Stephens, 2001). D'autres éléments le constituant tels

que les documents relatifs à ses opérations de gestion, la carte du site Web, la documentation technique relative aux programmes et logiciels utilisés pour produire ou visualiser son contenu informationnel, les journaux de modifications et les métadonnées doivent être capturés (Public Record Office Victoria, 2007: 13-15; NARA, 2005: 5-7; NSW, [s.d.]; Queensland State Archives, 2006b). Ces éléments permettent de préserver la structure du site Web, son contenu ainsi que les fonctionnalités qu'il offre au moment de sa capture.

Le site Web en tant qu'objet documentaire de nature numérique devrait être soumis aux mêmes règles de gestion des documents numériques. D'ailleurs, les politiques d'archivage analysées prescrivent l'archivage des sites Web tout en mettant l'accent sur l'importance de documenter les systèmes et applications ayant servi au développement et à la gestion du site Web (méta-archivage ou archivage de la documentation et des opérations de gestion relatives au site Web). La littérature professionnelle recommande, à cet effet, l'adoption du format WARC qui est un format normalisé pour contenir, sous un même format de fichier, tous les éléments d'un site Web archivé y compris les métadonnées qui les décrivent (Internet Archive, 2012a).

Le site Web est un objet documentaire composite qui subit des transformations graduelles et continues tout au long de son cycle de vie. Que faut-il archiver et à quel moment du cycle de vie du site Web? Les différentes mises à jour donnent naissance à plusieurs versions d'un même site. Ce qui nous amène à poser d'autres questions. À quel moment peut-on considérer qu'une modification est significative pour qu'elle puisse donner lieu à une nouvelle version du site, et combien de versions faut-il conserver? Selon quels besoins, pendant combien de temps et dans quel(s) format(s)? Voilà le genre de questions auxquels l'archivage du Web commande de répondre.

Il faut à cet effet avoir une bonne compréhension des caractéristiques technologiques d'un site Web, de son cycle d'évolution, des objectifs qu'il dessert et des impacts de son usage. Ces éléments forment les composantes à prendre en compte lors de l'analyse des risques associés au site Web, et ce, afin de recommander des

mesures appropriées de gestion des documents et, par conséquent, d'atténuer ces risques.

Cette démarche d'analyse des risques est bien connue dans le domaine de la gestion des archives et des collections (pour un exemple d'étude, voir Lawrence et al., 2000), et ce, pour les archives, peu importe leur support. Elle s'appuie sur des techniques d'évaluation et des mesures de contrôle des risques (BAC et al., 1999). Avant toute intervention, un archiviste devrait, entre autres, avoir une connaissance approfondie des activités et processus de son organisation, identifier les documents qui en découlent et évaluer les systèmes d'information et les systèmes de gestion documentaire existants (Couture, 1999). Ces connaissances doivent être régulièrement mises à jour. Dans un jargon archivistique plus classique, il s'agit de la fonction d'analyse des besoins. Lorsqu'elle est appliquée à l'échelle du site Web de l'organisation, elle permet à l'archiviste d'acquérir une bonne connaissance des documents créés ou reçus par le biais du système Web et des activités qu'ils documentent ainsi que de la situation générale de sa gestion afin de prendre des décisions éclairées en matière de gestion des archives Web (Couture, 1999: 14-15).

L'analyse des risques associés à la gestion et à l'usage du site Web public d'une organisation permet de déterminer les moments opportuns pour procéder à des captures du site Web, de spécifier le niveau de granularité du site Web qu'il faudrait viser et le niveau de pertes à accepter. Concernant ce dernier élément, nous savons qu'un site Web se distingue par une structure tentaculaire. De ce fait, il faudrait décider si, lors des collectes, les ressources qui sont pointées grâce à des liens externes seront également archivées ou non. L'analyse des risques permet ainsi d'entrevoir les conséquences de leur non-archivage ou de déterminer les démarches à faire avant de procéder à leur collecte. L'obtention des autorisations auprès des organismes responsables des pages auxquelles un site Web renvoie est recommandée afin, non seulement de respecter les droits de propriété intellectuelle, mais aussi de pouvoir communiquer et diffuser ultérieurement les archives constituées (BAnQ, 2012).

En transposant cette démarche dans le contexte Web, il apparaît en toute évidence que l'analyse des risques liés à l'usage du site Web est un préalable à la définition d'une stratégie pour son archivage. Malheureusement, il ne s'agit pas d'une pratique courante dans les milieux étudiés et, lorsqu'elle est mise en pratique, les choix sont établis par les webmestres. Ces choix ne correspondent pas toujours aux besoins de gestion documentaire.

Les archives Web doivent être identifiées, capturées dans un système d'archivage et préservées lorsqu'elles présentent une valeur permanente. Les archivistes doivent, plus que jamais, être impliqués dans le processus de développement et de gestion des sites Web. Le Web, de par sa nature, impose, toutefois, un processus d'évaluation continue et, par conséquent, un processus d'archivage en continu. Nous pouvons conclure que la nature inachevée des sites Web oblige à adapter les façons de faire mais ne remet pas en question les principes de gestion documentaire. Ainsi, les expériences et réflexions actuelles en matière de gestion des sites Web organisationnels ont permis de développer un cadre spécifique pour l'analyse des besoins de gestion des sites Web, des stratégies d'archivage bien adaptées aux différents types des sites Web et un format d'archivage qui préserve leurs caractéristiques. Ces façons de faire évolueront grâce aux retours d'expériences, aux efforts de collaboration dans la communauté archivistique et à la recherche scientifique.

Les résultats de notre recherche permettent d'ailleurs d'enrichir les composantes du modèle de gestion des risques associés aux sites Web. Le modèle initial vise principalement l'évaluation des niveaux de risques qui découlent des caractéristiques technologiques du site Web et des usages qui en sont faits dans un contexte organisationnel. Nous proposons l'ajout d'une composante qui porte sur la nature des documents diffusés sur les sites Web publics d'une organisation. Le choix d'une stratégie d'archivage dépend, en effet, de l'originalité des documents diffusés sur un site Web, de leur complexité et des liens qu'ils présentent avec les activités de l'organisation (ex. : les documents ayant un lien direct avec la mission de l'organisme

n'ont pas la même valeur que ceux produits à des fins de référence ou d'information).

Cette composante oblige à tenir compte :

- Du mode de création des documents diffusés sur le site Web afin de déterminer si les documents sont produits uniquement sur le site Web ou s'ils sont produits et gérés dans d'autres systèmes;
- Du potentiel d'accessibilité et d'exploitabilité des documents archivés dans le temps dans l'objectif d'identifier les documents produits dans un format propriétaire et d'envisager les solutions de migration vers des formats normalisés;
- Du lien entre les documents diffusés sur le Web et les fonctions de l'organisme afin d'indiquer ceux qui sont produits dans le cadre des activités propres de l'organisme et de prévoir les mesures nécessaires à leur capture et préservation.

Cette analyse permet aussi d'établir un inventaire des documents diffusés sur le site Web, d'identifier ceux qui sont le produit propre du site Web et les intégrer dans un calendrier de conservation, d'entrevoir la migration de certains documents vers des formats plus pérennes. Cette couche d'analyse précisera davantage les risques appréhendés et mènera à un choix de stratégie(s) d'archivage appropriée(s), non seulement aux caractéristiques technologiques des sites Web, mais aussi à la nature des archives Web. Par exemple, si une section du site Web fait l'objet de mise à jour fréquente, les documents déposés dans cette section seront alors de nature dynamique et probablement reliés à une fonction vitale dans l'organisme. L'archiviste pourrait devoir augmenter la fréquence de capture de cette section du site Web ou capturer uniquement les documents faisant l'objet de cette mise à jour. À titre d'exemple, The University of Melbourne a choisi de développer une liste pour ce genre de documents et programmer le robot de collecte à capturer ces documents lorsqu'il détecte une nouvelle mise à jour (The University of Melbourne, 2008g).

5.1.2 Le site Web en tant qu'objet archivistique : niveau de granularité à considérer

Le Web organisationnel ne sert pas uniquement d'entrepôt de documents à diffuser au sujet de l'organisation. Il s'agit aussi d'un espace où différentes transactions et interactions se produisent entre l'organisation et ses clients ou

partenaires (Government of Western Australia, 2002). Les résultats de l'observation le confirment : bon nombre des documents observés (70%) sont des documents de fonction, c'est-à-dire des documents relatifs à la mission propre de l'organisation. Le site Web (collections de pages Web) et la trace des différentes opérations Web (ex. : publication d'un document, réalisation d'une transaction, retrait d'une page, mise à jour d'une autre) constituent des archives et doivent être préservés pour des besoins de justification de droits et de témoignage (Brown, 2006).

Maintenant, reste à déterminer le statut d'un site Web : est-il une série, un dossier ou un document? À ce niveau, les politiques d'archivage ne spécifient pas explicitement son statut. En plus du macro-archivage (ensemble du site Web) et du méta-archivage (documentation des opérations de gestion du site Web), nous avons remarqué dans quelques sites d'étude, que les archivistes procèdent à un micro-archivage où des documents particuliers sont gérés dans d'autres systèmes documentaires avant leur publication sur le site Web (ex. : SIW02 et SIW09). Les politiques analysées abondent dans ce sens et suggèrent même l'établissement de règles de conservation propres à ces documents. Nous avons remarqué que dans les milieux étudiés, lorsque des portraits des sites Web sont réalisés, les archivistes n'établissent pas de règles de conservation propres à ces documents ce qui engendre, dans plusieurs cas, une accumulation de ces portraits sur des supports de stockage souvent non appropriés (CD-ROM ou DVD) dans l'incertitude absolue que ces documents puissent être préservés ou exploités à long terme. Nous suggérons d'intégrer les portraits des sites Web au calendrier de conservation et d'en spécifier le format et le support d'archivage ainsi que le sort final.

Les quelques règles de conservation recueillies lors des entrevues placent le site Web au niveau d'une série²⁶. Ceci nous semble tout à fait pertinent puisque le site Web englobe des documents en lien avec différentes activités, mais qui sont diffusés à des fins d'information et de transparence. Ainsi, il est au cœur de la stratégie

²⁶ La série est « une division de pièces ou d'articles couvrant les aspects d'une même fonction, activité ou sujet, à l'intérieur d'un fonds donné et définie par un cadre de classification. » (Rousseau et Couture, 1994: 121)

communicationnelle de l'organisation. La série « Site Web » comporte les composantes suivantes :

1. Le portrait du site Web (portrait du site Web ou copie à partir du répertoire racine).
2. Les documents et pages individuels (les documents générés de manière dynamique que le robot de collecte n'est pas capable de capturer, dans le cadre d'une fonction spécifique ou ceux qui évoluent de manière différente des autres documents du site). Exemple : la page d'accueil du site Web, les formulaires en ligne, les rapports des bases de données qui alimentent le site Web.
3. Les journaux historiques.
4. Les journaux de transactions.
5. Les documents et composantes traitant, de façon générale, de la gestion d'un site Web.
 - a. Les documents concernant les projets de développement du site Web, l'analyse des projets ainsi que les différentes étapes d'implantation de ceux-ci.
 - b. Les documents relatifs à la structure du site Web (ex. : la carte du site Web), son architecture et sa présentation visuelle.
 - c. Les documents relatifs aux informations diffusées sur le site Web (ex. : processus de publication, inventaire des documents diffusés sur le site Web, statistiques) et à la protection des droits d'auteurs.
 - d. Les documents relatifs aux moteurs de recherche, à l'indexation, aux hyperliens et au référencement du site.
 - e. La documentation et les fichiers de configuration relatifs aux programmes et logiciels utilisés par le site Web.
 - f. La documentation relative au robot de moissonnage.
 - g. Les documents relatifs à la perception du public face au site Web permettant l'évaluation de ce dernier : suggestions, commentaires, réactions du public, etc.
 - h. Les métadonnées descriptives des pages Web composant le site Web.
6. Les anciennes versions du site Web ou celles ayant subi une refonte majeure.
7. Les microsites (ou sites Web thématiques) et les sites Web événementiels associés au site Web public d'une organisation.

Cette liste correspond aux unités d'archivage identifiées lors de l'analyse de la recension des écrits et des politiques d'archivage (Barry, 2004; NAA, 2001a: 11-12; NARA, 2005: 5-7, 19-20; NSW, [s.d.]; Public Record Office Victoria, 2007: 13-15; Queensland State Archives, 2006b; TNA, [s.d.]a; Zwarich et Park, 2008: 5) et des règles de conservation obtenues des milieux d'étude. Cette liste a été complétée par l'analyse du modèle de règles de conservation proposé par le NARA (2005: 19-20) et les règles proposés sur le site Web de BAnQ (2011) et de la CREPUQ (2002).

Ce point nous amène à discuter également du sort réservé aux contenus organisationnels diffusés sur le Web social. À qui revient la responsabilité d'archivage de ces contenus? Est-il possible de procéder à la collecte de contenus qu'un organisme diffuse sur des plateformes gérées par un autre organisme? Les archivistes doivent être attentifs à l'usage fait des applications du Web 2.0, aux enjeux et risques auxquels s'exposent les organisations en publiant de l'information sur ces plateformes (CEFRIO, 2012; Dufour, 2008-2009). Afin de protéger la réputation de l'organisation et prévenir la perte d'informations, les archivistes doivent, dans une démarche proactive, identifier les documents publiés uniquement sur le Web social et ceux qui forment une duplication des documents existants dans d'autres environnements de production documentaire (NSW, 2009b: 7). Ces documents relatifs à une fonction ou à une activité doivent être retenus dans le calendrier de conservation. Des politiques et procédures doivent être établies afin de contrôler le processus de publication des informations sur ces applications et définir une stratégie pour leur archivage (ARMA International, 2011; NSW, 2009b). Les politiques suggèrent de faire appel à des applications maison (*in-house technology*) afin de s'assurer que le contenu publié demeure sous le contrôle de l'organisme producteur et de faciliter son processus d'archivage. Cette solution, si intéressante soit-elle, requiert un investissement important en ressources et en compétences. Des recherches doivent être entreprises afin d'étudier cette problématique émergente.

5.1.3 Le site Web comme lieu de production de nouveaux types d'archives

Les résultats de notre recherche confirment que l'usage des technologies Web dans les organisations provoque la migration progressive des documents existants sur la toile (ex. : rapports annuels, outils archivistiques, formulaires), notamment ceux qui renseignent sur la mission et les activités propres à l'organisme (ex. : supports d'apprentissage, rapports d'événement, documents relatifs aux activités de formation, de recherche ou de prévention dans un domaine d'intervention précis). Au cours de ce processus, ces documents acquièrent de nouvelles caractéristiques héritées d'un

nouveau mode de production et de diffusion des documents propre au Web (Bachimont et al., 2005; Masanès, 2006; Yates et Orlikowsk, 1992). L'usage des technologies Web est également à l'origine de nouveaux types de documents qui ne se retrouvent sur aucun autre support ou média d'information (à commencer par le site Web lui-même) et qui ne peuvent pas être compris en dehors de leur contexte de connectivité (Bachimont et al., 2005; Shepherd et Watters, 1998).

Rappelons que les archives Web sont tous les documents produits ou reçus par un système d'information Web dont certains n'ont pas d'équivalents dans d'autres environnements de production documentaire. L'observation des archives Web et l'analyse des politiques confirment cette assertion. Outre la carte du site Web, les portraits périodiques et les documents relatifs à la gestion d'un site Web, les politiques Web analysées n'offrent pas d'autres exemples d'archives Web. Toutefois, lors de l'analyse des données, nous avons pu identifier de nouveaux types d'archives, tels :

- L'historique du site Web;
- La page d'accueil et les pages d'orientation (sous-pages d'accueil);
- Les agendas en ligne;
- Les archives en ligne d'un site ou de certains contenus du site Web (ex. : les communiqués) ;
- Les cartes interactives;
- Les expositions virtuelles;
- Les formulaires en ligne;
- Les questions fréquemment posées (FAQ).

Ces documents se présentent généralement en format Web. Leur mise en forme et structure est différente de celles produites dans un autre environnement numérique. Ils sont créés dans un cadre bien spécifique : documenter l'évolution du site Web public de l'organisation, documenter une activité ou un événement, offrir de l'information sur une activité ou un produit, réaliser des transactions en ligne ou faciliter l'expérience de navigation de l'utilisateur à travers le site Web. Malgré leurs

caractéristiques, ces documents ne sont pas intégrés dans aucun des calendriers de conservation utilisés dans le cadre de cette recherche.

Les nouveaux types de documents issus des systèmes Web doivent être identifiés et intégrés au calendrier de conservation s'ils présentent une valeur archivistique individuelle, voilà ce que recommandent à ce sujet les politiques d'archivage. Le cadre d'analyse des risques liés au site Web enrichi par la composante Nature des documents que nous proposons servira à établir la liste de ces nouveaux types de documents, à analyser les valeurs qu'ils présentent et à choisir une stratégie appropriée pour leur archivage.

Ceci étant dit, l'identification et l'inventaire des documents diffusés ne sont pas suffisants. Il faudrait également identifier les composantes à capturer afin de pouvoir les reconstituer et préserver leurs qualités en tant que documents d'archives. Nous avons noté lors de l'observation des archives Web que certains documents ne comportent pas tous les éléments contextuels qui facilitent leur identification ou gestion. Ce constat s'applique surtout aux documents propres au Web. Les documents migrant sur le Web préservent en général leurs conditions de validité en tant que document d'archives. À ce sujet, les politiques d'archivage recommandent, et ce, depuis la création d'un document Web, de le décrire avec des métadonnées complètes et pertinentes. Nous regrettons que cette pratique ne soit pas appliquée et généralisée aux documents Web observés (description minimale et insuffisante des documents). Nous reviendrons sur ce constat et ses conséquences au point 5.2 (p. 191).

En résumé, le site Web est un objet documentaire assez complexe, mais il est évident qu'il présente une valeur archivistique qui découle des finalités du site Web et des usages qui en sont faits. La valeur qu'il présente pour l'organisme producteur et celle qui pourrait servir des besoins d'information et de recherche ne sont pas encore généralement admises. Ce phénomène se produit lorsqu'une nouvelle technologie est adoptée ou lorsque son usage évolue dans le temps (Yates et Orlikowski, 1992; Watters et Shepherd, 1997). La rareté de cas de litiges associés à la diffusion de certains renseignements sur les sites Web organisationnels, le peu d'études scientifiques qui s'intéressent au site Web en tant que matériel de recherche

ainsi que l'absence de cadre de référence pour son archivage pourraient être avancés comme arguments pour expliquer l'implication timide des archivistes dans les projets d'archivage du Web (Barry, 2004; Stephens, 2001; Stirling et al., 2012).

5.2 Archivage du Web : politiques et pratiques

L'analyse des données recueillies dans le cadre de cette recherche laisse transparaître deux niveaux d'archivage du Web, soit l'archivage du Web à l'échelle nationale, soit l'archivage du Web à l'échelle organisationnelle. Nous examinons dans ce qui suit les façons de faire à ces deux niveaux.

5.2.1 Pratiques des archives nationales

5.2.1.1 Modèles de gestion

L'étude des projets d'archivage (voir Annexe 1, p. i) et l'analyse des politiques d'archivage (voir Annexe 2, p. xvii) nous permettent de brosser un portrait général des choix opérationnels adoptés par les archives nationales quant à l'archivage du Web organisationnel, plus précisément du Web gouvernemental. Nous distinguons deux principaux modèles de gestion. Dans le premier, les archives nationales s'impliquent dans le processus d'archivage. Dans ce cas, elles prennent en charge l'archivage des sites Web gouvernementaux (ex. : BAC, 2007-2008), partagent cette responsabilité avec les organismes créateurs de sites Web publics (ex. : TNA, [s.d.]a) ou s'impliquent dans cette tâche en partageant des territoires de collecte avec d'autres institutions de mémoire telles que les bibliothèques nationales (ex. : NLA et NAA, 2002). Dans le deuxième, les archives nationales se contentent de jouer le rôle de conseiller.

Le premier modèle nous semble pertinent surtout dans le cas où certains organismes ne sont pas tenus de verser leurs archives définitives. Même si plusieurs organismes prennent en charge la gestion de leurs archives définitives et s'y investissent au plus haut point vu les intérêts économiques que ces archives pourraient leur apporter (ex. : société de radiodiffusion) ou leur importance dans la

reconstitution de l'histoire de l'organisme dans le temps, rien ne garantit que ce sera le choix pour lequel vont opter d'autres organismes. L'accessibilité de ces archives dépendra aussi des moyens mis en œuvre, et nous pouvons le confirmer, elle ne sera pas toujours dans les priorités d'une organisation. L'archivage des sites Web par les archives nationales est donc une bonne alternative au manque des ressources et de la volonté des décideurs dans une organisation. Ceci n'empêche pas ces institutions de mémoire de conduire des campagnes de sensibilisation visant à promouvoir l'intérêt historique et scientifique des archives du Web organisationnel. Ce premier modèle de gestion suggère aussi un partage des responsabilités selon les valeurs que présentent les archives Web. Les organismes sont tenus de conserver leurs sites Web à des fins administratives alors que les archives nationales le font dans une perspective patrimoniale, le site Web étant une source d'information qui témoigne de l'activité en ligne d'une organisation et qui pourrait servir ainsi à des fins de recherche et d'information. Ce mode opérationnel garantit un accès plus rapide aux archives Web. Les archives Web gouvernementales sont accessibles à partir des sites Web des archives nationales ayant un programme d'archivage du Web bien établi, telles que Bibliothèque et Archives Canada ou The National Archives (Royaume-Uni).

Dans le deuxième modèle, les archives nationales jouent seulement un rôle d'encadrement des opérations d'archivage du Web à l'échelle organisationnelle (ex. : BAC, 1999; NAA, [s.d.]; NARA, 2005). Les organismes doivent mettre en œuvre une démarche d'archivage du Web à l'interne dans le respect des lignes directrices développées par les archives nationales à ce sujet. Dans ce cas, les archives nationales pourraient exiger le versement des archives Web ayant une valeur permanente, ceci dépendant de l'obligation légale de versement à laquelle est assujéti l'organisme et des capacités techniques et logistiques des archives nationales à accepter de tels versements.

L'orientation prédominante actuellement correspond plutôt au deuxième modèle où les archives nationales ont la responsabilité de développer des orientations générales et des outils concernant l'archivage du Web et jouent principalement le rôle

de conseiller (ex. : BAC, 1999b; NAA, 2001a; NARA, 2005). La responsabilité d'archivage d'un site Web revient en premier lieu à l'organisme créateur.

Ces choix ressemblent d'une certaine manière à ceux établis par les bibliothèques nationales. En effet, les bibliothèques nationales ont procédé à l'archivage du Web en absence d'une loi sur le dépôt légal numérique qui encadre leurs actions et, ce afin, de prévenir la perte de ressources informationnelles. Lorsque la loi sur le dépôt légal n'est pas opérationnelle, ces organisations procèdent elles-mêmes à la collecte des sites Web (ex. : BnF) ou comptent sur le dépôt volontaire de la part de fournisseurs (ex. : don des archives de Twitter à la Library of Congress) et créateurs de contenus (ex. : After Life). Les territoires de collecte sont déterminés de la même manière que pour les autres publications, c'est-à-dire, en fonction de la politique de développement des collections (logique de continuité des collections), suite à un processus de veille (ex : collecte des sites Web événementiels) ou grâce à l'implication des usagers dans le processus de sélection (Archive-it, [s.d.]; Schneider et al., 2003; Stirling 2012).

De la même manière, les archives nationales procèdent à l'archivage des sites Web gouvernementaux de manière proactive afin de prévenir la perte d'un morceau considérable de la mémoire des organisations (ex. : dans le cas où un organisme public n'est pas tenu de verser ses archives), en attendant que ces dernières établissent leur propre programme d'archivage. Lorsque ces programmes sont bien établis, les archives nationales partagent alors la responsabilité de l'archivage du Web avec les organismes gouvernementaux ou publics.

Nous avons observé une forme de délégation intra-institutionnelle de l'archivage du Web organisationnel à BAnQ. Il s'agit d'une autre forme de collaboration entre deux unités organisationnelle au sein d'une même institution de mémoire, et ce, en fonction de l'expertise et des moyens disponibles. Dans ce cas, l'archivage du Web gouvernemental et événementiel (ex. : campagnes électorales de 2012, site Web de la commission Bouchard Taylor) se fait sous le chapeau du dépôt légal (BAnQ, 2012, Nicoud, 2012). Cette section conclut des ententes avec les organismes gouvernementaux et/ou publics afin d'obtenir l'autorisation d'archiver

leurs sites Web publics. Elle a également procédé à la récupération de manière rétrospective auprès de BAC des sites Web d'intérêt pour le Québec. La direction des archives au sein de BAnQ assume plutôt un rôle de conseiller auprès des organisations en leur proposant trois règles de conservation concernant les systèmes Web qu'elles produisent (BAnQ, 2011, 2012). La première, la règle 6409, définit le cycle de vie des documents relatifs à la conception des systèmes Web ainsi qu'à leurs spécifications techniques et graphiques. La deuxième, la règle 6410 porte sur les documents relatifs à la production et à l'implantation des systèmes Web. La dernière, la règle 6411, couvre les documents relatifs à la gestion et à l'administration quotidienne des systèmes Web en place ainsi qu'à leur fonctionnement.

L'approche d'archivage adoptée par BAnQ se reflète dans ses règles de conservation. Nous notons l'absence d'une règle qui porte sur les sites Web eux-mêmes en tant qu'unité documentaire à archiver, la collecte et la préservation des sites Web organisationnels étant de la responsabilité de la section du dépôt légal.

Ce modèle de gestion est à la fois original et intéressant. Nous regrettons le manque de visibilité des travaux réalisés par BAnQ au sujet de l'archivage du Web. Les projets entrepris par les bibliothèques nationales et quelques archives nationales sont bien documentés. Ces institutions publient, grâce à un site Web dédié au projet, une description des travaux entrepris, une présentation des outils développés ou utilisés et offrent l'accès aux collections constituées (voir Annexe 1, p. i).

La diffusion de ces informations permet de promouvoir le programme d'archivage du Web, d'attirer l'intérêt du public et de faire découvrir un nouveau matériel pour la recherche scientifique et historique (CRL, 2002-2005; Gomes et al., 2011; Groupe d'études sur l'usage des TIC dans les migrations, 2005; Internet Memory foundation, 2010; Nivard Bigourdan, 2005; Stirling et al., 2012). Grâce à l'interaction avec les usagers actuels et potentiels des collections Web, les archivistes peuvent améliorer les méthodes actuelles de recherche dans les archives Web (ex. : développer des collections thématiques) (Stirling et al., 2012). Cette démarche est bien intégrée dans les programmes d'archivage de quelques bibliothèques nationales

et universitaires (BnF, 2012b; British Library, [s.d.]; California Digital Library, 2008-2012; Harvard University Library, 2008; Schneider et al., 2003).

5.2.1.2 Le statut des sites Web dans les projets d'archivage du Web menés par des archives nationales : quelques réflexions

L'archivage du Web par les archives nationales s'effectue au niveau du site Web, c'est-à-dire que ce dernier est considéré comme unité d'archivage. Les actions menées par les archives nationales nous permettent de confirmer que les sites Web organisationnels détiennent, dès leur création, une double valeur archivistique, soit une valeur primaire et une valeur secondaire. Ces deux valeurs se superposent, le site Web étant archivé à des fins administratives en même temps que pour sa valeur patrimoniale. Ainsi, la valeur secondaire n'émerge pas seulement au déclin de la valeur primaire d'un document d'archives (Couture, 1999; Upward, 2005), mais peut être présente tout au cours de la vie du document. Rappelons que le site Web se distingue des autres archives par son évolution continue, par un statut de document inachevé dans le temps et l'espace (Bachimont et al., 2005; Masanès, 2006; Pédaque, 2006). Ainsi, les deux niveaux d'archivage (archivage à des fins administratives et archivage à des fins patrimoniales) peuvent s'opérer en parallèle. Nous croyons que la nature même des sites Web nous conduit vers le choix d'une responsabilité partagée au niveau de leur archivage où les archives nationales peuvent procéder à l'archivage des sites Web organisationnels à distance selon une fréquence régulière (généralement, on recommande deux captures par année) qui pourrait être adaptée au besoin (ex. : événement particulier ayant lieu et qui a un impact sur le contenu du site Web). Les organismes auront, quant à eux, la responsabilité d'un archivage à des fins administratives. Ceci exige la capture d'autres unités d'archivage, soit le site Web et/ou des documents particuliers diffusés/produits par le site Web ainsi que les documents qui découlent des opérations de sa gestion.

L'archivage du Web organisationnel change en quelque sorte les processus traditionnels de gestion des archives où les archives doivent servir d'abord à des fins administratives, ensuite, à des fins de témoignage. Les documents ne sont

généralement accessibles qu'à la fin de leur utilité administrative et après avoir été traités par le service responsable de leur conservation permanente. Or, les archives des sites Web servent ces deux fins en parallèle. Nous reprenons ici l'exemple des archives Web du gouvernement canadien ou britannique qui sont consultables en ligne dès leur capture sans devoir atteindre la fin de leur utilité administrative. À l'encontre des bibliothèques, les archives nationales n'ont pas été dans l'obligation de limiter l'accès aux archives Web pour deux principales raisons. Tout d'abord, l'observation des sites Web, dans le cadre de notre recherche concluent que les sites Web organisationnels (1) ne servent pas à diffuser des renseignements personnels ou sensibles et, si c'est le cas, (2) les archives nationales incitent les organismes, dans leurs politiques d'archivage, à signaler les pages à exclure du champ de collecte des logiciels moissonneurs ou leur recommandent d'ajouter, lors de la conception du site Web, une balise qui interdit au robot moissonneur la collecte qui pages comportant des données sensibles. Cette recommandation doit être généralisée et appliquée, car malheureusement cette pratique ne semble pas être courante dans les sites étudiés.

Le Web, source de maints maux en gestion documentaire, nous offre son infrastructure même comme solution d'archivage (Nelson et al., 2007; Secrétariat du Conseil du Trésor du Canada, 2012). Ainsi, qu'il soit archivé par l'organisme créateur ou par les archives nationales, le site Web archivé peut demeurer accessible. Une section du site Web public d'un organisme peut accueillir les archives de ce dernier. Cette solution peut être envisagée pour offrir un accès quasi instantané aux versions archivées d'un site Web et, aussi, pour régler les problèmes de préservation à long terme causé par un choix inapproprié de supports d'archivage des sites Web. Nous avons noté, au cours de cette recherche, que l'archivage des sites Web sur CD ou DVD, pratique à laquelle se livrent quelques sites étudiés, ne semble pas un choix qui assure un accès et la pérennisation de l'information archivée sur le long terme (TNA, 2009b).

5.2.2 Pratiques organisationnelles

5.2.2.1 Modèles de gestion

Trois principaux modèles de gestion se dégagent de l'analyse des pratiques d'archivage dans les 11 sites d'étude : la copie de sauvegarde (n=11), la copie locale du site Web (n=5) et le moissonnage à distance du site Web (n=1).

La moitié des sites d'étude ont adopté une stratégie d'archivage basée sur la capture d'une copie locale du site Web (archivage à la source), alors que les politiques d'archivage analysées mettent l'accent sur les avantages d'un archivage à distance. La technique d'archivage à distance épargne aux organisations les problématiques liées à la pérennisation des archives Web souvent stockées sur des CD-ROM ou DVD. Les archives Web peuvent être capturées dans un format normalisé, le format WARC, qui garantit non seulement la pérennité des composantes du site Web, mais aussi leur accessibilité et exploitabilité (ex. : mise en ligne des archives Web avec une interface personnalisée).

Les politiques d'archivage du site Web public des organismes à l'étude ne sont pas développées, les rôles et responsabilités ne sont pas clairement définis et l'analyse des risques n'est pas conduite pour soutenir et encadrer les pratiques observées. Or, la plupart des participants à notre recherche affirment que la démarche générale pour l'archivage ne devrait pas différer de celui des documents inscrits sur d'autres supports documentaires même s'ils reconnaissent le caractère complexe d'un site Web. Les principes restent les mêmes, mais les méthodes changent, nous disent-ils. Rares sont ceux qui sont impliqués dans le processus de développement et de gestion des sites Web publics de leur institution.

L'identification des différents éléments composant le site Web (objets documentaires) et essentiels à son fonctionnement (programmes, scripts), leur localisation au niveau de divers espaces documentaires, la précision de leur nature (document principal, document secondaire, copie de diffusion) et leur évaluation archivistique sont des actions de base que l'archiviste devrait poser peu importe le support de diffusion des documents qu'un organisme produit.

Pour pouvoir le capturer, le gérer et le conserver, un site Web doit non seulement respecter certaines normes d'édition Web, mais aussi, être bien identifié et décrit par des métadonnées. Les sites Web archivés respectent certaines normes d'édition, mais ils ne semblent pas être décrits adéquatement (peu de métadonnées pertinentes sont associées aux pages Web observées). L'attribution des métadonnées pertinentes pour la description du contenu et de la technologie permettant sa gestion facilite non seulement le processus de capture des contenus et leur exploitation, mais également leur processus de préservation à long terme (Brown, 2006). À cet effet, l'archiviste doit être impliqué dans le processus de création des sites Web afin d' :

- orienter les responsables de la gestion de l'infrastructure technologique de l'organisme vers le choix de technologies de développement et des formats de documents normalisés et pérennes;
- collaborer avec les webmestres lors du développement des règles de nommage et de description des pages Web;
- inciter les webmestres à documenter les opérations de gestion des sites Web et à bien gérer les documents qui en résultent.

Les politiques analysées soulignent aussi l'importance de bien gérer les URL considérées identifiant unique d'une page Web et de définir une stratégie de redirection pour les contenus qui ne sont plus d'actualité, mais auxquels les gens se réfèrent à des fins d'information.

Actuellement, les solutions techniques et technologiques ne manquent pas et ne semblent pas constituer une difficulté impossible à surmonter. Des outils libres et gratuits sont disponibles à cette fin et sont appropriés pour une capture à distance des sites Web (IIPC, 2012b). Quant aux documents relatifs aux opérations de gestion des sites Web, ils peuvent être localisés au niveau du répertoire où sont stockés les fichiers qui forment le site Web ou sur le serveur institutionnel. Ces documents sont relativement faciles à gérer puisqu'ils sont produits en grande majorité en formats papier ou numérique et non en format Web. Par contre, les sites Web dynamiques présentent un autre niveau de difficulté du fait qu'ils soient associés à une base de données ou un système de gestion de contenus. La *Loi concernant le cadre juridique des technologies de l'information* (chap. C-1.1, art. 3) considère comme document :

« toute banque de données dont les éléments structurants permettent la création de documents par la délimitation et la structuration de l'information qui y est inscrite ». Donc, l'archivage de la base de données est une obligation; reste à réfléchir, dans le cadre d'une autre étude, sur le rapport entre la base de données et le site Web et à définir les éléments à archiver ainsi que les solutions à envisager.

La capacité de reconstitution d'un site Web tel qu'il était au moment de sa capture est moins considérée lorsqu'un site Web est archivé à des fins du dépôt légal. Nous l'avons noté lors de l'analyse des initiatives d'archivage du Web dans une perspective documentaire où la perte de certains éléments de forme ou de contenu d'un site Web est tolérée. Les archives, par contre, doivent satisfaire des conditions d'authenticité, d'intégrité, de fiabilité et d'exploitabilité (Bachimont, 2000; Bourhis, 2000-2001; ISO, 2001a). Ces critères sont rappelés dans les politiques analysées. Ces dernières recommandent, à ce sujet, l'application des règles d'édition Web, l'adoption de formats ouverts, l'assignation de métadonnées complètes et pertinentes afin de documenter la technologie Web utilisée et les opérations effectuées sur le site Web (ex. : gestion et maintenance du site Web) afin de pouvoir documenter les modifications apportées au site Web, mais aussi pouvoir le reconstituer. La dispersion des composantes d'un site Web ne représente pas une contrainte majeure. Elle peut être maîtrisée en procédant à un archivage à distance. L'URL d'une page Web ou le domaine d'enregistrement d'un site Web devient dans ce cas l'identifiant unique qui permet de circonscrire les frontières d'un document dans un environnement Web. En le paramétrant, le robot de collecte aspire l'ensemble des composantes du site Web sans se soucier de leur localisation. De même, ces robots ont la capacité de convertir les liens absolus en des liens relatifs. Une copie navigable en local pourra ainsi être constituée dans un format d'archives Web normalisé, le format WARC (Groupe de coordination 46, 2009; Strodl, 2009). Ceci explique pourquoi les politiques d'archivage dans une perspective organisationnelle recommandent la technique d'archivage à distance même si les fichiers Web sont accessibles à la source.

5.2.2.2 Le statut des sites Web dans les sites d'études : les différentes perceptions

Les bibliothécaires ont très rapidement compris que le Web offre des matériaux intéressants pour la recherche et l'histoire. Ils se sont investis dans la collecte des publications en ligne depuis les débuts du Web, même si la loi sur le dépôt légal, dans certains pays, ne couvre pas ce genre de publications. Dans un souci d'exhaustivité et devant des frontières floues entre les différentes catégories de documents sur le Web, certaines bibliothèques nationales et universitaires incluent, dans leurs collections, des contenus qui relèvent plus du domaine des archives (ex. : les blogues sont souvent considérés comme des journaux intimes. Ces journaux se retrouvent habituellement dans les services d'archives et non dans la collection d'une bibliothèque) (Zeller, 2011). Les bibliothécaires se sont rapidement organisés et ont misé sur une approche collaborative interinstitutionnelle afin de surmonter les nombreux défis (Chaimbault, 2008; Illien, 2009). Le partage des collectes se fait dans une logique de continuité des collections, selon les champs d'intervention, l'expertise et les ressources des institutions.

Des archives nationales ont procédé à l'archivage, notamment celui des sites Web gouvernementaux (ex. : BAC, 2007-2008; TNA, [s.d.]a), d'autres ont établi des politiques nationales afin de responsabiliser les organismes gouvernementaux/publics et les orienter vers les meilleures pratiques d'archivage de leur Web organisationnel (ex. : NAA, 2001a; NARA, 2005; TNA, 2006, 2011). Certaines politiques spécifient le domaine d'intervention de chaque institution de mémoire tout en privilégiant une approche collaborative (ex. : NLA et NAA, 2002) ou établissent les critères qui permettent de distinguer les documents d'archives des publications (ex. : Queensland State Archives, 2008). Un site Web, lorsqu'il est produit dans le cadre de la mission propre d'une organisation et afin de satisfaire ses mandats a le statut d'archives. Ce lien n'a pas été aussi évident à établir pour plusieurs des archivistes interviewés. Même si les politiques d'archivage du Web reconnaissent à l'unanimité la valeur archivistique des sites Web organisationnels, les participants à notre recherche ne partageaient pas tous ce point de vue. Nous examinons dans les paragraphes suivants

les différentes perceptions du Web organisationnel des personnes rencontrées dans les différentes organisations à l'étude.

Bon nombre d'archivistes interviewés considèrent que le site de leur organisation (ou unité) est à la fois un produit et l'outil d'une stratégie de communication que l'organisation voulait mettre en œuvre afin de promouvoir ses activités et services. Ces archivistes estiment également que le site Web public de leur organisation est un support d'une importance capitale qui les soutient dans la réalisation de leurs mandats et tâches (n=5). Cette compréhension du site Web et de ses finalités explique pourquoi l'archivage du Web est envisagé, voire opérationnel dans les organisations où ces participants à l'étude œuvrent. Par contre, un groupe presque tout aussi nombreux n'est pas très convaincu de la nécessité de l'archivage d'un média qui diffuse principalement des documents produits dans d'autres environnements et gérés dans d'autres systèmes d'information (n=3), d'autant plus que, pour un participant, Internet Archive le fait déjà. Les critères d'identification des archives sont, par contre, établis et ne dépendent ni de la forme ni du support d'un document (Bearman et Hedstrom, 2000; Henry, 2000; Turnbaugh, 2002; Yusof et Chell, 1998; Zwarich et Park, 2008). La production du site Web d'une organisation a une fin en soi, il doit être préservé, entre autres, car il témoigne de cette fin. Ces mêmes archivistes ne sont pas responsables de la gestion du site Web de leur unité ou ne sont pas impliqués dans les travaux de l'équipe responsable du site Web organisationnel. Cette distance peut donc favoriser une méconnaissance de son importance. À plusieurs reprises, nous avons remarqué que les archivistes interviewés ne détenaient pas les informations relatives aux aspects techniques du site Web ni à ses composantes documentaires. Ils sont satisfaits, lorsqu'ils l'apprennent, de la solution mise en place par les webmestres, soit les copies de sauvegarde.

Les pratiques de gestion documentaire sont encore imprégnées d'une conception traditionnelle : le document doit être tangible, fixe, ses liens avec les fonctions et activités de l'organisation doivent être évidents. Pourtant, le cadre légal au Québec reconnaît l'écrit numérique (*Loi concernant le cadre juridique des technologies de l'information*, chap. C-1.1, art. 3). Le *Règlement sur la diffusion de*

l'information et sur la protection des renseignements personnels (chap. A-21.1, r.2, art. 5) exige même la diffusion en ligne de certains documents administratifs. Nous nous demandons comment les organisations peuvent prouver qu'elles ont bel et bien appliqué ce règlement en dehors de l'archivage de la portion du site Web où les documents devant être diffusés sont affichés. Ceci nous rappelle qu'au début de l'utilisation des systèmes de messagerie électronique, les messages électroniques n'ont pas été reconnus comme des documents d'archives (Henry, 2000: 572-573). Tel que démontré dans les travaux basés sur la notion de genre, ce n'est que lorsque ces types de documents sont acceptés socialement (ex. : publication des politiques d'archivage des courriers électroniques, reconnaissance juridique de l'écrit numérique), donc reconnus comme document, que leur archivage s'est imposé et généralisé (Shepherd et Watters, 1998; Yates et Summer, 1997).

Nous croyons que la reconnaissance de la valeur archivistique d'un site Web au sein de la communauté des archivistes et auprès des acteurs qui sont impliqués dans sa gestion n'est pas encore bien établie. L'absence de position officielle des archives nationales, l'absence d'une politique nationale en matière du Web à défaut d'un programme de formation continue pourraient expliquer pourquoi les sites Web organisationnels, notamment gouvernementaux et publics, ne sont pas archivés de manière systématique. Aussi, la décision finale d'investir les ressources et les moyens nécessaires pour accomplir l'archivage des sites Web dépend, dans plusieurs cas, de la volonté politique et du degré de persuasion des hauts responsables d'une organisation. Certaines politiques analysées, notamment celles développées par des archives nationales, indiquent que les dirigeants d'une organisation sont responsables de la décision d'élaborer une politique pour l'archivage du Web organisationnel. Ils doivent également veiller à ce qu'elle soit mise en œuvre et respectée.

Dans deux cas seulement, les archivistes ont évoqué les travaux de la CREPUQ sur l'archivage des sites Web comme cadre de référence à considérer dans leur démarche (CREPUQ, 2009). Nous recommandons l'intégration de la problématique d'archivage du Web dans le cursus de formation académique (écoles de formation) et professionnelle (associations). Les associations professionnelles dans

le domaine de la gestion des archives doivent jouer un rôle de sensibilisation à la problématique de conservation des sites Web organisationnels (ex. : organisation d'ateliers, de formations, de manifestations scientifiques sur le sujet), et ce, non seulement auprès des archivistes, mais aussi auprès des organisations où ils œuvrent. Les associations professionnelles devraient également offrir des activités de perfectionnement à leur membre dans ce domaine, développer des modèles d'outils pour l'archivage du Web ou orienter les intervenants dans ce processus vers des sources recommandées dans le domaine. L'IIPC offre des outils d'archivage libres et gratuits de même que la documentation nécessaire pour faciliter leur implantation et usage (IIPC, 2012b).

Dans le cas où une règle de conservation est établie (SIW03), l'archiviste reconnaît que cette règle n'a pas beaucoup de visibilité au sein des unités qui gèrent de manière décentralisée leur site Web. Dans ce genre de situation, nous recommandons la mise à profit des moyens de communication et des technologies Web pour inviter les créateurs de contenus et les gestionnaires des sites Web, d'une part, à appliquer les règles de conception et de gestion des sites et, d'autre part, à systématiser la gestion des sites Web tout au long de leur cycle de vie afin d'obtenir des archives Web de qualité au terme de la démarche.

Certains archivistes participant à cette recherche entendent Internet Archive comme une solution d'archivage et acceptent que des archives du site Web de leur organisation soient sous le contrôle d'un autre organisme. Malgré l'incomplétude de ses collections, Internet Archive est actuellement le seul entrepôt des archives des sites Web organisationnels qui est opérationnel depuis les débuts d'internet (Pennock et al., 2006). Les conséquences de cette paragnosie sont pourtant évidentes. La fréquence et la méthode de capture adoptées par Internet Archive ne représentent pas nécessairement l'évolution du site Web ni les besoins documentaires de l'organisation. Les organisations peuvent négocier auprès d'Internet Archive un archivage rétrospectif des anciennes versions de leurs sites Web, mais ne doivent pas faire de ce service une solution pour l'archivage de leurs sites Web actuels. Ils

doivent mettre en œuvre leur propre programme d'archivage des systèmes Web de leur organisation.

Malgré l'existence de modèles de règle de conservation pour le Web organisationnel au Québec, le manque d'orientation nationale à ce sujet (politique d'archivage) contribue, selon nous, à délaisser l'archivage des sites Web au profit d'autres « priorités » d'intervention des archivistes (ex. : mise en place d'un nouveau système de gestion documentaire, gestion des archives numériques, etc.) et ainsi à s'en remettre à d'autres instances pour effectuer ce travail, ce qui met en péril un fragment important de la mémoire organisationnelle ou expose les organisations à des litiges (ex. : diffusion de données sensibles). Internet Archive pourrait être envisagé lors d'un archivage rétrospectif des versions perdues du site Web public d'une organisation, mais ne pourrait constituer une solution d'archivage à long terme.

En résumé, la perception des archivistes est probablement influencée par une méconnaissance des systèmes Web et par la nature publique du site Web. La disponibilité des documents diffusés sur le site Web dans d'autres systèmes de gestion documentaire dans l'organisation dans un format ou sur un support plus conventionnel pourrait aussi expliquer l'absence d'archivage des sites Web dans certaines organisations. Pourtant, ces documents sont diffusés sur le site Web pour des raisons souvent différentes de celles à l'origine de leur création (ex. : le plan stratégique de l'organisation est une feuille de route que se donne une organisation pour réaliser sa vision à moyen et à long terme. Sur le Web, ce document devient un engagement public de l'organisation envers ses clients et témoigne de la transparence de ses choix et décisions). Le site Web représente un tout autre contexte d'usage des documents. Les politiques d'archivage analysées n'absolvent pas les organisations de l'archivage du site Web et des documents qui y sont produits ou diffusés même si ces derniers sont gérés dans d'autres espaces documentaires et vice-versa :

« If the organization is posting records to the website or intranet that exist in other formats, this does not absolve the organization from managing both the Web version of the record and the original version. The organization's website or intranet can be invaluable as a display space or delivery system for preexisting records, but staff must not

make the mistake of neglecting original records because they have been posted electronically. » (ARMA international, 2009: 6)

Les archivistes ne semblent pas être familiers avec l'évaluation des sites Web. L'absence d'une politique nationale ou organisationnelle en la matière n'aide pas à établir les critères de sélection des archives Web à des fins administratives ou encore patrimoniales. Des études menées auprès des usagers devraient aider les archivistes à appliquer les principes universaux de l'évaluation (voir revue de littérature, section 2.2.2.4, p. 40)²⁷. L'élaboration d'une politique nationale offrira un cadre de référence incitatif en vue de l'archivage des sites Web organisationnels.

En conclusion, cette multitude de façons de faire rend compte des défis qu'impose le Web tel que la redéfinition de la notion d'archives, la recherche de nouvelles méthodes et cadres de gestion, la précision des rôles et responsabilités. Nous avons noté beaucoup d'hésitations à considérer le site Web comme un objet archivistique. Pourtant, les politiques analysées tiennent les organisations responsables de la gestion de leurs archives et, par conséquent, de leur site Web puisqu'il forme une partie intégrante des archives d'une organisation. Souvent, on est trop centré sur l'objet plutôt que sur ses finalités. Le site Web ne ressemble pas aux documents papier, il faut repenser ces critères d'évaluation et d'archivage et revoir nos façons de faire pour les adapter aux caractéristiques des sites Web tout en tenant compte des exigences de l'archivistique en vue d'une saine gestion de l'information organisationnelle (Henry, 2000; Turnbaugh, 2002). Nous proposons dans la section 5.3 une démarche et des exemples d'outils qui peuvent aider les organisations à développer une stratégie d'archivage pour leur site Web public.

²⁷ Exemple : étude réalisée par la BnF : Chevalier, Philippe et Ilien Gildas (2011). *Les Archives de l'Internet : une étude prospective sur les représentations et les attentes des utilisateurs potentiels*. BnF, 30 p. Adresse URL : http://www.bnf.fr/documents/enquete_archives_web.pdf

5.3 Politique type pour l'archivage des sites Web organisationnels

Nous concluons de ce qui précède les principales caractéristiques d'une politique d'archivage. Nous rappelons dans les paragraphes suivants la définition et les composantes d'une politique d'archivage (ou d'une politique de gestion des archives). Nous présentons ensuite les composantes d'une politique d'archivage pour les sites Web organisationnels et les mesures de gestion documentaire à mettre en place tout au long du cycle de vie d'un site Web.

5.3.1 Caractéristiques d'une politique d'archivage

Selon la norme ISO 15489, les organismes doivent définir une politique en matière de *records management* afin de produire et d'organiser des documents authentiques, fiables, intègres et exploitables, qui documentent les activités de l'organisme aussi longtemps que nécessaire (ISO, 2001a). Selon Couture (1999: 10), une politique d'archivage a pour but « de planifier, de mettre en place, de tenir à jour et d'évaluer toute action, toute intervention archivistique permettant de coordonner de façon rationnelle, efficace et rentable la création, l'évaluation, l'acquisition, la conservation, le traitement, le repérage et l'utilisation des archives ». Tout organisme doit se doter d'un cadre pour assurer la bonne gouvernance de ses ressources informationnelles.

Certes, les organismes étudiés dans le cadre de cette recherche disposent des outils nécessaires à la gestion de leurs documents (ex. : plan de classification, calendrier de conservation). Les archives Web sont intégrées, dans certains cas, à ces outils (ex. : règle de conservation spécifique pour le site Web, règle spécifique pour les documents relatifs aux opérations de gestion du site Web). Cependant, une politique spécifique aux archives Web devrait être développée afin de servir d'outil de coordination de l'ensemble des actions à poser pour assurer une gestion uniforme du site Web et de ses composantes tout au long de leur cycle de vie. Cette politique

doit être cohérente avec la politique générale de gestion des archives (Gagnon, 1998: 107).

Une politique de nature générale visant la systématisation de la gestion des archives de l'organisme, tout support confondu, ne sera pas assez précise pour spécifier l'ensemble des activités nécessaires à la gestion complète, efficace et rentable du site Web public que produit l'institution dans son fonctionnement, depuis sa création jusqu'à sa conservation permanente ou son élimination (Couture, 1999: 13). Pour cette raison, les archives nationales, dont les politiques d'archivage du Web ont été analysées, élaborent une politique générale de gestion des documents qui vise à satisfaire les besoins d'une organisation en matière de preuve, de responsabilité et d'information dans la gestion de ses activités (FEPP, 2006: 8) et développent d'autres politiques spécifiques à des types de documents d'archives (ex. : courriers électroniques, documents sur support papier, documents numériques). Ces politiques comportent un renvoi vers les principes directeurs signalés dans la politique générale de gestion des archives et se consacrent aux actions à mettre en œuvre pour un type particulier d'archives. Ainsi, une politique sur l'archivage des systèmes Web de l'organisme précise les actions à mettre en œuvre et offre un cadre qui responsabilise les unités et les sensibilise à la nécessité d'archiver ce type de matériel :

« The aim of the policy and associated procedures are to help staff understand that websites create corporate University records and that they are therefore bound by a range of legal and recordkeeping requirements. [...] The material published to the web, is evidence of the activity taking place and therefore does become a record in its own right. » (The University of Melbourne, 2008f)

Nous avons noté, lors de l'analyse des politiques d'archivage du Web organisationnel, la dispersion des recommandations dans différents textes. Tout d'abord, l'approche générale de gestion d'un site Web est souvent publiée indépendamment des directives de sa mise en œuvre (ex. : Queensland State Archives, 2006a, 2006b). Ensuite, des directives spécifiques sont destinées à différents intervenants dans le processus d'archivage. Ainsi nous retrouvons des directives qui concernent les archivistes (ex. : TNA, 2006), d'autres visent les

webmestres (ex. : TNA, [s.d.]b). Enfin, les politiques nous renvoient vers des documents bien particuliers produits pour les besoins du site Web ou pour la gestion des archives numériques en général, et ce, afin d'avoir le détail relatif à une opération particulière (ex. : édition d'un site Web) ou qui nous renseignent sur les outils à utiliser (ex. : référentiel des métadonnées à intégrer au site Web). Cette dispersion nous a amenée à ajouter une vingtaine de documents supplémentaires au corpus initial des 33 politiques analysées sans quoi notre analyse restait incomplète. Cette dispersion de l'information n'offre pas une vision globale du processus complet de gestion et d'archivage des sites Web. Dans ces cas, en particulier, il serait important de pouvoir contrôler le processus continu de sa gestion ainsi que les différents états dans lesquels il se présente (site actif, site inactif, site archivé). Une politique doit non seulement mentionner la liste des outils à utiliser ou à développer, mais aussi en présenter un modèle. Ces caractéristiques n'étaient pas toujours prises en considération dans le corpus de politiques d'archivage que nous avons analysé.

Dans la littérature archivistique, il n'existe pas de modèle normalisé de politique pour la gestion des archives. Nous avons examiné, dans le cadre d'un projet de recherche, de nombreuses politiques d'archivage (Chebbi et Cherif, 2010). Nous constatons une grande diversité des éléments qui composent ces politiques. Toutefois, nous avons pu repérer quelques références qui indiquent la liste des composantes qu'une politique d'archivage devrait inclure (BARBICA, [s.d.]: 5-11; Conseil canadien des archives, 1995: 2-20; Couture, 1999: 10-13; FEED, 2006: 9-11; Gagnon, 1998: 106; ISO, 2001a: 5, 2001b: 1; Lemay, 2012a: Module 07; New Zealand, 2006b: 3-5). L'on y suggère les catégories de composantes suivantes : l'objectif qu'elle vise, les principes généraux qui sont à son origine, un plan d'action où l'on précise l'ensemble des activités nécessaires à la gestion complète, efficace et rentable des archives tout au long de leur cycle de vie dans le respect des lois et règlements ainsi que des politiques internes et externes, les rôles et responsabilités des personnes visées par les actions à entreprendre pour appliquer la politique et les ressources nécessaires pour la mettre en œuvre (ex. : budget, équipement). Ce dernier élément,

indiqué dans les manuels archivistiques, ne semble être détaillé dans aucune des politiques analysées dans le cadre de cette recherche.

Selon les références consultées, une politique doit être datée, approuvée par les hauts responsables de l'organisme et régulièrement mise à jour. Elle doit être communiquée et implantée à tous les niveaux de l'organisation. La politique doit être supportée par des procédures qui expliquent de manière détaillée comment elle doit être appliquée. Un mécanisme de contrôle devrait être mis en place afin de vérifier la cohérence entre les objectifs visés par la politique et les résultats de son application.

Avant de structurer la politique d'archivage et de la diffuser au niveau de l'organisation, l'archiviste doit bien planifier sa conception. Pour ce faire, les archivistes doivent établir les exigences de gestion documentaire, analyser les pratiques actuelles de gestion documentaire et consulter les décideurs et tout intervenant impliqué dans la création et la gestion des documents (Archives New Zealand, 2006b: 2-3).

Les Archives New Zealand proposent une liste de contrôle « *Checklist for developing a recordkeeping policy* » qui guide l'archiviste dans le processus d'élaboration d'une politique d'archivage en vérifiant l'accomplissement de toutes les étapes relatives à sa planification, sa structuration et sa communication (Archives New Zealand, 2006b: 7).

5.3.2 Composantes d'une politique type pour l'archivage du Web organisationnel

Afin d'identifier les composantes propres à une politique d'archivage du Web, nous avons analysé la structure des politiques étudiées dans le cadre de cette recherche (voir Annexe 3, p. xxii), en plus de quelques ouvrages sur les politiques documentaires à l'ère de l'internet (Bonnenfant, 2008; Calenge, 2008; Carbone et Cavalier, 2009).

Nous suggérons, ci-après, les composantes d'une politique pour l'archivage du Web organisationnel. Les composantes que nous proposons décrivent l'énoncé

général de la politique, ses objectifs, son champ d'application, le cadre juridique dans lequel elle se situe, une définition des principaux termes qui s'y rapportent, les principes directeurs sur lesquels elle se fonde, les rôles et responsabilités des intervenants devant l'appliquer, les mesures qui doivent être appliquées afin d'atteindre les objectifs qu'elle vise ainsi qu'un ensemble de champs descriptifs de la politique (approbation, date d'entrée en vigueur, révision, contacts et annexes). L'ordre de présentation de ces éléments est à titre suggestif. Les intitulés des champs peuvent être différents puisque les termes utilisés pour les décrire ne sont pas normalisés. Nous décrivons le contenu de chaque champ composant la politique en l'illustrant avec des exemples tirés des politiques analysées.

- **Énoncé général/objet de la politique** : ce champ justifie la raison d'être de la politique. Il sera pertinent de rappeler, à ce niveau, que le site Web est un moyen et un produit qui découle de la conduite des affaires de l'organisme et que, à ce titre, il importe de s'assurer, par une gestion adéquate, de son archivage à des fins de justification de droits, pour la continuité des affaires de l'organisation ou les besoins de témoignage et d'information. Une politique pour les archives Web offre des balises pour encadrer le processus de gestion de ces documents.

Exemple :

« Many organisations are not aware that it can be important to create and maintain records of their website. Key information is made available via websites and significant business transacted there but often no recordkeeping is performed to support and document these activities. This section provides some guidance on managing web records. » (NSW, [s.d.])

- **But et objectifs** : ce champ décrit les finalités de la politique poursuivies par sa mise en œuvre. Ces finalités doivent être cohérentes avec la politique générale de gestion des archives et les cadres de gestion documentaire au niveau institutionnel, national et international.

Exemple :

« This guideline aims to:

- assist public offices in determining what constitutes a web record,
- suggest how to create and capture web records, and
- discuss how to maintain web records. » (NSW, 2009a)

- **Champ d'application** : ce champ sert à spécifier les organismes ou les entités organisationnelles qui sont soumis à la politique. Ce champ doit également identifier les systèmes Web visés par la politique (site Web public, intranet, extranet, etc.). Les sites éphémères, les sites inactifs, les sites dont le contenu est partagé avec une autre institution et les sites dont le contenu appartenait à un organisme qui n'existe plus ou qui a été fusionné avec l'organisme responsable du site Web actuel doivent être inclus dans le champ d'application.

Exemple :

« All public offices including those that outsource their website management are obliged to make and keep adequate web records. This guideline is relevant to web records produced for both public office internet or intranet sites and it is important to remember that both types of sites create State records. » (NSW, 2009a)

- Ce champ indique aussi les documents visés par la politique. Une liste détaillée des documents qui forment les archives Web peut être fournie (voir l'exemple dans NARA, 2005: 5-7; voir aussi la liste des archives Web établie dans la section 5.1.2, p. 185).

Exemple :

« Web site-related records include (1) web content records, which represent information presented on a web site, and (2) web site administrative records, which provide evidence of the management and operations of the web site. » (NARA, 2005: 5)

- **Définition des termes** : ce champ donne la définition des principaux termes archivistiques ou technologiques employés dans la politique, garantissant ainsi que tous les intervenants visés par la politique auront la même compréhension et interprétation des termes clés liés à l'archivage du Web organisationnel. Le contenu de ce champ pourrait figurer en annexe de la politique.

Les termes suivants doivent être impérativement définis : archivage, archives Web, les différents types de systèmes Web (intranet, extranet, site Web public, système de gestion de contenus, site Web dynamique, site Web statique), portrait du site Web (*snapshot*) et services en ligne (pour un exemple, voir NAA, 2001b, p. 9-10). Nous avons également fourni dans le chapitre Revue de la littérature la définition de quelques-uns de ces termes.

- **Contexte juridique et réglementaire** : ce champ rappelle, voire énumère les lois et règlements qui sont à l'origine de la politique ainsi que les politiques (internes ou externes) qui ont servi à son établissement. Ce champ permet aussi de situer la politique d'archivage du Web organisationnel par rapport à la politique générale des archives de l'organisme. Chaque organisme évolue dans un contexte juridique et réglementaire qui lui est propre.

Exemple :

« Web records are covered by the same legal disposal requirements that apply to all other government records. » (NSW, [s.d.]

« This policy forms part of a wider framework that aims to ensure best practice recordkeeping and information management occurs in Queensland public authorities. » (Queensland State Archives, 2006b: 3)

Il existe cinq niveaux d'application du cadre juridique et réglementaire qu'il faudrait prendre en compte, à savoir les lois relatives aux archives (ex : *Loi sur les archives*), les autres textes juridiques et réglementaires complémentaires aux lois sur les archives (ex. : *Loi concernant le cadre juridique des technologies de l'information*, *Règlement sur le calendrier de conservation, le versement, le dépôt et l'élimination des archives publiques* de BAnQ), les lois et règlements qui concernent le domaine d'activité de l'organisme (ex. : *Loi sur l'enseignement privé*), les lois, normes et politiques pouvant contenir des obligations relatives aux documents (ex. : Code civil du Québec, Standard sur le format des documents diffusés dans le Web) ainsi que les normes et conventions internationales (ex. : ISO 15489:2001, ISO 16175:2010) (Lemay, 2012b: Module 03; Leroux et Lemay, 2009: 37-39). Ces niveaux peuvent servir à structurer ce champ.

L'archiviste doit mettre en place un processus de veille juridique qui lui permet de signaler la parution de nouveaux textes juridiques et d'évaluer leur impact sur la politique d'archivage du Web existante (McClure, 2000).

- **Principes directeurs** : ce champ présente un ensemble de règles générales de gestion documentaire inspirées des principes fondamentaux de l'archivistique. Ces principes servent de cadre de référence dans la définition des mesures de gestion afin qu'elles soient en conformité avec le contexte juridique et réglementaire et en cohérence avec les principes archivistiques.

Les principes directeurs peuvent se présenter sous la forme d'une liste de contrôle (*checklist*)²⁸ (pour un exemple, voir Australian Government, 2004). Les politiques analysées (Archives New Zealand, 2006; NAA, 2001a: 15, 19; Public Record Office Victoria, 2008; Queensland State Archives, 2006b: 6, 10, 12) recommandent de rappeler les principes suivants :

²⁸ Pour un modèle de liste de contrôle, consulter : Australian Government. Information Management Office. (2004). *Archiving web resources. Better practice checklist - 7: practical guide for effective use of new technologies in government. Archiving web resources*. 12 p. Adresse Url: <http://web.archive.org/web/20090614181325/http://www.finance.gov.au/e-government/better-practice-and-collaboration/better-practice-checklists/docs/BPC7.pdf> (Page consultée le: 23 avril 2007).

- o Un organisme produit des archives lorsqu'il utilise des systèmes Web afin de diffuser des informations ou conduire des activités et transactions en ligne;
- o Les archives Web doivent être capturées dans un système d'archivage afin de préserver leurs authenticité, fiabilité, intégrité et exploitabilité;
- o Les choix de gestion documentaire doivent correspondre aux besoins de l'organisme, considérer la nature technologique des systèmes Web implantés, respecter les caractéristiques des archives produites et être conformes au contexte réglementaire dans lequel évolue l'organisme. Une analyse des risques doit être établie à cette fin;
- o L'obsolescence et la dépendance technologiques doivent être contrôlées. Des politiques, des procédures et des outils doivent être définis pour minimiser les risques qu'elles présentent sur la pérennisation de la mémoire organisationnelle;
- o Les conditions de transfert des archives Web aux archives nationales doivent être vérifiées et respectées.

Cette liste n'est pas exhaustive. Elle présente une synthèse des principes sur lesquels se fondent les politiques d'archivage analysées.

- **Rôles et responsabilités** : ce champ détermine la responsabilité de chacun des intervenants impliqués dans la conception et la maintenance du site Web, la création et la diffusion des contenus sur le site Web et l'archivage du site Web. Trois principaux intervenants sont impliqués dans le processus d'archivage des sites Web : les créateurs de contenus, les webmestres et les archivistes (McClure et Sprehe, 1998). Cette catégorisation suffit lorsqu'il s'agit d'une politique nationale. Nous recommandons, lorsqu'il s'agit d'une politique organisationnelle, de présenter une catégorisation plus détaillée. Les groupes suivants ont été identifiés suite à l'analyse des politiques sur l'archivage du Web organisationnel :
 - Les hauts responsables de l'organisation;
 - Les auteurs de contenus Web;
 - Les webmestres;
 - Le personnel responsable de la gestion de l'infrastructure technologique et communicationnelle de l'organisation;
 - Les archivistes;
 - Les contractants, le cas échéant;

Les rôles et responsabilités ne sont pas mutuellement exclusifs puisque, dans certains cas, le même employé pourrait avoir la charge de différentes tâches liées au site Web. La description des postes et l'analyse des activités réalisées par chaque intervenant permettraient de préciser celles qu'il doit assumer par rapport au site Web. Nous avons exposé, dans le chapitre Résultats, les rôles

et responsabilités des principes intervenants dans le processus de gestion et d'archivage d'un site Web public d'une organisation de même que les rôles et responsabilités des archives nationales dans ce processus. Un partage collaboratif des responsabilités est préconisé dans les politiques analysées.

Il faudrait aussi tenir compte, lors de la définition de ce champ, de l'approche générale de gestion et d'archivage des systèmes Web de l'organisation. Une approche de gestion centralisée n'implique pas les mêmes rôles et responsabilités qu'une approche de gestion décentralisée.

Exemple :

« Web Administrator

- Manages information technology aspects of web sites
- For static materials, applies departmental record-keeping
- For dynamic materials, consults with content manager and records officer to determine whether materials qualify as records and what area is responsible for the content of the materials
- Maintains historical log and local site repository.

Content Manager

- Decides whether posting is a 'record' or a 'publication'
- For static material, ensures that copy of the posted 'record' and its metadata, as well as any administrative records associated with the posting process are transferred to a record keeping system (e.g. University of Toronto File plan) either before or at the time of posting
- For dynamic material, determines whether the materials qualify as records and how the departmental record keeping system will capture the records.

Records Officer

- Consults University of Toronto File Plan for approved retention periods, or other approved retention schedules, or contact UTARMS for advice on final disposition of unscheduled records
- Contacts University Archivist for transfer of archival web site records and publications and associated metadata. » (University of Toronto, 2006)

- **Mesures à mettre en place:** ce champ sera détaillé dans le point 5.3.3 (p. 217) puisqu'il constitue le noyau central de la politique. Il indique les actions et les mécanismes découlant des principes directeurs ainsi que des objectifs de la politique et qui doivent être mises en œuvre pour gérer le processus de création, de capture, de conservation et de préservation des archives Web (Couture, 1999: 10; FEPP, 2006: 7-11).

La structure que nous proposons s'avère appropriée à la description de l'ensemble des actions à mettre en œuvre pour la gestion des archives Web, et ce, tout au long du cycle de vie du site Web. Des mesures particulières visent

la normalisation de la technologie utilisée et des documents diffusés sur le site Web. D'autres ciblent plutôt le contrôle du processus de publication de l'information (mise en ligne, mise à jour et retrait des documents). Ces deux catégories de mesures faciliteront l'archivage du site Web et garantiront son exploitabilité. Enfin, certaines mesures encadrent le processus d'archivage du site Web lorsqu'il passe du statut actif à un statut inactif.

Les mesures que nous suggérons ne sont pas exhaustives. Le choix de l'une ou de l'autre dépend des résultats de l'analyse des risques associés à l'usage du site Web. Elles doivent, par contre, être définies depuis la création du site Web, voire même sa conception.

- **Approbation** : La politique d'archivage doit être approuvée par les hautes instances de l'organisation afin qu'elle soit reconnue et mise en œuvre (ISO, 2001a: 5). Cette approbation officialise le document et fait de lui une preuve attestant de la conduite des actions relatives à la gestion et à l'archivage des systèmes Web d'une organisation (Gagnon, 1998: 107). Cette information pourrait figurer sur la page de couverture de la politique.
- **Date d'entrée en vigueur** : ce champ indique la date d'entrée en vigueur de la politique et renseigne aussi sur l'actualité des mesures préconisées. Cette information pourrait figurer sur la page de couverture de la politique.
- **Révision de la politique** : ce champ vise à garder la traçabilité des différentes révisions apportées à la politique. Cette dernière doit être révisée régulièrement en fonction de l'évolution du cadre juridique dans lequel elle se situe, de l'évolution des technologies d'information qui affectent à la fois les systèmes Web et la nature des documents qu'ils véhiculent et, enfin, à la lumière des retours d'expériences.

La révision de la politique doit également s'établir dans l'objectif d'évaluer l'efficacité des mesures qu'elle propose par rapport aux objectifs qu'elle se fixe. Ces mesures doivent aussi refléter les besoins courants de l'organisme (Couture, 1999; ISO, 2001a). Les politiques analysées recommandent de procéder à la mise à jour d'une politique sur l'archivage du Web tous les six mois (Australian Government, 2004; TNA, 2009b). Lorsqu'une mise à jour est effectuée, il importe d'aviser les différents intervenants visés par la politique. Les révisions qui ont été appliquées à la politique peuvent être indiquées sur la page de couverture.

Exemple :

« This policy may therefore be reviewed and revised in the light of comments from users of the records or from archive professionals, the experience of departments in using the policy, or as a result of newly discovered information. » (TNA, 2006)

« A review should be conducted within six months of implementing a strategy for webbased recordkeeping. The review should identify the

effectiveness of the approach, measure performance, identify any corrective action or enhancements required, minimise the agency's exposure to risk through procedural failure and develop a program for conducting regular reviews. » (Australian Government, 2004)

- **Contact et retours d'expériences** : ce champ indique les coordonnées d'un service ou d'une personne qui pourrait être contactée au sujet de la politique ou afin de lui rendre compte des difficultés rencontrées lors de la mise en application de la politique. Les données recueillies permettront de faire évoluer la politique et la mettre à jour.
- **Annexes** : il est déconseillé d'annexer des documents à une politique, car cela alourdit le texte de la politique et rend difficile sa mise à jour (FEED, 2006: 11). Toutefois, nous jugeons important qu'elle soit accompagnée, sans qu'ils soient annexés, de certains documents nécessaires à la bonne interprétation et à la mise en œuvre de la politique (ex. : modèle de gestion des risques, modèle de règle de conservation, modèle vierge de formulaire pour l'inventaire des documents à archiver, etc.).

La plupart des politiques analysées sont disponibles en format Web ainsi qu'en format PDF. La version PDF peut présenter la forme allégée de la politique. La version en format Web est souvent localisée dans une section du site Web qui décrit le projet et fait un lien vers l'ensemble des documents et outils qui gravitent au tour de la politique d'archivage (voir par exemple la section sur l'archivage du Web gouvernemental sur le site The National Archives (Royaume-Uni)) (l'exemple est disponible sur TNA, ([s.d.]a)).

Notes :

Nous avons analysé de nombreuses politiques d'archivage du Web développées par des archives nationales. Des éléments supplémentaires doivent être précisés lorsqu'il s'agit de développer une politique nationale pour l'archivage du Web organisationnel. À titre d'exemple, il faut bien établir les frontières entre une publication et un document d'archives dans un environnement Web, et ce, dans le but de délimiter les territoires de collecte entre les archives nationales et les bibliothèques nationales (pour un exemple, voir Archives New Zealand, 2009: 8).

Exemple : définition des archives Web

« Web records can be defined as records generated, captured or received using web technologies. Web records on an organisation's website can generally fall into three categories 1) web content records, including introductory or descriptive information about the organisation, its activities or particular areas it manages, but also publications that it makes available on the web, 2) inputs/outputs of transactions performed via the web e.g. data submitted via online forms and 3) website administration records, such as statistics,

diagnostics etc, which provide evidence of website operations and management. » (NSW, 2009a)

Si les archives nationales entreprennent l'archivage du Web organisationnel, elles doivent également articuler leurs actions avec celles entreprises dans les organisations. Souvent, les archives nationales procèdent, de manière volontaire, à la capture périodique de portraits des sites Web gouvernementaux à des fins historiques et les mettent en ligne à la disposition du public. Il importe, dans ce cas, d'indiquer comment l'archivage des sites Web organisationnels et l'accessibilité en ligne des documents archivés sont soutenus par la loi.

Exemple :

« La Section 8(2) de la Loi sur la Bibliothèque et les Archives du Canada donne à BAC le mandat de constituer des échantillons de sites Web accessibles pour le public qui sont d'un intérêt pour les Canadiens et les Canadiennes. » (BAC, 2007-2008)

Dans d'autres cas, les archives nationales exigent le versement de ces portraits par l'organisme responsable du site Web. Dans cette situation, les archives nationales doivent développer des règles pour régir le transfert de ces portraits, préciser leur périodicité et les documents qui doivent les accompagner.

Les différentes composantes d'une politique sur l'archivage du Web organisationnel ayant été présentées, nous décrivons maintenant de manière détaillée les mesures à mettre en place.

5.3.3 Mesures à mettre en place

Nous présentons dans cette section la composante « Mesures à mettre en place » de la politique pour assumer la gestion d'archives Web fiables, intègres, authentiques et exploitables. Nous rappelons que cette politique type s'applique principalement au site Web public d'une organisation. D'autres recherches sont nécessaires pour vérifier si ces mesures s'appliquent de la même manière à d'autres plateformes Web (ex. : intranet, extranet).

Les mesures indiquées sont accompagnées d'exemples d'outils à utiliser ou à développer pour la mettre en œuvre. Elles ne sont pas exhaustives, elles découlent des travaux menés depuis quelques années par des organismes avec des référentiels archivistiques différents de ceux du Québec (ex. : États-Unis, Australie, Nouvelle-

Zélande). Ceci étant dit, ces mesures restent toutefois pertinentes puisque ces politiques font référence à des normes internationales de gestion documentaire (ex. : norme ISO 15489). En outre, le Web, de par sa nature, impose les mêmes défis de gestion à tous. Les politiques analysées se rejoignent au niveau des principes de gestion et des stratégies d'archivage. Les outils diffèrent selon l'expérience acquise par l'organisme dans ce domaine.

Avant d'établir des choix techniques, des principes directeurs pour la gestion des archives Web doivent être définis et une analyse des risques associés à la diffusion en ligne des informations doit être menée. Cette démarche permet de vérifier si les mesures mises en œuvre suite à l'analyse des risques associés au site Web public de l'organisme sont conformes aux principes archivistiques, aux lois ainsi qu'aux normes de gestion documentaire. Aussi, elle permet d'évaluer, une fois que la politique est appliquée, si les mesures proposées sont efficaces (ex. : mesures qui diminuent ou éliminent les risques).

5.3.3.1 Analyse des risques

Les mesures à mettre en place pour gérer le processus d'archivage du Web découlent des principes directeurs, mais dépendent avant tout des niveaux de risques auxquels l'organisme accepte de s'exposer en toute connaissance de cause lorsqu'il publie des informations sur son site Web public (BAC, 1999; McClure et Sprehe, 1998; NARA, 2005). La conduite d'une analyse des risques est recommandée dans l'ensemble des politiques d'archivage analysées. Elle consiste à étudier : « les niveaux de risques associés à la gestion des renseignements contenus dans chaque site Web et recommande des mesures appropriées de gestion documentaire. » (BAC, 1999)

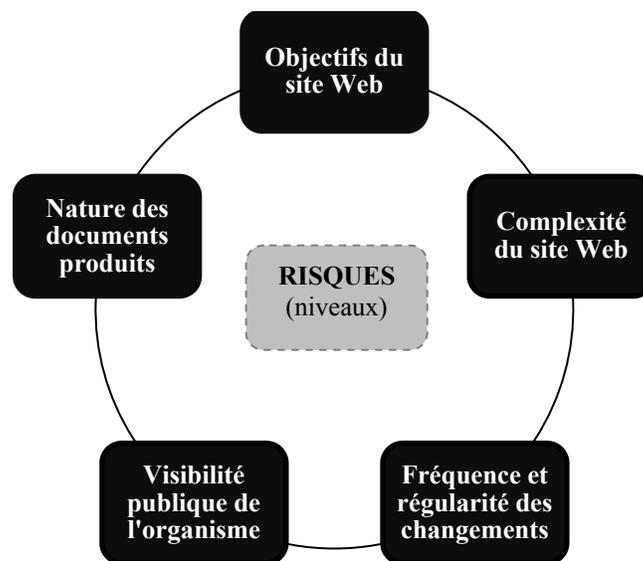
Lorsqu'un organisme ne s'assure pas d'une saine gestion de son actif documentaire, il risque l'inaccessibilité à ses documents, le manque de crédibilité et de confiance dans l'information nécessaire à la prise de décision et son incomplétude, un manque de performance, une perte financière et le non-respect des lois et règlements en conséquence (Lemieux, 2004: 2). Lorsqu'il s'agit du site Web

organisationnel, l'organisme risque, de plus, une mauvaise réputation, un manque de transparence, le manque de confiance dans les informations, services et produits qu'il offre sur internet et la perte de documents à valeur patrimoniale documentant un phénomène unique (Archives New Zealand, 2009: 10).

Empruntée du monde de la gestion, la méthodologie de l'analyse des risques consiste à identifier, décrire, évaluer, traiter, contrôler les risques réels et potentiels. Ces risques doivent être ensuite catégorisés (risque faible, risque moyen, risque élevé). L'organisme peut choisir d'éviter ces risques, réduire leur probabilité ou impacts ou tout simplement les accepter. Cette démarche permet donc de vérifier si les mesures actuelles doivent être améliorées ou si elles sont tout simplement absentes. L'organisme pourrait ainsi établir des priorités au niveau des mesures à prendre et décider de la meilleure stratégie pour traiter ces risques (Lemieux, 2004: 10-24).

La Figure 6 (p. 219) présente les facteurs à prendre en considération pour asseoir une analyse des risques d'un site Web public.

Figure 6 : Modèle enrichi de McClure et Sprehe (1998) sur l'analyse des risques – facteurs à considérer.



L'analyse des risques doit être réalisée de manière collaborative et impliquer la participation des producteurs de contenus, des archivistes, des webmestres et du personnel responsable de la gestion de l'infrastructure technologique du site Web (Queensland State Archives, 2006a: 34). Elle s'appuie sur l'évaluation des facteurs suivants :

1. **Les objectifs du site Web** : indiquer les fins pour lesquelles le site Web est utilisé afin de pouvoir spécifier les documents qui en découlent : « Websites serve a variety of purposes [...]. For every purpose the web is used for, a different record may be generated » (NSW, 2009b). Les sites peuvent servir aux fins suivantes :
 - a. Le site Web est un entrepôt en ligne de documents relativement stables,
 - b. Le site Web est un entrepôt fluide d'informations et de documents produits par l'organisation,
 - c. Le site Web est un moyen de communication qui offre des options minimales d'interaction,
 - d. Le site Web sert d'interface pour accéder aux différents services et informations offerts par l'organisation;

2. **La complexité du site Web** : identifier la nature du site Web. Les sites Web dynamiques sont plus difficiles à archiver : « dynamic data access websites raise some additional issues for agency recordkeeper » (NAA, 2001b: 12). Les organismes implantent des sites Web présentant des niveaux de complexité différents (NAA, 2001b: 11-12; NARA, 2005: 3; NSW, [s.d.]) :
 - a. Site Web de nature statique (entrepôt relativement statique ou fluide d'informations),
 - b. Site Web de nature dynamique,
 - c. Site Web géré par un système de gestion de contenus,
 - d. Site avec des microsites (thématiques ou événementiels),
 - e. Site Web hébergé à l'interne ou à l'externe,
 - f. Un mélange de toutes ces options;

3. **La fréquence et la régularité des changements** : préciser la nature des changements qui affectent le contenu diffusé sur le site Web. Lorsque les changements sont irréguliers, mais fréquents, ils présentent un niveau de risque plus important : « Websites that change frequently but irregularly are likely to pose the most risk to an agency, because such changes are the most

difficulty to track.» (Australian Government, 2004: 3). Les changements peuvent être :

- a. fréquents et réguliers,
 - b. irréguliers, mais fréquents,
 - c. irréguliers et non fréquents,
 - d. réguliers, mais non fréquents;
4. **La visibilité publique du site Web** : indiquer si le site Web constitue une source principale ou alternative pour diffuser l'information auprès des clients (ou citoyens) et partenaires. Quand l'organisme publie des informations sur des affaires publiques ou lorsqu'il offre une variété de services en ligne, il s'expose à des risques plus importants : « When a public authority's business has a high profile or level of public scrutiny, or they provide a large range of services to the public, they normally have a higher level of visibility and associated risks. » (Queensland State Archives, 2006a: 31);
5. **La nature des documents qui découlent de l'usage du site Web** : analyser le mode d'alimentation du site Web (modes de création), dans quels formats le contenu est diffusé et la nature des liens qui existent entre les documents produits et les fonctions de l'organisme.

Suite à cette analyse, l'organisme détermine si :

- Le contenu de son site Web est publié ou non dans d'autres sources d'information;
- Certains contenus ou l'ensemble de son site Web sont gérés par une tierce personne ou dans des applications externes;
- Les documents de gestion de son site Web sont intégrés ou non dans un système d'archivage (Archives New Zealand, 2009: 7-8).

Cette analyse des risques permettrait d'établir ensuite :

- Le volume et la complexité des archives Web à gérer;
- La capacité des systèmes existants à gérer la complexité des archives Web;
- Les niveaux de risques (faible, moyen, élevé) encourus par l'organisation (ex. : perte de l'information);

- Les exigences de gestion et de conservation des archives Web;
- Le(s) niveau(x) de granularité à considérer lors de l'archivage du site Web (NAA, 2001b: 15, 19).

Les informations recueillies serviront, enfin, à définir des mesures appropriées de gestion documentaire, autrement dit, proposer une stratégie d'archivage des archives Web ainsi que les outils archivistiques qui concourent à sa mise en œuvre (Archives New Zealand, 2006: 9; BAC, 1999).

Ci-après est présentée une compilation des mesures à mettre en place, tout au long du cycle de vie d'un site Web (création, gestion, retrait et conservation à des fins administrative ou historiques). Nous avons préférée cette présentation à celles qui représente le cycle de vie des archives (archives actives, archives intermédiaires et archives historiques) dont les frontières sont rigides, difficiles à établir pour un document aussi évolutif que le site Web qui peut être à la fois en phase active alors que quelques-uns des documents qui y sont encore diffusés sont à leur phase de semi-activité. Ils sont archivés sur le site Web à des fins de référence et d'information. Cette représentation rejoint plus le modèle du Records continuum qui nous offre une vue sur les différents états d'un document tout au long de son cycle de vie.

5.3.3.2 Autres mesures à mettre en place

La liste des mesures que nous suggérons n'est pas exhaustive. Le choix de l'une ou l'autre dépend, en effet, des résultats de l'analyse des risques et du contexte dans lequel évolue l'organisme. La politique développée par le State Records Authority of New South Wales (NSW, 2009a) liste tous les scénarios et mesures possibles pour l'archivage des sites Web organisationnels. Les guides de mise en œuvre développés par les National Archives of Australia (2001b), le Public Record Office Victoria (2008) et les Queensland State Archives (2006a) discutent les avantages et les inconvénients pour chacune d'entre elles tout en indiquant dans quel contexte le choix de l'une ou de l'autre est pertinent. Rappelons que les mesures à mettre en œuvre suite à l'analyse des risques doivent être conformes au contexte réglementaire externe et interne d'une organisation de même qu'aux normes de gestion documentaire.

Nous avons jugé important d'indiquer les outils qui peuvent découler de la mise en place d'une mesure particulière de gestion. Nous avons signalé les outils indiqués dans les politiques d'archivage et quelques autres identifiés suite à une recherche d'information sur les sites Web de projets d'archivage du Web organisationnel.

Les mesures à mettre en œuvre sont spécifiées selon le cycle de vie d'un site Web. Un site est conçu et ensuite il est mis en ligne. Il reste actif selon les besoins de l'organisation. Au cours de cette phase, il est continuellement mis à jour (ajout, modification, mise à jour ou retrait des informations diffusées, modification de l'aspect visuel du site, etc.). Certains documents sont retirés dans un autre espace ou archivés en ligne dans une section particulière du site Web, ces documents passent à un stade de semi-activité. Enfin, le site Web n'est plus actif. Il est remplacé par une nouvelle version. Le site Web et les documents qui le composent ne sont plus actifs. Ils sont archivés à des fins administratives ou historiques, voire éliminés.

Nous présentons, pour chacune de ces phases, les mesures de gestion appropriées.

5.3.3.2.1 Conception et mise en ligne du site Web

Les mesures énoncées ci-après visent à permettre la création des sites Web selon une technologie et dans un format qui n'entravent pas ou ne compliquent pas son processus d'archivage. Une fois le site actif, des mesures particulières doivent être appliquées afin de déterminer, au tout début de son cycle, les documents qu'ils génèrent, d'identifier ceux qui doivent être capturés dans un système d'archivage et de spécifier selon quelles stratégies ils doivent être capturés. L'organisme doit :

- **Concevoir des sites Web simples et développer des politiques ainsi que des procédures qui encadrent le processus de développement des sites Web et de mise en ligne des documents** : « keep it simple » (TNA, 2009a), afin d'éviter des problèmes lors de leur capture ou préservation. La technologie AJAX est non recommandée, à titre d'exemple puisque les robots de collecte ne sont pas capables actuellement d'identifier et de moissonner les pages générées par cette technologie (Public Record Office Victoria, 2008).

Exemples d'outils à développer :

- o Lignes directrices pour la gestion des URL et l'usage des technologies de redirection (TNA, 2009b).
 - o Lignes directrices pour la conception de la carte du site Web (TNA, [s.d.]a).
 - o Pour avoir une idée sur les règles à respecter lors de l'écriture des URL qui permettrait d'éviter des problèmes lors du moissonnage du site Web, voir Public Record Office Victoria, 2008: 20-24).
- **Adopter des solutions technologiques ouvertes et bien documentées.**

« When designing and building websites, agencies should plan to use software tools and applications that meet accepted (or de facto) standards and that are readily available and fully supported. »
(Australian Government, 2004: 8)

Exemple d'outils à développer :

- o Liste des formats documentaires à utiliser sur le site Web
- **Documenter les technologies de développement** : décrire l'architecture des systèmes et les versions des logiciels utilisés (instructions, mode d'emploi), évaluer le volume de l'information qui est y stockée, indiquer les systèmes de gestion des documents utilisés (NARA, 2005).

Exemples d'outils à développer :

- o Liste des systèmes Web dans l'organisation. Cette liste, qui pourrait prendre la forme d'une feuille de calcul, devrait inclure les éléments suivants :
 - Titre du site et son URL (avec identification des sous-sites)
 - Type du site Web et noms de domaines de l'organisation
 - Cycle de révision du site
 - Nom de la personne responsable qui figure en bas de page de chaque page Web
 - Noms des personnes qui mettent à jour le contenu du site Web
 - Noms des personnes qui approuvent le contenu et ses mises à jour
 - Noms des personnes qui alimentent le contenu du site Web (ceux qui fournissent un contenu dans un format différent du format HTML) (University of Tasmania; 2010b)
- **S'assurer que des copies de sauvegarde sont prises quotidiennement.**
Cette technique garantit la conservation des documents éphémères et

représente une stratégie de récupération des données en cas de problèmes techniques (Archives Office of Tasmania, 2009).

« As with other electronic records, procedures for regular backups of all web sites and associated repositories such as the local site repository and the e-library should be undertaken. » (University of Toronto, 2006)

La règle des trois générations de sauvegardes est conseillée : les trois dernières copies de sauvegarde devraient être préservées tout le temps. Les supports de conservation de ces copies doivent être préservés à un endroit sécuritaire.

- **Assigner des métadonnées pertinentes et complètes aux documents créés ou diffusés en ligne, notamment ceux qui existent uniquement sur le Web** (ex. : page d'accueil, carte du site). Deux groupes de métadonnées doivent être assignés : des métadonnées descriptives des ressources mises en ligne et des métadonnées qui gardent la traçabilité des opérations d'archivage du document (ex. : nom du robot de collecte, date et heure de la collecte). Il faut également décider si un outil particulier doit être développé, choisir un mode centralisé ou décentralisé pour la gestion des métadonnées et établir le mode de gestion de ces métadonnées et leur accessibilité aux moteurs de recherche internes et externes :

« Records of websites and webpages should be managed over time: The issues of system changes and media and file format obsolescence, and the continued linkage of recordkeeping metadata with individual records need to be managed to ensure appropriate records are maintained for as long as necessary. » (Archives Office of Tasmania, 2009)

Pour un exemple des métadonnées à appliquer aux documents Web, voir NARA (2005) et BAC (1999).

- **Identifier et inventorier les documents produits ou issus des technologies Web** : « Public authorities should also consider conducting a content audit of their online resources to determine exactly what information is available online. » (Queensland, State Archives, 2006: 6)

« A register of the resources displayed online, such as documents, graphics and forms, along with associated metadata is maintained and linked to the URL where they have been available. » (Australian Government, 2004: 23)

Pour un exemple d'outil, consulter The University of Melbourne, 2008f :

« Records Services has created a spreadsheet in Excel that contains the UOM's Enterprise Classification Scheme (ECS). A number of 'seed' urls have been listed (a 'seed' means the initial starting point of a

Crawl in the process of creating a Collection of websites) against each function. This means that in theory, any url that is attached to a specific function/activity combination will also have a disposal option attached to it as well.

The spreadsheet however, is not comprehensive. It is not comprehensive because it does not attempt to list and appraise every single website.

The goal of the spreadsheet is to contain a sample listing of urls that can be checked against quarterly web crawls to ensure that information (and web records) that are related to a core University function, are being captured. It does provide evidence of Records Services ability to assess the functions delivered or documented on the website. » (The University of Melbourne, 2008f)

- **Déterminer les documents et les composantes du site Web devant être archivés** : « Having determined what parts of a website need to be captured, it is then possible to make technical decisions about how that part is to be captured. » (Public Record Office Victoria, 2008). Ces documents doivent être intégrés au calendrier de conservation

« The appraisal of web records requires all elements of a website be considered, including pages, content items, files, background processes, look and feel, automated or manual transactions, and evidence of business activity in creating and maintaining websites. The appraisal process generally does not consider the format of the record as this does not usually have a bearing on its value. However some web records may be defined by format. In this case, the appraisal process should identify where the format of the record provides specific recordkeeping requirements. For example, information that is part of the organisational record as a static document, but which is also offered on the website with interactive functions, such as user comments acting as consultation feedback, may need to be kept as a record because of the additional functionality provided by the website. » (Archives New Zealand, 2006b: 11)

Exemple d'outils à développer :

- o Liste des pages Web soumises à des mises à jour fréquentes et irrégulières
- o Liste des documents à exclure de la collecte (ex. : les agendas en ligne doivent être exclus de l'archivage afin d'éviter les *crawler trap* lors de la collecte)
- **Choisir une stratégie d'archivage appropriée aux caractéristiques des documents à archiver.** Il existe différentes approches et techniques

d'archivage. Les politiques d'archivage du Web organisationnel recommandent généralement une combinaison de ces approches :

« It is important to note that the strategies are not mutually exclusive; often a combination of approaches will be the most appropriate strategy for ensuring full and accurate records are created and retained. » (Queensland State Archives, 2006: 8).

« There is no generic solution for creating and maintaining records of web-based activity. The most appropriate recordkeeping strategy is likely to use a combination of approaches, depending on the outcome of an analysis of an agency's particular circumstances. » (Australian Government, 2004: 3).

Les politiques préconisent les approches suivantes : moissonnage du site Web afin d'en créer un portrait à des fréquences prédéterminées, micro-archivage (archivage de documents particuliers) et documentation des changements ainsi que des transactions. Les techniques suggérées sont principalement l'archivage à distance et la création d'un entrepôt local du site qui sert à répliquer le contenu du site Web au moment de sa publication :

« The local site repository is contained in a separate storage device. Its purpose is to provide a means of reconstructing the information content of a high-risk web site, and helps a department dispose of recorded information through 1) controlled deletion and 2) transfer of archival records to the University Archives. The local site repository is a stopgap method. It is not the preferred method for managing Web-based records or their legal disposition. » (University of Toronto, 2006)

Pour une liste complète des approches et techniques possibles de même que les avantages et inconvénients qu'elles présentent, consulter Public Record Office Victoria, 2008 : 20-33.

- **Définir une stratégie de préservation des archives Web** : choix d'un support de stockage approprié et définition d'une technique de préservation (ex. : migration des données).
- **Former le personnel** aux règles de création et de gestion de documents authentiques et fiables (ex. : formation sur les métadonnées à associer à un document Web au moment de sa création) et les sensibiliser à l'importance de documenter leurs activités et opérations de gestion des sites Web.

Remarque générale :

Lorsqu'une politique générale pour la gestion d'archives existe, nous suggérons d'y inclure la liste des formats des documents créés et gérés par

l'organisme. Sur cette liste devraient figurer les sites et documents Web (pour un exemple, voir : University of Tasmania, 2010c) :

« Managing the records of online resources and services needs to be one part of an information and records management strategy that encompasses all the information created or received by a public authority that is evidence of its business activity. » (Queensland State Archives, 2006: 9)

5.3.3.2.2 *Modification, mise à jour, retrait des contenus*

Les mesures suivantes visent à documenter toutes les actions accomplies sur le site public d'une organisation et les opérations de sa gestion. Pour ce faire, l'organisme doit :

- **Gérer toutes les modifications dans un journal historique.**
- **Capturer toutes les transactions dans un journal des transactions.**
- **S'assurer que les anciennes versions du site Web ou des documents qui le composent font l'objet d'une règle de conservation et que, au moment de leur retrait, elles soient déposées dans un dépôt.**

5.3.3.2.3 *Conservation à des fins administratives et historiques*

Les mesures indiquées dans cette section portent sur les règles à respecter lors de l'application de la stratégie d'archivage. L'organisme doit :

- **Demander un avis juridique avant de capturer le site Web ou les transactions ayant lieu sur le site Web** (cette mesure est très importante pour les archives nationales qui souhaitent entreprendre un projet d'archivage du Web de grande envergure) :

« Agencies may consider taking legal advice on capturing transactions, and certainly should ensure that this information is held securely. » (Public Record Office Victoria, 2008)

- **Autoriser le service des archives à réaliser les portraits périodiques du site Web :**

« The Service allows Records Services staff (only) to login and create web collections ("Collections"), catalogue the websites associated with a Collection, archive websites in the Collection, monitor the archiving process, search and browse the Collections when complete, and administer access to these Collections. » (The University of Melbourne, 2008f)

- **Capturer les métadonnées relatives au robot et aux objets collectés au cours de la collecte** (Public Record Office Victoria, 2008: 24-25).
- **Intégrer ces portraits au calendrier de conservation à titre de documents d'archives par le biais d'une règle de conservation appropriée.**

« Quarterly captures may become a record in their own right, with long term historical value to researchers who wish to track the history of the UOM website or to track particular changes in organizational structure or campus life. » (The University of Melbourne, 2008f)

- **S'assurer que les portraits du site Web organisationnel rencontrent les critères de qualité d'un document d'archives.** Les portraits réalisés doivent préserver le contenu, la forme et les fonctionnalités du site Web au moment de la capture :

« Ensuring that each snapshot is a full and accurate record of resources and services at a particular point in time requires capturing all of the content, layout and functionality of the website. Therefore, snapshots need to collect:

- the text or documents displayed;
 - any scripts that run on the page; and
 - any programs, browser software and plug-ins that are used. »
- (Queensland State Archives, 2006: 24)

Certaines fonctionnalités doivent être par contre modifiées afin de préserver l'authenticité du portrait du site Web telles que les statistiques sur le nombre des visiteurs (Queensland State Archives, 2006).

La réalisation de portraits Web est une approche pertinente lorsque le site Web à moissonner est de nature statique. Afin de diminuer ces inconvénients, la technique du moissonnage à distance doit être combinée aux journaux historiques.

En cas de refonte générale du site Web, il faut aviser l'équipe responsable de l'archivage des sites Web dans l'organisation et ne pas ajouter de nouveaux changements tant que le site Web n'a pas été capturé. Il ne faut pas interdire la collecte du site Web :

« Please ensure that your sites are set up to allow access to our crawler. Sites are crawled (archived) at a very polite request rate. This means that our crawler should not cause you any difficulties when archiving sites. » (TNA , [s.d.]b).

- **Détruire les documents qui ne présentent pas de valeur permanente lorsqu'ils atteignent la fin de leur cycle de vie.** Ces documents sont souvent une copie qui reprend une information gérée dans un autre système ou ne

présentent aucune valeur administrative ou historique (Archives Office of Tasmania, 2009).

- **Verser, dans le respect de la procédure de versement si elle existe, les sites Web retirés et les portraits des sites Web actuels aux archives nationales, le cas échéant** (Archives New Zealand, 2006: 12; NAA, [s.d.]: 2).

« Agencies should take snapshots of their entire website. These triggers are determined by the agency in consultation with the National Archives of Australia. However, all agencies should take snapshots when there is major change to, or decommissioning of, the website. These snapshots are ‘Retain as national archives’ records.

Once an agency has no further need of these snapshots, they may be transferred to the National Archives. » (NAA, [s.d.]: 2)

Pour un exemple de procédure de transfert, voir NARA, 2004 et Public Record Office Victoria, 2008.

5.3.3.3 Exemples de règles de conservation

Afin de dépasser une logique d’accumulation et de constituer des archives Web représentatives de l’évolution du site, des activités qui y ont eu lieu et des événements et actions qu’ils documentent, l’analyse des risques doit conduire à l’élaboration d’un ensemble de règles de conservation propres au site Web. Nous avons démontré qu’un site Web est un objet complexe et composite. Une seule règle ne pourrait pas couvrir toutes ses composantes. Ces dernières évoluent de manière différente. De même, l’archivage des documents individuels diffusés sur un site Web ou des documents relatifs à sa gestion n’est pas suffisant pour rendre compte des activités en ligne d’une organisation. De plus, l’accumulation des portraits du site Web va à l’encontre de la logique sélective des archives, le but n’étant pas la quantité, mais la qualité et la représentativité des archives constituées. Nous avons noté qu’il n’existe pas qu’un seul acteur responsable de l’archivage du Web organisationnel. Dans certains cas, les archives nationales partagent cette responsabilité avec les organismes en prenant la charge d’archiver les portraits Web à des fins historiques.

Partant de ces constats, nous suggérons l’élaboration de six règles de conservation. La première porte sur les portraits Web. Qu’ils soient gérés par l’organisme ou par les archives nationales, les portraits d’un site Web sont des documents d’archives « provoqués ». L’organisme responsable du site Web ne

pourrait accumuler à l'infini ces archives. Une fois que ces archives ne sont plus nécessaires aux fins administratives, elles peuvent être éliminées, conservées dans l'organisation ou versées aux archives nationales si elles présentent une valeur historique, le cas échéant. Dans les textes de politiques étudiées, les portraits du site Web sont considérés ayant une valeur historique :

« These snapshots are 'Retain as national archives' records (i.e. they have been assessed as having ongoing archival value) under AFDA class 1935. Once an agency has no further need of these snapshots, they may be transferred to the National Archives. » (NAA, [s.d.]: 2)

La fréquence de l'archivage est très variable. Elle semble dépendre en toute évidence des risques que présente le site Web.

Lorsque ces archives présentent une valeur historique, il faudrait déterminer, dans la masse accumulée, les portraits les plus représentatifs des actions et des événements ayant lieu dans l'organisme et qui sont documentés par le site Web. Le processus d'évaluation de ces portraits se précisera dans le temps, il faudrait toutefois l'encadrer par l'application d'une règle de conservation. Cette remarque vaut autant pour les organismes responsables de la gestion de leurs archives définitives que pour les archives nationales si elles partagent cette responsabilité d'archivage avec l'organisme producteur.

La deuxième règle traduit un processus de micro-archivage. Cette règle doit porter sur les documents diffusés sur le site Web. Nous l'avons démontré, les documents propres au Web qu'ils soient de nature statique ou dynamique ainsi que ceux produits dans d'autres environnements et qui ont migré sur le Web ont acquis un nouveau contexte d'usage et de nouvelles fonctionnalités. Ces documents doivent être archivés. La technique de moissonnage à distance a ses limites. Les portraits du site Web, pris selon une certaine périodicité qui reflète de manière générale l'évolution de l'ensemble du site Web, ne peuvent pas garder la trace de certains changements importants affectant une section du site Web ou capturer l'ensemble des documents qui sont produits.

La troisième règle couvre la documentation Web (niveau méta-archivage). Cette règle est un moyen pour mettre en évidence l'importance des documents qui gardent la traçabilité des actions menées sur le site Web tant par ceux qui sont responsables de sa gestion (ex. : affichage, mise à jour, retrait de l'information) que par ses usagers (ex. : soumission de commentaires ou de requêtes, réalisation d'une transaction).

La quatrième règle couvre la documentation de la technologie et des programmes qui permettent de gérer le site Web. Cette règle porte à la fois sur des composantes technologiques (ex. : scripts, applications) et sur la documentation relative à la gestion de ces composantes technologiques.

La cinquième règle concerne les sites Web inactifs. Cette catégorie d'archives est souvent négligée dans les politiques analysées; or les archives d'un site révolu peuvent présenter un intérêt historique pour l'organisme et la société.

Finalement, la sixième porte sur les microsites associés au site Web public principal. Souvent ces microsites sont de nature événementielle très éphémère. Cette règle doit également préciser l'unité responsable de l'archivage, car, dans bien de cas, les microsites sont sous une responsabilité partagée entre deux ou plusieurs organismes.

Il importe de choisir des supports de conservation appropriés aux caractéristiques des documents archivés et d'établir une stratégie de pérennisation des documents destinés pour une conservation permanente.

Les archives Web constituées doivent être soumises à un processus d'évaluation continu encadré par une règle de conservation. Les règles de conservation doivent être, de temps à autre, remaniées. Les critères actuels de sélection de ces portraits évolueront au rythme de l'évolution du site Web lui-même. Les règles de conservation doivent être mises à jour en conséquence.

5.3.4 Limites de la politique type

Le modèle de politique que nous avons proposé présente quelques limites. Tout d'abord, les mesures indiquées sont inspirées des politiques analysées. Ces politiques proviennent généralement d'une culture archivistique qui diffère de la culture archivistique québécoise. Les politiques analysées sont établies en cohérence avec la stratégie nationale de gestion des archives d'un pays, des ressources disponibles aux archives nationales ainsi que des besoins et des ressources organisationnels. Ces conditions doivent être prises en compte lors du développement d'une politique d'archivage du Web pour son organisation (Brown, 2006; McClure, 2005).

Quelques-unes des politiques utilisées à titre de référence sont publiées depuis quelques années et n'ont pas été revues. Toutefois, nous croyons que les archives Web posent les mêmes difficultés de gestion (quand et comment intervenir) que les archives traditionnelles (archives sur support papier ou archives numériques bureautiques) et confrontent les organisations aux mêmes exigences de préservation à long terme. La technologie Web est une technologie mouvante, la mise à jour d'une politique d'archivage du Web organisationnel s'impose de manière plus récurrente. C'est pour cette raison que nous avons énoncé des mesures très générales qui ne portent pas sur une technologie particulière.

Enfin, la politique type porte sur les sites Web publics uniquement même si quelques politiques analysées appliquent les mêmes mesures d'archivage aux sites intranet et extranet. Certes, les principes directeurs et la démarche d'analyse des risques énoncés dans la politique type s'appliquent à l'ensemble des systèmes Web d'une organisation. Il sera prudent de mener une étude plus approfondie de ces autres types de systèmes, des usages qui en sont faits dans différents contextes organisationnels et de la nature des documents qui y sont produits ou diffusés pour s'assurer que les mesures développées pour un site Web public s'appliquent de la même manière aux intranets et extranets, par exemple.

6. Conclusion de la recherche

Ce dernier chapitre vise à établir le bilan de notre recherche. Dans un premier temps, nous présentons un résumé de notre étude. Ensuite, nous exposons les principales contributions de notre recherche au domaine de la gestion des archives. Enfin, nous établissons les limites de la recherche et indiquons les possibilités de réalisation de recherches futures.

6.1 Résumé de la recherche

La présente recherche avait pour but de développer une meilleure compréhension des archives Web et de décrire les pratiques d'archivage du Web organisationnel. Elle visait à répondre aux questions de recherche suivantes :

1. Que recommandent en général les politiques d'archivage du Web organisationnel?
2. Quelles sont les principales caractéristiques des archives Web?
3. Quelles pratiques d'archivage du Web organisationnel sont mises en place dans des organisations au Québec?
 - a. Quels sont les principaux intervenants impliqués dans le processus d'archivage du Web organisationnel?
 - b. Quelles sont les approches et les techniques d'archivage du Web employées dans un contexte organisationnel?
 - c. Quels facteurs influencent la mise en œuvre d'une politique d'archivage du Web organisationnel?

L'approche méthodologique adoptée pour cette recherche est de nature qualitative. Notre étude est descriptive cherchant à explorer des pratiques documentaires liées à la gestion et à l'archivage des sites Web publics. Trois principaux modes de collecte de données ont été utilisés, soit l'analyse d'une sélection de politiques d'archivage du Web organisationnel, l'observation du site Web public de 11 organismes au Québec et des types de documents qui y sont

diffusés et l'entrevue en profondeur en personne auprès des intervenants impliqués dans le processus de gestion et d'archivage des sites Web publics. Les données ont été recueillies à l'aide de grilles d'analyse et ont fait l'objet d'une analyse de contenu qualitative. Quatre unités d'analyse ont été considérées, soit : les sites Web publics observés, les archives Web produites grâce à ces sites Web publics, les stratégies d'archivage de ces systèmes Web et les intervenants impliqués dans leur processus d'archivage.

Les résultats obtenus montrent qu'il existe une dissimilitude entre les recommandations des politiques d'archivage développées par des institutions archivistiques nationales ou ayant mené un projet d'archivage, et les pratiques observées dans les milieux d'étude. Selon ces politiques, l'archivage des systèmes d'informations Web, en général, et du site Web public d'une organisation, en particulier, est une obligation, d'une part, lorsque l'organisme est soumis aux dispositions de la loi sur les archives et, d'autre part, pour assurer la continuité des activités de l'organisation et pour témoigner de sa présence en ligne. Les politiques affirment la nature archivistique des sites Web publics et identifient, à ce niveau, les documents devant être archivés et les acteurs organisationnels devant être impliqués dans ce processus. Ces politiques n'imposent pas une stratégie particulière de capture des sites Web. Elles décrivent les avantages et les inconvénients de différentes méthodes et techniques d'archivage. Elles préconisent l'analyse des risques pour établir un choix réfléchi de la stratégie d'archivage la plus appropriée aux besoins de l'organisme.

Les politiques soulignent également l'importance d'adopter une démarche systématique fondée sur les principes de base en gestion des documents, de développer des outils de gestion adaptés et de prévoir une stratégie de préservation des sites Web capturés. Cette position révèle que, dans le contexte du Web, les principes de gestion documentaire restent les mêmes, mais les méthodes de leur mise en œuvre changent.

Alors que les institutions archivistiques, à travers les politiques qu'elles diffusent, reconnaissent la valeur archivistique d'un site Web et l'intérêt de procéder

à son archivage, sur le terrain, la réalité est différente. Deux cas de figure se présentent. Dans le premier cas, les organismes ne procèdent pas à l'archivage de leur site Web organisationnel puisque, d'un côté, tout ce qui est publié sur ce dernier est géré dans un autre environnement, et de l'autre, le site n'est qu'un média supplémentaire pour la diffusion de l'information. Les organismes ne croient pas à l'obligation d'archiver leurs sites Web. Les copies de sauvegarde réalisées par les webmasters répondent aux besoins actuels de l'organisation. Or, les résultats de l'observation des sites Web et des documents qui y sont diffusés ont bien démontré que le site Web public d'une organisation comporte certainement des documents produits dans d'autres environnements, mais leur présence sur le site Web leur procure des caractéristiques héritées du média qui les incorpore (ex. : format Web, fonctionnalités Web) et un autre contexte d'usage. Plus encore, de nouveaux types de documents propres au Web organisationnel ont été également identifiés. Ces documents résultent de la conduite des affaires de l'organisation ou des fonctionnalités d'interaction qu'elle offre à ses clients et partenaires. N'oublions pas aussi de noter que le développement d'un site Web et ses opérations de gestion génèrent des documents qu'il importe de conserver à deux finalités : témoigner de l'évolution du site Web et assurer l'authenticité des archives Web constituées.

Dans le deuxième cas, les organismes procèdent à l'archivage de leur site Web public grâce à l'implication active des archivistes dans ce processus. Les archivistes s'assurent qu'en plus des copies de sauvegarde qui sont produites, un portrait du site Web est réalisé à des intervalles réguliers. Ce portrait est réalisé par archivage à la source (copie du répertoire du site Web) ou par moissonnage à distance (le portrait est réalisé grâce à un robot de collecte). Il est à souligner que les choix d'archivage sont établis souvent par convenance et ne s'appuient pas sur une analyse des risques que présente le site Web. Seulement quelques-uns des sites étudiés ont développé quelques règles de conservation ou des procédures d'archivage de leur site Web. Les archives du site Web ainsi constituées sont conservées sur des supports optiques (CD-ROM ou DVD). Les archivistes rencontrés étaient conscients des

limites de ce choix et des problèmes de pérennité et d'exploitabilité des archives Web qu'il engendrait.

L'existence d'un cadre de référence à l'échelle nationale, la perception des archivistes du site Web en tant qu'objet archivistique, le niveau de collaboration entre les différents acteurs impliqués dans la gestion et l'archivage d'un site Web de même que les ressources disponibles ont été identifiés comme facteurs affectant la mise en œuvre d'un programme d'archivage pour le Web organisationnel. À cela, nous pouvons aussi ajouter que les politiques de référence dans le domaine ne sont pas faciles à adapter. D'une part, les recommandations sont dispersées, il est très difficile d'entrevoir les actions à mettre en place tout au long du cycle de vie du site Web. D'autre part, le modèle proposé pour la gestion des risques ne prend pas en considération la nature des documents diffusés sur le site Web (ex. : modes de création, valeurs).

L'ensemble de ces constats nous a amenée à discuter de la valeur archivistique d'un site Web organisationnel et à proposer un modèle de politique type pour l'archivage des sites Web adapté aux caractéristiques des archives Web qui tient compte des scénarios de gestion des sites Web tout au long de leur cycle de vie tout en faisant appel à plusieurs standards documentaires.

Les apports de cette recherche sont donc de nature théorique, méthodologique et pratique. Nous les exposons au point suivant.

6.2 Apports théoriques de la recherche

Cette recherche contribue à la reconnaissance de la valeur archivistique des sites Web organisationnels et à la nécessité d'archiver ces nouveaux objets documentaires, sans quoi une portion de la mémoire organisationnelle, voire même sociale, sera condamnée à l'oubli ou encore sous le contrôle d'autres entités. Il s'agit d'un message de sensibilisation que nous souhaitons lancer à la communauté archivistique à travers cette recherche.

Cette recherche démontre que l'usage des technologies Web dans les organisations provoque la genèse de nouveaux types de documents propres au Web organisationnel et la migration en ligne de différents types de documents où il acquiert de nouvelles caractéristiques. Ces documents sont une composante de la mémoire documentaire des organisations.

Certes les technologies Web sont en constante évolution et les politiques analysées sont le fruit d'une réflexion en cours sur les méthodes et techniques d'archivage du Web qui évolueront grâce aux expériences de différents milieux organisationnels. Néanmoins, l'analyse comparative de ces textes nous a permis de circonscrire la notion des archives Web en identifiant les différentes unités d'archivage et en définissant le statut de certains documents (ex. : les portraits périodiques d'un site sont des documents d'archives qu'il faudrait gérer en établissant des règles de conservation). L'investigation sur le terrain nous a offert l'opportunité de développer une meilleure connaissance des pratiques d'archivage peu ou rarement documentées dans la littérature. Les résultats obtenus ne nous permettent pas toutefois de brosser un portrait général de la situation dans le milieu organisationnel.

Notre recherche a permis d'enrichir le modèle de gestion des risques qui se centre sur les caractéristiques technologiques des sites Web organisationnels sans porter attention à la nature des documents qui y sont diffusés. Les résultats de cette démarche auront des retombées pratiques dans la mesure où l'ajout de la composante Nature des documents au modèle permettrait d'établir un choix plus réfléchi de la stratégie d'archivage.

6.3 Apports méthodologiques

Les outils de collecte des données qui ont été développés ont permis d'obtenir des données riches sur le phénomène à l'étude. Ils pourront également être intégrés à la démarche d'analyse des risques et adaptés afin d'établir un portrait des caractéristiques techniques et documentaires des sites Web organisationnels. Ces outils pourront être utilisés conjointement par les différents intervenants impliqués dans la gestion et l'archivage des sites Web organisationnels et servir de tableau de

bord pour suivre l'évolution de ces sites sur le plan technologique, mais aussi au niveau de l'actif documentaire et mettre à jour, en conséquence, la politique d'archivage.

6.4 Apports pratiques

L'apport majeur de cette recherche réside dans le développement d'un modèle de politique pour l'archivage des sites Web organisationnels. Ce modèle pourrait ainsi soutenir les archivistes lors de la mise en place d'une démarche d'archivage du Web au sein de leur organisation en l'absence d'un cadre de référence en la matière. Il prescrit une démarche d'archivage décentralisée (le modèle pourrait être adopté par différentes unités documentaires) et de nature collaborative (la mise en œuvre de ce modèle devrait impliquer différents acteurs organisationnels). Ce modèle est inspiré de politiques développées par des institutions nationales dans le domaine des archives, mais aussi par des organisations ayant un programme d'archivage du Web bien établi. Ce modèle se veut général afin qu'il puisse être adapté à différents contextes organisationnels.

Les résultats obtenus permettent de dégager un ensemble de recommandations qui visent les archives nationales, les organisations, les associations professionnelles et les écoles de formation.

Les archives nationales sont invitées à établir une politique de référence en matière d'archivage du Web organisationnel afin de remédier aux nombreux doutes et hésitations que cette absence de position claire et officielle a créés. Cette politique pourrait ainsi contribuer à sensibiliser les archivistes à l'importance de gérer cet objet documentaire. Nous avons noté qu'en l'absence d'une telle politique, les organismes ne se sentent pas dans l'obligation d'assurer l'archivage de leur site Web ou lorsqu'ils souhaitent le faire, ils ne sont pas encadrés dans leur démarche. Les choix qui en résultent ne sont pas toujours pertinents et affectent, à long terme, la qualité des archives constituées (ex. : choix du support ou de la technique d'archivage).

Les archives nationales doivent aussi recommander, voire offrir des outils d'archivage et encadrer le processus de transfert des archives Web définitives. Les organisations manquent souvent de ressources matérielles, humaines et financières pour entretenir à long terme les archives Web. Dans ce cas, une démarche collaborative entre les organismes et les archives nationales est fortement recommandée.

Quant aux organisations, il leur faut adapter une démarche systématique et intégrer les sites Web aux systèmes de gestion documentaire (ex. : règles de conservation). Une politique d'archivage du Web doit être développée et son application doit faire l'objet d'un contrôle continu. Les archivistes doivent jouer un rôle plus actif à ce niveau en s'impliquant davantage dans le processus de conception et gestion des systèmes Web. Les rôles et les responsabilités doivent être clairement définis.

Les associations professionnelles doivent jouer un rôle proactif et offrir une formation continue sur l'archivage du Web. Ce sujet devrait faire l'objet de nombreuses études, rencontres, retours d'expériences, sondages et surtout encadrement. C'est ainsi que les connaissances pourront évoluer. L'Association des archivistes du Québec a offert quelques formations sur le sujet et inclut depuis quelques années la problématique d'archivage du Web dans son programme de congrès annuel. Des lignes directrices, des modèles de procédures, des outils peuvent augmenter ces efforts.

À l'examen de quelques programmes d'études en sciences de l'information au Québec, nous avons noté que l'archivage du Web est peu présent dans les cursus de formation. Pourtant à l'ère du numérique, on ne peut plus négliger cette question. Les étudiants doivent être sensibilisés à cette problématique. Les associations et les écoles de formation se compléteront à ce niveau.

Nous tenons à souligner la rareté des écrits sur ce sujet en langue française. La rareté des recherches sur l'archivage du Web peut contribuer à ce désintérêt pour l'archivage du Web tant dans les programmes de formation que dans les milieux

organisationnels. Des études, notamment des études expérimentales et longitudinales doivent être menées pour assister les organisations dans l'implantation d'une stratégie d'archivage, pour établir les critères d'évaluation de l'efficacité des solutions choisies, pour évaluer la qualité des archives constituées. Ces recherches témoigneront, par la même occasion, de l'évolution des pratiques d'archivage du Web et du Web 2.0 dans les organisations et constitueront un support au matériel pédagogique ou de formation sur le sujet. Des analyses comparatives doivent être également menées de manière rigoureuse afin d'étudier les forces et faiblesses des solutions techniques disponibles.

La notion de gestion de risques, empruntée du domaine des affaires, a séduit le domaine des archives. Rares sont les écrits qui décrivent comment cette notion a été ou peut être adaptée en archivistique (Lemieux, 2004). La méthodologie de l'analyse des risques doit être aussi considérée et expérimentée dans des recherches sur le terrain.

6.5 Limites de la recherche

Le recours à une approche qualitative s'est avéré un bon choix pour étudier en profondeur les pratiques d'archivage du Web dans un contexte organisationnel. L'étude d'autres systèmes et d'autres milieux devrait être envisagée afin de vérifier la transférabilité des résultats de cette recherche dans d'autres contextes de même que la transférabilité du modèle de la politique que nous avons proposé aux autres types de systèmes Web.

Notre étude, exploratoire et descriptive, porte sur les sites Web publics. Nous étions contrainte à l'analyse des sites Web public, l'accès aux autres systèmes Web n'étant pas aisé. De même, l'observation à distance ne nous a pas permis d'étudier d'une manière approfondie les caractéristiques techniques des sites Web publics et les liens qu'ils présentent avec d'autres systèmes (ex. : intranet, extranet, base de données, système de gestion de contenu, etc.). À cause de la difficulté à trouver des sites d'étude et les délais pour réaliser cette recherche, nous avons décidé d'interviewer uniquement les archivistes et les webmestres. D'autres intervenants

dans le processus de gestion et d'archivage du Web auraient pu être rencontrés (ex. : créateurs de contenus Web, spécialistes des systèmes d'information). Peu de documentation a pu être recueillie puisque les pratiques de gestion d'un site Web ne sont pas souvent documentées dans les milieux étudiés. Bien que nous ayons veillé à sélectionner, selon des critères bien définis, un échantillon des types de documents qui peuvent se retrouver à différents niveaux de profondeur d'un site Web, les pages Web observées ne reflètent pas nécessairement tous les types des documents qui peuvent se retrouver sur un site Web organisationnel. L'étude d'un échantillon plus important de documents Web contribuera à établir une typologie plus complète des archives Web.

Malgré ces contraintes, les instruments de collecte développés pour cette recherche nous ont permis de collecter des données riches et de qualité qui nous renseignent sur les caractéristiques des archives Web et des pratiques de leur archivage.

6.6 Recherches futures

Trois axes de recherches futures se dessinent. Premièrement, les résultats de cette recherche doivent être complétés par l'étude des autres types de systèmes Web organisationnels (ex. : intranets, extranets) du côté serveur. Les résultats d'une telle étude nous permettront d'étudier l'ensemble des systèmes Web d'une même organisation, de dresser un portrait global des types d'archives qui pourraient s'y retrouver et d'entrevoir les stratégies possibles de leur archivage. Ainsi, une compréhension plus globale du processus d'archivage nous permettrait de vérifier la pertinence du modèle de politique type que nous proposons dans le cadre de cette étude et sa transférabilité aux autres systèmes Web.

Le Web 2.0 étant de plus en plus utilisé à différentes fins par les organisations (CEFRIO, 2012; Dufour, 2008-2009), des politiques d'archivage du Web 2.0 sont à développer (ARMA International, 2011; NARA, [s.d]a; NSW, 2009b). Il importe de mener une étude sur les différentes plateformes Web 2.0 utilisées dans les organisations et les possibles valeurs archivistiques que véhiculent ces contenus.

D'autres recherches peuvent être menées afin de systématiser la reconnaissance de nouveaux types de documents qui émergent suite à l'implantation d'une nouvelle technologie ou l'évolution de son usage. Des recherches qui associent la notion de genre et de diplomatique pour l'étude des archives numériques sont en cours. Nous pouvons nous associer à ces recherches en considérant les sites Web organisationnels comme cas d'étude. L'étude des conditions de validité d'un document d'archives produit par des systèmes Web devrait faire suite à cette recherche.

Enfin, la sélection des archives pour une conservation permanente dépend de quelques critères, dont leurs possibles usages pour la recherche et le témoignage. Une des raisons qui expliquent le manque d'intérêt pour archiver les sites Web vient du fait qu'actuellement ces archives ne sont pas utilisées comme matériaux de recherche scientifique ou historique. À cet effet, nous souhaitons mener des études d'utilisateurs dans le but de préciser les besoins pour lesquels ces archives doivent être préservées.

L'archivage du Web, nouveau champ de spécialisation en archivistique, doit faire l'objet d'une étude plus poussée.

Bibliographie

- (2012). *WAC - Web Archive Collective at Harding University*. [En ligne]. Adresse Url: <http://www.harding.edu/fmccown/research/wac/> (Page consultée le: 05 mai 2011).
- AFNOR - Association française de normalisation. (2004). *Records management: gestion et conservation des documents en toute sécurité*. Saint-Denis La Plaine: AFNOR. 144 p.
- AFNOR - Association française de normalisation. (2009). *NF Z42-013 - Archivage électronique: spécifications relatives à la conception et à l'exploitation de systèmes informatiques en vue d'assurer la conservation et l'intégrité des documents stockés dans ces systèmes*. St Denis La Plaine. 44 p.
- Alberts, I. (2009). *Exploitation des genres de textes pour assister les pratiques textuelles dans les environnements numériques de travail : le cas du courriel chez des cadres et des secrétaires dans une municipalité et une administration fédérale canadiennes*. Thèse de doctorat. Montréal: Université de Montréal, École de bibliothéconomie et des sciences de l'information. 383 p.
- Albertsen, K. (2003). « The Paradigma Web harvesting environment: the Paradigma project at the National Library of Norway. » Dans: *3rd ECDL workshop on web archives in conjunction with the 7th european conference on research and advanced technologies for digital libraries*, Trondheim, Norvège, 21 août, pp. 14. Adresse Url: <http://bibnum.bnf.fr/ecdl/2003/> (Page consultée le: 17 octobre 2006).
- Andrieu, O. (2005). « L'archivage du Web: utopie ou projet stratégique? » *Technologies Internationales* (115), pp. 35-40.
- Archive-it. ([s.d.]). *K-12 Web archiving*. [En ligne]. Adresse Url: <http://www.archive-it.org/k12/> (Page consultée le: 27 février 2010).
- Archives New Zealand. (2006a). *Guide to developing recordkeeping strategies for websites, issued June 2004/ revised June 2006*. Recordkeeping guides. G20. 16 p. Adresse Url: <http://web.archive.org/web/20100525160610/http://continuum.archives.govt.nz/files/file/guides/g20/index.html> (Page consultée le: 27 janvier 2009).
- Archives New Zealand. (2006b). *A guide to developing a recordkeeping policy*. Recordkeeping guides. 20 p. Adresse Url: http://archives.govt.nz/sites/default/files/g6_6.pdf (Page consultée le: 11 décembre 2012).
- Archives New Zealand. (2009). *Guide to managing web records*. [En ligne]. Adresse Url: <http://archives.govt.nz/advice/continuum-resource-kit/continuum-publications-html/g20-guide-managing-web-records> (Page consultée le: 27 janvier 2009).

- Archives Office of Tasmania. (2009). *State records guideline No. 15: recordkeeping strategies for websites and web pages*. [En ligne]. Adresse Url: http://www.linc.tas.gov.au/_data/assets/pdf_file/0007/334978/State_Records_Guideline_No_15_Revised.pdf (Page consultée le: 27 janvier 2009).
- ARMA International. (2009). *Website records management*. Overland Park. 22 p.
- ARMA International. (2011). *Implications of web-based, collaborative technologies in records management : an American national standard*. Overland Park. 21 p.
- Arvidson, A. et al. (2000). « The Kulturarw3 Project - The Royal Swedish Web Archiw3e - an example of "complete" collection of web pages. » Dans: *66th IFLA council and general conference*, Jerusalem, Israel, 13-18 août. Adresse Url: <http://www.ifla.org/IV/ifla66/papers/154-157e.htm> (Page consultée le: 17 octobre 2006).
- Aschenbrenner, A. (2004). *Web Archiving - bibliography*. Austrian On-Line Archive. [En ligne]. Adresse Url: <http://www.ifs.tuwien.ac.at/~aola/link/WebArchiving.html> (Page consultée le: 16 octobre 2006, 26 février 2011).
- Australian Government. ([s.d.]). *Web guide: archiving websites*. [En ligne]. Adresse Url: <http://webguide.gov.au/recordkeeping/archiving-a-website/> (Page consultée le: 23 avril 2007).
- Australian Government. Department of finance and deregulation. Australian government information management office. (2007). *Guide to minimum Website standards - electronic record keeping and archiving*. [En ligne]. Adresse Url: <http://www.agimo.gov.au/archive/mws/archiving.html> (Page consultée le: 23 avril 2007).
- Australian Government. Information Management Office. (2004). *Archiving web resources. Better practice checklist - 7: practical guide for effective use of new technologies in government*. Archiving web resources. 12 p. Adresse Url: <http://web.archive.org/web/20090614181325/http://www.finance.gov.au/e-government/better-practice-and-collaboration/better-practice-checklists/docs/BPC7.pdf> (Page consultée le: 23 avril 2007).
- Babbie, E. R. (2005). *The basics of social research*. 3e éd. Belmont, CA: Thomson/Wadsworth. 519 p.
- BAC - Bibliothèque et Archives Canada. (1999). *Gestion de l'information sur Internet et les intranets pour assurer l'accès à long terme et la reddition de comptes: Guide de mise en œuvre*. Ottawa: Forum sur la gestion de l'information (GI). Groupe de travail (GT) sur les réseaux Internet / intranets. 75 p. Adresse Url: <http://www.lac-bac.gc.ca/gouvernement/produits-services/007002-2032-f.html> (Page consultée le: 15 octobre 2005).
- BAC - Bibliothèque et Archives Canada. (2007-2008). *Archives du Web du gouvernement du Canada - FAQ*. [En ligne]. Adresse Url: <http://www.collectionscanada.gc.ca/archivesweb/faq/index-f.html> (Page

consultée le: 23 avril 2007).

- BAC - Bibliothèque et Archives Canada. Groupe de coordination des collections électroniques. (1998). *Politiques et directives relatives aux publications électroniques diffusées en réseau*. Bibliothèque nationale du Canada. Adresse Url: http://epe.lac-bac.gc.ca/100/200/301/nlc-bnc/networked_epubs-ef/9/8/index-f.html (Page consultée le: 28 décembre 2006).
- Bachimont, B. (2000). « L'archive numérique, entre authenticité et interprétabilité. » *Archives*, **32**(1), pp. 3-15.
- Bachimont, B. et al. (2005). « Documenter et partitionner une archive du Web: vers le dépôt légal d'un domaine media. » Dans: *ICHIM 05 - Digital culture & heritage / Patrimoine & culture numérique*, Paris, Bibliothèque nationale de France, 21-23 septembre. Archives & Museum Informatics Europe, pp. 21. Adresse Url: <http://www.archimuse.com/publishing/ichim05/Bachimont.pdf> (Page consultée le: 15 octobre 2006).
- Balasubramanian, V. et Bashian, A. (1998). « Document management and web technologies: Alice marries the Mad Hatter. » *Communications of the ACM*, **41**(7), pp. 107-115.
- BAnQ - Bibliothèque et Archives nationales du Québec. (25 septembre 2012). *Campagne électorale 2012: Bibliothèque et Archives nationales du Québec collecte et conserve les sites Web*. [En ligne]. Adresse Url: http://www.banq.qc.ca/a_propos_banq/salle_de_presse/communiques_de_presse/2012/communiqu.html?c_id=d4f40db4-e009-4f6b-a1bc-0602783db997 (Page consultée le: 26 octobre 2012).
- BAnQ - Bibliothèque et Archives nationales du Québec. (29 novembre 2010). *Renouvellement de l'entente entre Bibliothèque et Archives nationales du Québec et la Bibliothèque nationale de France*. [En ligne]. Adresse Url: http://www.banq.qc.ca/a_propos_banq/salle_de_presse/communiques_de_presse/2010/com_2010_11_26.html (Page consultée le: 12 janvier 2011).
- BAnQ - Bibliothèque et Archives nationales du Québec. (2001). « Dépôt légal et Internet: deux univers à concilier. » *À rayons ouverts*, **55**(juillet-septembre).
- BAnQ - Bibliothèque et Archives nationales du Québec. (2007). *Publications gouvernementales du Québec en ligne*. [En ligne]. Adresse Url: <http://services.banq.qc.ca/sdx/pgq/static.xsp?page=presentation> (Page consultée le: 15 octobre 2006).
- BAnQ - Bibliothèque et Archives nationales du Québec. (2011). *Recueil de règles de conservation des documents communs des ministères et des organismes gouvernementaux*. [En ligne]. Adresse Url: http://www.banq.qc.ca/documents/services/archivistique_ged/Recueil_MO_Ressources_informatiques.pdf (Page consultée le: 26 octobre 2012).
- Barry, R. (2004). « Web sites as recordkeeping & "recordmaking" systems. » *The Information Management Journal* (novembre/décembre), pp. 26-32.

- Beagrie, N. (2003). *National preservation initiatives: an overview of developments in Australia, France, the Netherlands, and the United Kingdom and of related international activity*. Washington, D.C.: Council on Library and Information Resources, Library of Congress. 61 p. Adresse Url: <http://www.clir.org/pubs/reports/pub116/reports/pub116/pub116.pdf> (Page consultée le: 15 octobre 2006).
- Beagrie, N. et al. (2008). *Digital preservation policies study: part 1 Final report*. Charles Beagrie Limited, JISC. 60 p. Adresse Url: http://www.jisc.ac.uk/media/documents/programmes/preservation/jiscpolicy_p1finalreport.pdf (Page consultée le: 22 janvier 2009).
- Bearman, D. et Hedstrom, M. (2000). « Part eight: electronic records - 24. Reinventing archives for electronic records: alternative service delivery. » Dans: Jimerson, R.C., éd. *American archival studies: readings in theory and practices*. Chicago: The Society of American Archivists, pp. 549-567.
- Beghtol, C. (2001). « The concept of genre and its characteristics. » *Bulletin of the American Society for Information Science and Technology*, **27**(2), pp. 17-19.
- Beijing University. Computer networks and distributed systems laboratory. (2004). *Web InfoMall*. [En ligne]. Adresse Url: <http://www.infomall.cn/index-eng.htm> (Page consultée le: 22 janvier 2009).
- Benbasat, I. et al. (1987). « The case research strategy in studies of information systems. » *MIS Quarterly*, pp. 369-386.
- Bergeron, P. (1992). « La gestion des archives électroniques : quelques questions-clés à considérer. » *Archives*, **23**(3), pp. 51-70.
- Bergeron, P. (1997). « A qualitative case study approach to examine information resources management. » *Canadian Journal of Information and Library Science*, **22**(3/4), pp. 1-19.
- Bergquist, M. et Ljungberg, J. (1999). « Genre in action: negotiating genres in practice. » Dans: *HICSS-32, proceeding of the 32nd annual Hawaii international conference on system sciences*, Maui, Hawaii, 05-08 janvier, pp. 11. Adresse Url: <http://computer.org/proceedings/hicss/0001/00012/00012009.PDF> (Page consultée le: 07 janvier 2006).
- Bermès, É. (2009). « La conservation des documents numériques » Dans: Carbone, P. et Cavalier, F., éd. *Les collections électroniques, une nouvelle politique documentaire*. Paris: Éditions du Cercle de la librairie, pp. 209-228.
- Bermès, E. et Oury, C. (2009). « Web 2.0 et mémoire : De la conversation à la conservation. » *Documentaliste-Sciences de l'Information*, **46**(1), pp. 61-63.
- Beunen, A. C. et Schiphof, T. A. (2006). *Legal aspects of web archiving from a Dutch perspective*. Leiden: National Library in The Hague. 34 p. Adresse Url: http://www.kb.nl/hrd/dd/dd_projecten/webarchivering/documenten/KB_Legal_Aspects_WebArchiving_EN.pdf (Page consultée le: 22 janvier 2009).

- Biblioteca de Catalunya. (2011). *PADICAT (Digital Heritage of Catalonia): The Web Archive of Catalonia* [En ligne]. Adresse Url: <http://www.padicat.cat/en> (Page consultée le: 22 janvier 2009).
- Bibliotheca Alexandrina. (2008). *Web Archive*. [En ligne]. Adresse Url: <http://www.bibalex.org/English/initiatives/internetarchive/web.htm> (Page consultée le: 22 janvier 2009).
- Blatner, D. ([s.d.]a). *Things to do on the web when you're dead*. AfterLife.org. [En ligne]. Adresse Url: <http://www.afterlife.org/info.html> (Page consultée le: 21 janvier 2009).
- Blatner, D. ([s.d.]b). *AfterLife.org*. [En ligne]. Adresse Url: <http://www.afterlife.org/info.html> (Page consultée le: 21 janvier 2009).
- BnF - Bibliothèque nationale de France. (2008). *Dépôt légal de l'Internet*. [En ligne]. Adresse Url: http://www.bnf.fr/pages/infopro/depotleg/dl-internet_intro.htm (Page consultée le: 21 janvier 2009).
- BnF - Bibliothèque nationale de France. (2009). *Blog Lecteurs de la BnF: du journal intime au blog* [En ligne]. Adresse Url: <http://blog.bnf.fr/lecteurs/index.php/2009/04/06/du-journal-intime-au-blog/> (Page consultée le: 27 octobre 2012).
- BnF - Bibliothèque nationale de France. (2012a). *Archives de l'Internet*. [En ligne]. Adresse Url: http://www.bnf.fr/fr/collections_et_services/livre_presse_medias/a.archives_internet.html (Page consultée le: 13 mai 2012).
- BnF - Bibliothèque nationale de France. (2012b). *Parcours guidés des archives de l'Internet*. [En ligne]. Adresse Url: http://www.bnf.fr/fr/collections_et_services/anx_pres/a.parcours_guides_arch_internet.html (Page consultée le: 27 octobre 2012).
- Bonnenfant, B. (2008). *Définir une politique d'archivage du Web régional en bibliothèque municipale: l'exemple du web forézien*. Mémoire d'étude. Lyon: École nationale supérieure des sciences de l'information et des bibliothèques. 83 p.
- Boullier, D. et Ghitalla, F. (2004). « Le Web ou l'utopie d'un espace documentaire. » *Revue I3*, 4(1), pp. 173-189.
- Bourhis, T. (2000-2001). « Les archives électroniques: la question de l'intégrité. » *Archives*, 32(1), pp. 17-36.
- British Library. ([s.d.]). *UK Web Archive: F.A.Q.* [En ligne]. Adresse Url: <http://www.webarchive.org.uk/ukwa/info/faq> (Page consultée le: 22 janvier 2009).
- Brodie, N. (1997). « L'archivage des publications électroniques: le rôle de la Bibliothèque nationale du Canada. » *Nouvelles de la Bibliothèque nationale*, 29(10).
- Brody, R. (2000). « Word bites and user-defined documents. » *Econtent*, 23(5), pp. 16-18.

- Brown, A. (2006). *Archiving websites: a practical guide for information management professionals*. London: Facet. 238 p.
- Brügger, N. (2005). *Archiving websites: general considerations and strategies*. The Centre for Internet Research. Adresse Url: http://cfi.au.dk/fileadmin/www.cfi.au.dk/publikationer/archiving_underside/archiving.pdf (Page consultée le: 07 janvier 2006).
- Brygfjeld, S. A. (2002). « Access to web archives: the Nordic Web archive access project. » Dans: *68th IFLA council and general conference*, Glasgow, 18-24 août. Adresse Url: <http://ifla.queenslibrary.org/IV/ifla68/papers/090-163e.pdf> (Page consultée le: 16 octobre 2006).
- Bütikofer, N. et al. (2006). *Managing and archiving records in the digital era : changing professional orientations*. Baden: Hier + Jetzt. 131 p.
- Butler, T. (2003). « An institutional perspective on developing and implementing intranet- and internet-based information systems. » *Information Systems Journal*, **13**(3), pp. 209-231.
- Buzina, T. et Willer, M. (2007). « Croatian Digital Web Archive: from project to service of the National and University Library in Zagreb. » *DigItalia*, pp. 20-34.
- California Digital Library. (2008-2012). *Web Archiving Service*. [En ligne]. Adresse Url: <http://webarchives.cdlib.org/> (Page consultée le: 22 janvier 2009).
- Campbell, L. E. et Dulabahn, B. (2010). « Digital preservation: the Twitter archives and NDIIPP. » Dans: *iPRES 2010, proceedings from the 7th international conference on preservation of digital objects*, Vienne, Autriche, 19-24 Septembre, pp. 4. Adresse Url: <http://www.ifs.tuwien.ac.at/dp/ipres2010/papers/campbell-27.pdf> (Page consultée le: 30 janvier 2009).
- Cantello, G. et Stegenga, J. (2008). « Le contenu Web de la documentation gouvernementale au Canada: perspective d'archivage du Web par la Bibliothèque nationale (traduction française). » Dans: *World library and information congress: 74th IFLA general conference and council*, Québec, Canada, 10-14 août. Adresse Url: <http://www.ifla.org/IV/ifla74/index.htm> (Page consultée le: 22 janvier 2009).
- Capra, R. et al. (2007). « Selection and context scoping for digital video collections: an investigation of YouTube and blogs. » Dans: *JCDL'08*, Pittsburgh, Pennsylvania, USA, juin 16-20.
- Carbone, P. et Cavalier, F. (2009). *Les collections électroniques, une nouvelle politique documentaire*. Paris: Éditions du Cercle de la librairie.
- CEFRIO. (2012). *Gérer les enjeux et risques juridiques du Web 2.0*. 125 p. Adresse Url: <http://www.cefrio.qc.ca/publications/detail-dune-publication/publication/5329/45/> (Page consultée le: 02 février 2012).
- Céllard, A. (1997). « L'analyse documentaire. » Dans: Poupart, J. et al., eds. *La*

- recherche qualitative: enjeux épistémologiques et méthodologiques*. Québec: Morin, pp. 251-272.
- Chabin, M.-A. (2000). *Le management de l'archive*. Paris: Hermès science publications. 246 p.
- Chabin, M.-A. (2004). *MoReq: Modèle d'exigences pour l'organisation de l'archivage électronique, spécifications MoReq*. Paris: Archive 17. 140 p. Adresse Url: http://www.adbs.fr/records-management-29391.htm?RH=OUTILS_DOSTHEM&RF=1202925005805 (Page consultée le: 06 novembre 2005).
- Chabin, M.-A. (2005). « Document trace et document source. La technologie numérique change t-elle la notion de document? » *Information-Interaction-Intelligence*, 4(1), pp. 141-157.
- Chabin, M.-A. (2007). *Archiver, et après?* Paris: Djakarta. 159 p.
- Chabin, M.-A. et al. (2006). *Dématérialisation et archivage électronique: mise en oeuvre de l'ILM (Information Lifecycle Management)*. Paris: Dunod. 207 p.
- Chaimbault, T. (2008). *L'archivage du Web*. France: École nationale supérieure des sciences de l'information et des bibliothèques. Dossiers documentaires. 54 p. Adresse Url: <http://www.enssib.fr/bibliotheque-numerique/document-1730> (Page consultée le: 21 janvier 2009).
- Chapman, S. (2006). « Chapitre 7. It's money that matters in long-term preservation. » Dans: Deegan, M. et Tanner, S., eds. *Digital preservation*. London: Facet, pp. 133-146.
- Charlesworth, A. (2003). *Legal issues relating to the archiving of Internet resources in the UK, EU, USA and Australia: a study undertaken for the JISC and Wellcome Trust: Version 1.0*. Centre for IT and Law, University of Bristol. 79 p. Adresse Url: http://www.jisc.ac.uk/uploaded_documents/archiving_legal.pdf (Page consultée le: 14 octobre 2006).
- Chebbi, A. (2007-2008). « Archivage du Web: quelques leçons à tirer. » *Archives*, 39(2), pp. 19-46.
- Chebbi, A. (2009). « Nouvelles alliances, nouvelles façons de faire pour préserver le patrimoine numérique en ligne. » Dans: *Actes du premier congrès de l'Association Internationale Francophone des Bibliothécaires et Documentalistes (AIFBD) et satellite IFLA*, 03-06 août 2008, pp. 427-437.
- Chebbi, A. et Cherif, A. (2010). *Banque d'outils pour la gestion des archives*. Montréal: École de bibliothéconomie et des sciences de l'information. Adresse Url: <http://www.delicious.com/RessourcesArchiviques> (Page consultée le: 27 mars 2012).
- Chen, K.-h. et al. (2008). « Developing national Taiwan university web archiving system. » Dans: *IWAW08, 8th international web archiving workshop*, Aarhus, Danemark, 18-19 septembre. Adresse Url: <http://webarchive.lib.ntu.edu.tw/>

iwaw_paper.pdf (Page consultée le: 22 janvier 2009).

- Christensen, N. H. (2005). « Preserving the bits of the Danish Internet. » Dans: *IWAW05, 5th international web archiving workshop in conjunction with the 8th european conference on research and advanced technologies for digital libraries*, Vienna, Austria, 22-23 septembre, pp. 15. Adresse Url: <http://netarkivet.dk/wp-content/uploads/iwaw05-christensen.pdf> (Page consultée le: 22 janvier 2009).
- Cleroult, R. (2006). *L'archivage des sites Web au gouvernement du Québec: éléments de problématique v3.3*. [En ligne]. Adresse Url: <http://www.webmaestro.gouv.qc.ca/publications/08-soutien/WebEducation2006/WE2006-10/6-WE2006-10-ArchivageA.doc> (Page consultée le: 27 avril 2007).
- Collin, P. (2002). « Définitions. » Dans: Collin, S.M.H., éd. *Dictionary of information technology*. pp. 442.
- Commissariat EASI-WAL ([s.d.]). *L'archivage des sites web*. [En ligne]. Adresse Url: http://easi.wallonie.be/easi/col_gauche_niveaux_fr/easi-wal/dossiers-thematiques/gestion-du-web/archivage/index.html?LANG=fr (Page consultée le: 26 février 2011).
- CommonCrawl. ([s.d.]). [En ligne]. Adresse Url: <http://commoncrawl.org/> (Page consultée le: 12 juillet 2012).
- Coufal, L. et Žabička, P. (2007). « Strategies and approaches to building thematic collections in WebArchiv. » Dans: *INFuture2007, 1st international conference: the future of information sciences 'Digital information and heritage'*, 07-09 novembre, Sheraton Zagreb Hotel / Faculty of Humanities and Social Sciences, Zagreb, pp. 6. Adresse Url: <http://en.webarchiv.cz/files/dokumenty/konference/InFuture07.pdf> (Page consultée le: 22 janvier 2009).
- Couture, C. (1996). « Le concept de document d'archives à l'aube du troisième millénaire » *Archives*, 27(4), pp. 3-19.
- Couture, C. (1999). *Les fonctions de l'archivistique contemporaine*. Sainte-Foy: Presses de l'Université du Québec. 559 p.
- CREPUQ (Conférence des recteurs et des principaux des universités du Québec). Sous-comité des archivistes. (2002). *Recueil des règles de conservation des documents des établissements universitaires québécois*. [En ligne]. Adresse Url: <http://www.crepucq.ca/documents/arch/recueil/pt.htm> (Page consultée le: 26 octobre 2012).
- CREPUQ (Conférence des recteurs et des principaux des universités du Québec). Sous-comité des archivistes. (2009). *Mesures transitoires et bonnes pratiques de gestion des documents numériques: Proposées et adoptées par le Sous-comité des archivistes*. 32 p. Adresse Url: http://www.crepucq.ca/IMG/pdf/mesures_transitoires_bonnes_pratiques_GDN-10juin.pdf

- CRL - Center for Research Libraries. (2002-2005). *Political communication web archiving project*. [En ligne]. Adresse Url: <http://www.crl.edu/Content/PolitWeb.htm> (Page consultée le: 27 février 2010).
- Crowston, K. et Williams, M. (1999). « The effects of linking on genres of web documents. » Dans: *Proceedings of the 32nd Hawaii international conference on system sciences (HICSS '99)*, Maui, Hawaii, 5-8 janvier. IEEE, pp. 2006. Adresse Url: <http://csdl2.computer.org/comp/proceedings/hicss/1999/0001/02/00012006.PDF> (Page consultée le: 07 janvier 2006).
- Cruse, P. et al. (2003). *Web-based government information: evaluating solutions for capture, curation, and preservation. An Andrew W. Mellon funded initiative of the California Digital Library*. California Digital Library. 79 p. Adresse Url: http://www.cdlib.org/programs/Web-based_archiving_mellon_Final.pdf (Page consultée le: 23 avril 2007).
- Cunningham, A. et Phillips, M. (2005). « Accountability and accessibility: ensuring the evidence of e-governance in Australia. » *Aslib Proceedings New Information Perspectives*, **57**(4), pp. 301-317.
- Czech National Library. (2009). *WebArchiv: overview of the WebArchiv project* [En ligne]. Adresse Url: <http://en.webarchiv.cz/information/> (Page consultée le: 22 janvier 2009).
- Daffini, C. (2010). *Dépôt légal numérique: l'archivage des blogues adolescents*. Mémoire d'étude. Lyon: École nationale supérieure des sciences de l'information et des bibliothèques. 106 p.
- Day, M. (2003a). *Collecting and preserving the World Wide Web: a feasibility study undertaken for the JISC and Wellcome Trust. Version 1.0*. UKOLN, University of Bath. 91 p. Adresse Url: http://www.jisc.ac.uk/uploaded_documents/archiving_feasibility.pdf (Page consultée le: 15 octobre 2006).
- Day, M. (2003b). *Preserving the fabric of our lives: a survey of web preservation initiatives*. Bath, Royaume-Uni: UKOLN, University of Bath. 12 p. Adresse Url: <http://opus.bath.ac.uk/23597/> (Page consultée le: 15 octobre 2006).
- Deegan, M. et Tanner, S. (2002). « Chapitre 1. Digital futures in current contexts. » Dans: *Digital futures : strategies for the information age*. New York; London: Neal-Schuman Publishers; Library Association, pp. 1-29.
- Denzin, N. K. et Lincoln, Y. S. (2003). *The landscape of qualitative research: theories and issues*. 2e éd. Thousand Oaks, Calif.: Sage. 684 p.
- Deslauriers, J.-P. et Mayer, R. (2000). « Chapitre 6. L'observation directe. » Dans: Mayer, R. et al., éd. *Méthodes de recherche en intervention sociale*. Québec: Morin, pp. 135-158.
- Detlor, B. (2000). *Facilitating organizational knowledge work through web information systems: an investigation of the information ecology and information behaviours of users in a telecommunications company*. Thèse de doctorat. Toronto: University of Toronto. 320 p.

- Dillon, A. et Gushrowski, B. A. (2000). « Genres and the web: is the personal home page the first uniquely digital genre? » *Journal of the American Society for Information Science*, **51**(2), pp. 502-505.
- Dmitriev, P. et al. (2005). « Finding the boundaries of information resources on the web. » Dans: *International world wide web conference*, Chiba, Japan, mai 10-14. ACM Press, pp. 1124 - 1125. Adresse Url: <http://www2005.org/cdrom/docs/p1124.pdf> (Page consultée le: 07 janvier 2006).
- DNB - Deutsche National Bibliothek. (2006). *Archive server DEPOSIT.D-NB.DE*. [En ligne]. Adresse Url: http://deposit.ddb.de/index_e.htm (Page consultée le: 21 janvier 2009).
- Documentation centre for Dutch political parties. (2000). *The project plan*. [En ligne]. Adresse Url: <http://www.archipol.nl/english/project/projectplan.html> (Page consultée le: 21 janvier 2009).
- Ducharme, D. (2007). Théorie des trois âges ou Records continuum? Réflexion sur le cycle de vie des documents d'archives. 36e congrès de l'AAQ: d'hier à demain 40 ans de théories et de pratiques archivistiques. Quelles pistes pour l'avenir? Sainte-Adèle, Québec, 31 mai-2 juin.
- Ducharme, D. (2000-2001). « Identification des critères d'évaluation pour les archives informatiques: enquête auprès d'archivistes québécois. » *Archives*, **32**(2): 17-32.
- Dufour, C. (2003). *Étude du rôle des professionnels de l'information dans les systèmes d'information Web du gouvernement fédéral canadien*. Thèse de doctorat. Montréal: Université de Montréal. 191 p.
- Dufour, C. (2008-2009). « Web 2.0, organisations et archivistique. » *Archives*, **40**(2), pp. 3-26.
- Dumont, M. (2005). *Internet, entre archivage et mémoire*. EspacesTemps.net. Adresse Url: <http://espacestemps.net/document1725.html> Page consultée le: 15 mai 2010).
- Duranti, L. (2001). « Concepts, principles, and methods for the management of electronic records. » *The Information Society*, **17**, pp. 271-279.
- Duranti, L. (2003). « Pour une diplomatie des documents électroniques. » *Bibliothèque de l'école des chartes*, **161**(161-2), pp. 603-623
- Duranti, L. (2004). « La conservation à long terme des documents dynamiques et interactifs: InterPARES 2. » *Document Numérique*, **8**(1), pp. 73-86.
- Duranti, L. et al. (2002). *Preservation of the integrity of electronic records*. Dordrecht ; Boston: Kluwer Academic. 172 p. (The Archivist's library).
- Eisenhardt, K. M. (1989). « Building theories from case study research. » *Academy of Management Review*, **14**(4), pp. 532-550.
- European Archive. (2010). *International Web Archiving Workshop*. [En ligne]. Adresse Url: <http://www.iwaw.net/> (Page consultée le: 27 février 2010).

- Farrell, S. et al. (2010). *A guide to web preservation: practical advice for web and records managers based on best practices from the JISC-funded PoWR project*. University of London Computer Centre et UKOLN, University of Bath. 60 p. Adresse Url: <http://jiscpowr.jiscinvolve.org/wp/files/2010/06/Guide-2010-final.pdf> (Page consultée le: 15 mai 2010).
- FCCN - National Foundation for Scientific Computing. (2000-2009). *Portuguese Web Archive*. [En ligne]. Adresse Url: http://arquivo-web.fccn.pt/portuguese-web-archive-2?set_language=en (Page consultée le: 22 janvier 2009).
- FÉEP - Fédération des établissements d'enseignement privés et BANQ - Bibliothèque et Archives nationales du Québec. (2006). *Guide de gestion des archives à l'intention des établissements d'enseignement privés du Québec*. 140 p. Adresse Url: http://www.banq.qc.ca/documents/services/archivistique_ged/GuideGestionEtablissementsEnsPrives_2006.pdf (Page consultée le: 27 septembre 2012).
- Fitch, K. (2003). « Web site archiving - an approach to recording every materially different response produced by a website. » Dans: *AusWeb03, the thirteenth australasian Word Wide Web conference*, 05-09 juillet, Queensland, Australia. Adresse Url: <http://ausweb.scu.edu.au/aw03/papers/fitch/paper.html> (Page consultée le: 07 septembre 2007).
- Foo, S. et Lim, E. P. (1997). « Managing world wide web publications. » *Asian Libraries*, 6(3/4), pp. 166-177.
- Forde, H. (2007). *Preserving archives*. London: Facet. 320 p. (Principles and practice in records management and archives).
- Fortin, M.-F. (1996). *Le processus de recherche: de la conception à la réalisation*. Québec, Ville Mont-Royal: Décarie Éditeur. 379 p.
- Frost, J. P. (2001). « Web technologies for information management. » *The Information Management Journal*(octobre), pp. 34-37.
- Gagnon, L. (1998). *Typologie des documents des organisations: de la création à la conservation*. Sainte-Foy: Presses de l'Université du Québec. xvi, 432 p.
- Gagnon, L. (2000). « Les archives des organisations: de la gestion à la mémoire. » *Document Numérique*, 4(3-4), pp. 265-275.
- Gagnon, L. (2002). « Pour une typologie des archives électroniques. Les archives électroniques: une mémoire orpheline ou en mutation? » Dans: *Les archives électroniques: une mémoire orpheline ou en mutation? Actes du 4ème symposium du GIRA*, pp. 21-31. Adresse Url: http://www.gira-archives.org/documents/GIRA_2002.pdf
- Gagnon, Y.-C. (2005). *L'étude de cas comme méthode de recherche: guide de réalisation*. Sainte-Foy: Presses de l'Université du Québec. 128 p.
- Garon, R. (2003). « Archives, bibliothèques et cybergouvernement. » *Archives*, 34(3), pp. 89-105.

- Gatenby, P. (2004). « Collecter et gérer les ressources Internet pour un accès pérenne: moissonnage et directives pour l'aide à la conservation (Actions de l'ICABS 3.3 et 3.4). » Dans: *World library and information congress: 70th IFLA general conference and council*, Buenos Aires, Argentina, 22-27 août, pp. 6. Adresse Url: http://archive.ifla.org/IV/ifla70/papers/026f_trans-Gatenby.pdf (Page consultée le: 17 octobre 2005).
- Gharsallah, M. (2002). *Archivage du Web français: Dépôt légal des publications électroniques et préservation patrimoniale*. Thèse de doctorat. Paris: Université Paris 8. 199 p.
- Gilliland-Swetland, A. J. (2000). *Enduring paradigm, new opportunities: the value of the archival perspective in the digital environment*. Washington: Council on Library and Information Resources. 37 p. Adresse Url: <http://www.clir.org/pubs/reports/pub89/contents.html>
- Gladney, H. M. (2007). *Preserving digital information*. Berlin ; New York: Springer. 314 p.
- Gomes, D. et al. (2006). « Design and selection criteria for a national web archive. » Dans: *ECDL 2006, 10th european conference*, Alicante, Espagne, 17-22 septembre, pp. 196-207. Adresse Url: <http://dl.acm.org/citation.cfm?id=2111175> (Page consultée le: 27 février 2010).
- Gomes, D. et al. (2011). « A survey on web archiving initiatives. » Dans: *TPDL'11, proceedings of the 15th international conference on theory and practice of digital libraries: research and advanced technology for digital libraries*, Berlin, Heidelberg, 26-28 septembre, pp. 408-420. Adresse Url: <http://dl.acm.org/citation.cfm?id=2042590> (Page consultée le: 22 janvier 2009).
- Gomes, D. et al. (2008). « Introducing the Portuguese web archive initiative. » Dans: *IWAW08, 8th international web archiving workshop*, Aarhus, Danemark, 18-19 septembre, pp. 9. Adresse Url: <http://iwaw.net/08/IWAW2008-Gomes.pdf> (Page consultée le: 21 janvier 2009).
- Gouvernement du Québec. (2011). *Politique gouvernementale relative à l'emploi et à la qualité de la langue française dans les administrations*. 12 p. Adresse Url: http://www.spl.gouv.qc.ca/fileadmin/medias/pdf/Politique_gouvernementale.pdf (Page consultée le: 26 octobre 2012).
- Government of Western Australia. (2002). *Guidelines for the management of web information as per The State Records Act, 2000: Addendum to Section 5 of the guidelines for state government web sites, version 1.0*. West Perth: The Department of Industry and Technology. 74 p. Adresse Url: <http://www.indetech.wa.gov.au/govt/polguides/websites/webinfo.htm> (Page consultée le: 17 octobre 2005).
- Gregory, L. A. (2009). *The practice and perception of web archiving in academic libraries and archives*. Master's Paper for the M.S. in L.S degree. Chapel Hill, North Carolina: School of Information and Library Science of the University

of North Carolina. 58 p.

- Grimard, J. (2005-2006). « L'archivistique à l'heure du paradigme de l'information ou la Révolution numérique à l'âge archivistique. » *Archives*, **37**(1), pp. 59-87.
- Groupe d'études sur l'usage des TIC dans les migrations. (2005). « Archiver le Web sur les migrations: quelles approches techniques et scientifiques pour créer un observatoire du champ migratoire on-line. » *Migrance* (23), pp. 72-93.
- Groupe de coordination 46. (25 juin 2009). *Publication de la norme ISO 28500: format de fichier pour l'archivage de l'Internet (WARC)*. [En ligne]. Adresse Url: <http://gc46.wordpress.com/2009/06/25/publication-de-la-norme-iso-28500-format-de-fichier-pour-l%E2%80%99archivage-de-l%E2%80%99internet-warc/> (Page consultée le: 27 février 2010).
- Haettiger, M. (2003a). *L'archivage des sites Web d'intérêt régional*. Mémoire d'étude. Lyon: École nationale supérieure des sciences de l'information et des bibliothèques. 122 p.
- Haettiger, M. (2003b). « Vers la conservation des sites web régionaux. » *BBF*, **T48**(4), pp. 77-84.
- Hallgrímsson et Bang, S. (2003). « Nordic Web Archive. » Dans: *3rd ECDL workshop on web archives*, Trondheim, Norvège, 21 août. Adresse Url: <http://bibnum.bnf.fr/ecdl/2003> (Page consultée le: 22 janvier 2009).
- Hare, C. et McLeod, J. (2003). *Guide pratique: mettre en place le records management dans son organisation (traduction française pour archimag)*. Paris: IDP Sarl
- Harries, S. (1999). « Capturing and managing electronic records from websites and Intranets in the government environment. » Dans: *DLM- Forum-European citizens and electronic information: the memory of the Information society*, Bruxelles, Belgique, 18-20 octobre. Adresse Url: http://europa.eu.int/ISPO/dlm/fulltext/full_harr_en.htm (Page consultée le: 15 janvier 2005).
- Hartley, J. (2004). « Case study research. » Dans: Cassell, C. et Symon, G., édés. *Essential guide to qualitative methods in organizational research*. London: SAGE Publications, pp. 323-333.
- Harvard University Library. (2008). *Web archiving resources*. [En ligne]. Adresse Url: <http://hul.harvard.edu/ois/systems/wax/resources.html> (Page consultée le: 21 janvier 2009, 26 février 2011).
- HEC Montréal. (2010). *Politique relative à la gestion des documents actifs, semi-actifs et inactifs*. 21 p. Adresse Url: http://www.hec.ca/direction_services/secretariat_general/juridique/reglements_politiques/documents/politique_gestion_documents_actifs_semi-actifs.pdf (Page consultée le: 11 décembre 2012).
- Hedstrom, M. (2002). « Archives, memory, and interfaces with the past. » *Archival Science*, **2**(1-2), pp. 21-43.
- Heinze, N. et Hu, Q. (2006). « The evolution of corporate web presence: a

- longitudinal study of large American companies. » *International Journal of Information Management*, **26**(4), pp. 313-325.
- Henry, L. (2000). « Part eight: electronic records - 25. Schellenberg in cyberspace. » Dans: Jimerson, R.C., éd. *American archival studies: readings in theory and practices*. Chicago: The Society of American Archivists, pp. 569-587.
- Hofman, H. (1996). « Lost in cyberspace- where is the record? » Dans: *The concept of record : report from the second Stockholm conference on archival Science and the concept of record*, Stockholm: Swedish National Archives, 30-31 mai.
- Holck, J. (2003). « 4 perspectives on web information systems. » Dans: *Proceedings of the 36th Hawaii international conference on system sciences (HICSS'03)*, 6-9 Janvier, Big Island, Hawaii, 6-9 janvier, pp. 8. Adresse Url: <http://csdl2.computer.org/comp/proceedings/hicss/2003/1874/08/187480265b.pdf> (Page consultée le: 15 octobre 2006).
- Huang, L. et al. (2008). « Building a story tracer out of a web archive. » Dans: *Joint conference on digital libraries*, Pittsburgh, Pennsylvania, USA, 16-20 juin, pp. 455. Adresse Url: <http://dl.acm.org/citation.cfm?id=1379000> (Page consultée le: 15 octobre 2006).
- Huang, Z. Y. et al. (2002). « Integrating web-based data into a data warehouse. » *Information Systems Management*, **19**(1), pp. 23-34.
- ICA - International Council of Archives. (2012). *Keynote speakers*. [En ligne]. Adresse Url: <http://www.ica2012.com/pages/keynote-speakers.php> (Page consultée le: 28 août 2012).
- IIPC - International Internet Preservation Consortium. (2004-2008). *Web archiving: what, why and how*. [En ligne]. Adresse Url: <http://www.netpreserve.org/about/index.php> (Page consultée le: 03 mars 2008).
- IIPC - International Internet Preservation Consortium. (2012a). *Member archives*. [En ligne]. Adresse Url: <http://www.netpreserve.org/resources/member-archives> (Page consultée le: 26 février 2011).
- IIPC - International Internet Preservation Consortium. (2012b). *Web archiving: tools and software*. [En ligne]. Adresse Url: <http://netpreserve.org/web-archiving/tools-and-software> (Page consultée le: 26 février 2011).
- Illien, G. (2009). « 18. Patrimoine, culture et Internet. Les mémoires de la Toile: enjeux de la coopération internationale pour l'archivage de l'internet, le consortium IIPC, une chance pour les bibliothèques francophones ? » Dans: *Actes du premier congrès de l'Association Internationale Francophone des Bibliothécaires et Documentalistes (AIFBD) et satellite IFLA*, 03-06 août 2008, Montréal, pp. 439-446.
- Illien, G. et Game, V. (2006). « Le dépôt légal d'Internet à la Bibliothèque nationale de France : Cadre juridique, modèle de collecte, évolutions des métiers. » *BBF*(3), pp. 82-85.

- Illien, G. et Youry, C. (2009). « Quelle politique documentaire pour l'archivage des sites Web. » Dans: Carbone, P. et Cavalier, F., édés. *Les collections électroniques, une politique documentaire en mouvement*. Paris: Éditions du Cercle de la librairie, pp. 157-178.
- IM Forum Internet and Intranet Working Group. (1999). *Approche à la gestion de l'information sur Internet et Intranet pour assurer l'accès à long terme et la reddition de comptes*. Ottawa: Office of Government Records. National Archives of Canada. 75 p. Adresse Url: <http://web.archive.org/web/20091025070625/http://www.imforumgi.gc.ca/archives/groups-groupes/iiwg-gtrii/implement/implement-eng.pdf> (Page consultée le: 23 novembre 2009).
- INA - Institut national de l'audiovisuel. (2003). *Inter-Ateliers: L'archivage du web*. 45 p.
- Institute of Chinese Studies. University of Heidelberg. (2008). *About DACHS*. [En ligne]. Adresse Url: <http://www.sino.uni-heidelberg.de/dachs/intro.htm> (Page consultée le: 22 janvier 2009).
- Internet Archive. (2012a). *Internet Archive Blogs*. [En ligne]. Adresse Url: <http://blog.archive.org/2012/10/26/80-terabytes-of-archived-web-crawl-data-available-for-research/> (Page consultée le: 26 octobre 2012).
- Internet Archive. (2012b). *WayBack machine*. [En ligne]. Adresse Url: <http://archive.org/index.php> (Page consultée le: 11 février 2012).
- Internet Archive et Archive-It. (2011-2012). *Jasmine Revolution - Tunisia 2011* [En ligne]. Adresse Url: <http://www.archive-it.org/collections/2323> (Page consultée le: 27 octobre 2012).
- Internet Memory foundation. (2010). *Web archiving in Europe*. 13 p. Adresse Url: http://internetmemory.org/images/uploads/Web_Archiving_Survey.pdf (Page consultée le: 05 mars 2011).
- iPRES2011. *International Conference on Preservation of Digital Objects, 1-4 novembre 2011, Singapore*. [En ligne]. Adresse Url: <http://ipres2011.sg/> (Page consultée le: 30 mai 2012).
- Isakowitz, T. et al. (1998). « Web information systems. » *Communications of the ACM*, **41**(7), pp. 78-80.
- ISO - Organisation internationale de normalisation. (2001). *ISO 15489-1, Information and documentation: "records management". Partie 1: principes directeurs*. Geneva: ISO. 19 p. (Norme internationale).
- ISO - Organisation internationale de normalisation. (2001b). *ISO 15489-1, Information and documentation: "records management". Partie 1: guide pratique*. Geneva: ISO. 39 p. (Norme internationale).
- Jaccoud, M. et Mayer, R. (1997). « L'observation en situation et la recherche qualitative. » Dans: Poupart, J. et al., édés. *La recherche qualitative: enjeux épistémologiques et méthodologiques*. Québec: Morin, pp. 211-250.

- Jacobsen, G. (2007). « Archivage d'internet au niveau international: l'interopérabilité de l'avenir? Traduction de Rosa María Gómez de Regil » Dans: *World library and information congress: 73th IFLA general conference and council*, Durban, South Africa, 19-23 août, pp. 9.
- JISC-Joint Information Systems Committee. *JISC PoWR: preservation of web resources, a JISC-funded project, [Archived blog]*. [En ligne]. Adresse Url: <http://jiscpowr.jiscinvolve.org/wp/> (Page consultée le: 12 février 2012).
- Jodelis, R. (2003). « Harvesting and archiving electronic resources in Lithuania: towards virtual library. » Dans: *Inforum 2003, 9th conference of professional information resources*, 27-29 mai. Adresse Url: http://www.inforum.cz/archiv/inforum2003/prispevky/Jodelis_Remigijus.pdf (Page consultée le: 21 janvier 2009).
- Karjalainen, A. et al. (2000). « Genre-based metadata for enterprise document management. » Dans: *Proceedings of the 33rd Hawaii international conference on system sciences (HICSS)*, 2000, Maui, Hawaii, 4-7 janvier. Adresse Url: <http://csdl.computer.org/comp/proceedings/hicss/2000/0493/03/04933013.pdf> (Page consultée le: 07 janvier 2006).
- Kavcic-Colic, A. (2006). *Selection criteria for archiving the slovenian web*. 3 p. Adresse Url: www.webarchiv.cz/files/culture2000/Slovenianweb.doc (Page consultée le: 07 janvier 2006).
- Kavcic-Colic, A. et Grobelnik, M. (2004). « Archiving the slovenian web: recent experiences. » Dans: *IWAW04, 4th international web archiving workshop*, Bath, UK, 16 septembre, pp. 10. Adresse Url: <http://iwaw.europarchive.org/04/Kavcic.pdf> (Page consultée le: 07 janvier 2006).
- Kenney, A. R. et al. (2002). « Preservation risk management for web resources: Virtual Remote Control in Cornell's project prism » *D-Lib Magazine*, **8**(1).
- Khouaja, B. et Couture, C. (2003-2004). « La gestion et l'archivage des sites Web institutionnels. » *Archives*, **35**(3/4), pp. 17-41.
- Koehler, W. (1999). « An analysis of web page and web site constancy and permanence » *Journal of the American Society for Information Science*, **50**(2), pp. 162 - 180
- Koehler, W. (2004). « A longitudinal study of Web pages continued: a consideration of document persistence. » *Information Research*, **9**(2).
- Koerbin, P. (2004). « Managing web archiving in Australia: a case study. » Dans: *IWAW04, 4th International Web archiving workshop in conjunction with the ECDL2004*, 16 septembre, Bath, UK, 16 septembre, pp. 24. Adresse Url: <http://iwaw.europarchive.org/04/Koerbin.pdf> (Page consultée le: 15 octobre 2006).
- Koerbin, P. (2005). *Report on the crawl and harvest of the whole australian web domain undertaken during june and july 2005*. Australie: Digital Archiving

- Branch, National Library of Australia. 22 p. Adresse Url: http://pandora.nla.gov.au/documents/domain_harvest_report_public.pdf (Page consultée le: 15 octobre 2006).
- L'Hostis, D. et al. (2006). *Archives ouvertes – Vers une obligation de dépôt? Synthèse sur les réalisations existantes, les pratiques des chercheurs et le rôle des institutions, Version 2*. France: INRA. 45 p. Adresse Url: <http://hal.archives-ouvertes.fr/docs/00/13/83/02/PDF/Note-AO-version2-051206-diff-Externe.pdf> (Page consultée le: 27 février 2010).
- Lafouge, T. et al. (2002). *Clustering du web en vue d'extraction du corpus homogène*. [En ligne]. Adresse Url: http://archivesic.ccsd.cnrs.fr/documents/archives0/00/00/03/88/sic_00000388_00/sic_00000388.pdf (Page consultée le: 15 octobre 2006).
- Lainé-Cruzet, S. (2005). « Documents, ressources, données: les avatars de l'information numérique. » *Revue 3I*, 4(1), pp. 105-119.
- Lamos, C. et al. (2004). « Archiving the Greek Web. » Dans: *IWAW04, 4th international Web archiving workshop in conjunction with the ECDL2004*, Bath, UK, 16 septembre, pp. 12. Adresse Url: <http://www.iwaw.net/04/> (Page consultée le: 22 janvier 2009).
- Laperrière, A. (2003). « L'observation directe. » Dans: Gauthier, B., éd. *Recherche sociale : de la problématique à la collecte des données*. Sainte-Foy: Presses de l'Université du Québec, pp. 269-292.
- Laudon, K. C. et al. (2001). *Les systèmes d'information de gestion : organisations et réseaux stratégiques*. Saint-Laurent (Québec): Éditions du renouveau pédagogique. 914 p.
- Lavoie, B. (2006). « Chapitre 6. The cost of digital preservation. » Dans: Deegan, M. et Tanner, S., éd. *Digital preservation*. London: Facet, pp. 106-132.
- Lawrence, G. W. et al. (2000). *Risk management of digital information: a file format investigation*. Washington, D.C: Council on Library and Information Resources. 81 p. Adresse Url: <http://milleniata.com/doc/resources/Risk%20Management%20of%20Digital%20Info%20copy.pdf> (Page consultée le: 26 octobre 2005).
- Lefurgy, W. G. (2001). « Records and archival management of World Wide Web sites. » *Government Record News, the newsletter of the Government Records Section of the Society of American Archivists*, 2(April).
- Léger, D. 2003. L'implantation du dépôt légal des publications diffusées sur Internet à la Bibliothèque nationale du Québec: Texte d'accompagnement de la présentation PowerPoint DLPI20030620.ppt. Bibliothèque nationale du Québec. Section du dépôt légal.
- Leleu-Merviel, S. (2004). « Effets de la numérisation et de la mise em réseau sur le concept de document. » *Revue I3*, 4(1), pp. 121-140.

- Lemaître, F. (2001). *La gestion des documents numériques dans les bibliothèques nationales*. Mémoire de maîtrise en sciences de l'information et documentation. Paris: Université Paris I-Panthéon Sorbone. 180 p.
- Lemay, Y. (2012a). *SCI6111 Politique de gestion des archives, automne 2012: module 07*. [En ligne]. Adresse Url: <http://cours.ebsi.umontreal.ca/planscours/diffusion/index.asp?sigle=sci6111> (Page consultée le: 10 décembre 2012).
- Lemay, Y. (2012b). *SCI6111 Politique de gestion des archives, automne 2012: module 03*. [En ligne]. Adresse Url: <http://cours.ebsi.umontreal.ca/planscours/diffusion/index.asp?sigle=sci6111> (Page consultée le: 10 décembre 2012).
- Lemieux, V. L. (2004). *Managing risks for records and information*. Lenexa, Kansas: ARMA International. 94 p.
- Leroux, É. et Lemay, Y. (2009). « Les professions et les institutions » Dans: Salaün, J.-M. et Arsenault, C., eds. *Introduction aux sciences de l'information*. Montréal: Les Presses de l'Université de Montréal, pp. 15-52.
- Lessard-Hébert, M. et al. (1995). *La recherche qualitative: fondements et pratiques*. 2e éd. Montréal: Éditions nouvelles. 124 p.
- Levy, D. M. (1994). « Fixed or fluid? Document stability and new media. » Dans: *Proceedings of the 1994 ACM european conference on hypermedia technology*, Edinburgh, Scotland, septembre 19-23, pp. 24-31. Adresse Url: <http://adamkikal.org/courses/ht/files/readings/p24-levy.pdf> (Page consultée le: 23 septembre 2005).
- Liao, C. et al. (2006). « Website practices: a comparison between the top 1000 companies in the US and Taiwan. » *International Journal of Information Management*, **26**(3), pp. 196-211.
- Library of Congress. (2011). *Minerva: Law library legal blawgs web archive*. [En ligne]. Adresse Url: <http://lcweb2.loc.gov/diglib/lcwa/html/lawlb/lawlb-overview.html> (Page consultée le: 27 octobre 2012).
- Liu, Z. (2004). « The evolution of documents and its impacts. » *Journal of Documentation*, **60**(3), pp. 279-288.
- Llueca, C. et al. (2010). *CAT (Curator Archiving Tool): improving access to web archives*. Vienna: International Internet Preservation Consortium meeting. Adresse Url: http://www.recercat.net/bitstream/2072/85525/1/Padicat_iipc_2010_ENG.pdf (Page consultée le: 26 octobre 2006).
- Locher, H. (2008). *Archives Web Suisse: une collection commune de sites Web patrimoniaux des bibliothèques cantonales suisses et de la Bibliothèque nationale suisse, version 1.1 (traduction française)*. Suisse: Office fédéral de la culture OFC, Bibliothèque nationale suisse. 21 p. Adresse Url: http://www.nb.admin.ch/slb/slb_professionnel/01693/01698/01987/index.html?lang=fr (Page consultée le: 22 janvier 2009).
- Lounamaa, K. et Salonharju, I. (1999). « EVA: the acquisition and archiving of

- electronic network publications in Finland. » *Tietolinja News*(1).
- Maftai, M. ([s.d.]). *Conditions requises pour évaluer et maintenir l'authenticité des documents d'archives électroniques*. International Research in Permanent Authentic Records in Electronic Systems (InterPARES), Groupe de travail sur l'authenticité. 13 p. Adresse Url: http://www.cdncouncilarchives.ca/ATF_Requirements_FR.pdf (Page consultée le: 22 mai 2008).
- Mannerheim, J. (2000). « Le World Wide Web et la préservation de notre patrimoine numérique - De nouvelles missions pour les bibliothèques (traduction). » Dans: *66th IFLA council and general conference*, Jerusalem, Israel, 13-18 août. Adresse Url: <http://ifla.queenslibrary.org/IV/ifla66/papers/158-157e.htm> (Page consultée le: 17 octobre 2006).
- Marill, J. et al. (2004). « Tools and techniques for harvesting the world wide web. » Dans: *Proceedings of the 2004 Joint ACM/IEEE conference on digital libraries (JCDL'04)*, Tucson, Arizona, USA, 7-11 juin, pp. 403. Adresse Url: <http://dl.acm.org/citation.cfm?id=996469> (Page consultée le: 15 octobre 2006).
- Masanès, J. (2000). *L'archivage des sites Internet: Rapport de stage*: École nationale supérieure des sciences de l'information et des bibliothèques. 24 p.
- Masanès, J. (2005). « Web archiving methods and approaches: a comparative study. » *Library Trends*, **54**(1), pp. 72-90.
- Masanès, J. (2006). « Chapitre 5. L'archivage du Web. » Dans: Calderan, L. et al., éd. *Pérenniser le document numérique*. Paris: ADBS éditions, pp. 101-122.
- Masanès, J. E. (2006). *Web archiving: issues and methods*. Springer Berlin Heidelberg. 234 p.
- McClure, C. R. (2000). « Federal information policies affecting public access to web-based federal information. » Dans: *SIST 2000, proceedings of the 63rd ASIST annual meeting* Chicago, IL, 13-16 novembre, pp. 224-245. Adresse Url: <http://www.asis.org/proceedings.html> (Page consultée le: 12 novembre 2005).
- McClure, C. R. et Sprehe, T. J. (1998). *Analysis and development of model quality guidelines for electronic records management on state and federal websites*. Washington: The national historical publications and records commission. 214 p. Adresse Url: http://web.archive.org/web/20080302132340/http://www.ii.fsu.edu/~cmclure/nhprc/nhprc_toc.html
- McCown, F. et Nelson, M. L. (2009). « What happens when facebook is gone? » Dans: *JCDL'09, proceedings of the 9th ACM/IEEE-CS joint conference on Digital libraries*, Austin Texas, USA, 15-19 juin. Adresse Url:
- McDonald, A. et al. (2005). *Les archives électroniques: manuel a l'usage des archivistes*. Paris: CIA - Conseil international des archives. Comité pour les archives courantes dans un environnement électronique,. ICA Études 16. 69 p. Adresse Url: <http://old.ica.org/biblio/ICA%20Study16%20FR.pdf> (Page consultée le: 17 octobre 2006).

- McDonald, J. (2005). *Vers une stratégie nationale sur l'information numérique: revue des initiatives internationales pertinentes*. Bibliothèque et Archives Canada. 22 p. Adresse Url: <http://www.collectionscanada.ca/scin/012033-602-f.html> (Page consultée le: 21 janvier 2009).
- McKemmish, S. (2001). « Placing records continuum theory and practice. » *Archival Science*, **1**, pp. 333-359.
- Memento. ([s.d.]). [En ligne]. Adresse Url: <http://www.mementoweb.org/>
- Metamatrix // Interfolio. ([s.d.]). *Deja Vu: The web as we remember it*. [En ligne]. Adresse Url: <http://www.dejavu.org/> (Page consultée le: 22 janvier 2009).
- Miles, M. B. et Huberman, A. M. (2003). *Analyse des données qualitatives*. 2e éd. Paris: De Boeck Université. 626 p. (Méthodes en sciences humaines).
- Milinic, M. et Topolscak, N. (2008). « The architecture of DAMP: a system for harvesting and archiving Web publications. » *WIDWISAWN, Who Is Doing What In Scotland And What's Needed*, **3**(3), pp. 1-3.
- Millar, I. (2001). « Nouvelles orientations pour la British Library. » *BBF*, **t.46**(6), pp. 86-92.
- Mukamurera, J. et al. (2006). « Des avancées en analyse qualitative : pour une transparence et une systématisation des pratiques. » *Recherches qualitatives*, **26**(1), pp. 110-138.
- NAA - National Archives of Australia. (2001a). *Archiving web resources: a policy for keeping records of web-based activity in the Commonwealth government*. Government recordkeeping standards and policy. 18 p. Adresse Url: http://www.naa.gov.au/Images/archweb_policy_tcm2-902.pdf (Page consultée le: 11 octobre 2005).
- NAA - National Archives of Australia. (2001b). *Archiving Web Resources: guidelines for keeping records of web-based activity in the Commonwealth of Australia*. Government Recordkeeping Standards and Policy. 41 p. Adresse Url: http://www.naa.gov.au/Images/archweb_guide_tcm16-47165.pdf (Page consultée le: 11 octobre 2005).
- NAA - National archives of Australia. ([s.d.]). *Archiving websites: advice and policy statement*. [En ligne]. Adresse Url: http://www.naa.gov.au/Images/Archiving%20websites%20advice%20and%20policy%20statement_tcm2-902.pdf (Page consultée le: 27 janvier 2009).
- Nahuet, R. (2009). « L'archivistique contemporaine à l'âge adulte: pertinence et actualité du respect des fonds. » *Archives*, **41**(1), pp. 45-60.
- NARA - National Archives and Records Administration. (2004). *Expanding acceptable transfer requirements: transfer instructions for permanent electronic records - Web content records*. [En ligne]. Adresse Url: <http://www.archives.gov/records-mgmt/initiatives/web-content-records.html> (Page consultée le: 26 mars 2011).

- NARA - National Archives and Records Administration. (2005). *NARA guidance on managing web records*. 28 p. Adresse Url: <http://www.archives.gov/records-mgmt/pdf/managing-web-records-index.pdf> (Page consultée le: 27 janvier 2009).
- NARA - National Archives and Records Administration. ([s.d.]a). *Implications of recent web technologies for NARA Web guidance*. [En ligne]. Adresse Url: <http://www.archives.gov/records-mgmt/initiatives/web-tech.html> (Page consultée le: 23 avril 2007).
- NARA - National Archives and Records Administration. ([s.d.]b). *Tips for scheduling potentially permanent web content records*. [En ligne]. Adresse Url: <http://www.archives.gov/records-mgmt/publications/web-tips.pdf> (Page consultée le: 23 avril 2007).
- National and University Library in Zagreb. (2005). *Legal deposit of web resources*. [En ligne]. Adresse Url: <http://www.nsk.hr/DigitalLibEN.aspx?id=755> (Page consultée le: 21 janvier 2009).
- National library board Singapore. (2009). *Web Archive Singapore: Frequently Asked Questions*. [En ligne]. Adresse Url: <http://was.nl.sg/faq.html> (Page consultée le: 21 janvier 2009).
- National Library of Latvia. ([s.d.]). *WWW Archive*. 4 p. Adresse Url: <http://webarhivs.lndb.lv/> (Page consultée le: 22 janvier 2009).
- National Library of New Zealand. (2008). *New Zealand web harvest FAQs*. [En ligne]. Adresse Url: <http://www.natlib.govt.nz/about-us/news/20-october-2008-web-harvest-faqs/?searchterm=webarchiving> (Page consultée le: 27 février 2010).
- National Library of Sweden. (2008). *Swedish Websites: Kulturarw3 – The Swedish Web Archive* [En ligne]. Adresse Url: <http://www.kb.se/english/find/internet/websites/> (Page consultée le: 22 janvier 2009).
- NDSA - National Digital Stewardship Alliance. (2012). *Web archiving survey report*. 23 p. Adresse Url: http://www.digitalpreservation.gov/ndsa/working_groups/documents/ndsa_web_archiving_survey_report_2012.pdf
- Nelson, M., L. et al. (2007). « Using the web infrastructure to preserve web pages. » *International Journal of Digital Libraries*, 6(4), pp. 327-349.
- netarchive.dk. (2008). *Netarchive.dk collects and preserves the Danish portion of the internet*. [En ligne]. Adresse Url: <http://netarchive.dk/index-en.php> (Page consultée le: 22 janvier 2009).
- New Zealand Government. ([s.d.]). *Web standards: Managing web information as records and archiving websites*. [En ligne]. Adresse Url: <http://www.webstandards.govt.nz/managing-web-information-as-records-and-archiving-websites-2/> (Page consultée le: 27 janvier 2009).
- Nicholls, C. et Williams, J.-P. (2002). « Identifying roadkill on the information

- superhighway: a Website appraisal case study. » *Archives and Manuscripts*, **30**(2), pp. 96-111.
- Nicoud, A. (12 octobre 2012). « Commission Bouchard-Talylor: le site internet fermé. » *La presse.ca*. Adresse Url : <http://www.lapresse.ca/actualites/quebec-canada/politique-quebecoise/201210/12/01-4582611-commission-bouchard-talylor-le-site-internet-ferme.php> (Page consultée le: 12 octobre 2012).
- Niu, J. (2012). « An overview of web archiving. » *D-Lib Magazine*, **18**(13/4).
- Nivard Bigourdan, J. (2005). « L'archivage de l'Internet chinois... à Heidelberg. » *Migrance*(23), pp. 94-99.
- NLA - National Library of Australia. (2008). *Online australian publications: a selection guidelines for archiving and preservation by the National Library of Australia*. [En ligne]. Adresse Url: <http://pandora.nla.gov.au/selectionguidelines.html> (Page consultée le: 27 janvier 2009).
- NLA - National Library of Australia et NAA - National Archives of Australia. (2002). *Keeping government publications online: a guide for commonwealth agencies*. 6 p. Adresse Url: <http://web.archive.org/web/20090709110933/http://www.nla.gov.au/guidelines/govpubs.pdf> (Page consultée le: 11 octobre 2005).
- NLA - National Library of Australia et partners. (2007). *PADI: Preserving Access to Digital Information*. [En ligne]. Adresse Url: <http://pandora.nla.gov.au/pan/10691/20110824-1153/www.nla.gov.au/padi/topics/92.html> (Page consultée le: 23 avril 2007).
- NLA - National Library of Australia et partners. ([s.d.]). *PANDORA. Australi'S Web Archive: blogs*. [En ligne]. Adresse Url: <http://pandora.nla.gov.au/subject/29>
- NLC - National Library of Canada. (1996). *Electronic Publications Pilot Project (EPPP): summary of the final report*. Adresse Url: http://epe.lac-bac.gc.ca/100/200/301/nlc-bnc/eppp_summary-e/ereport.htm (Page consultée le: 15 octobre 2006).
- NSW - State Records Authority of New South Wales. (2009a). *Guideline 20 - Keeping web records*. [En ligne]. Adresse Url: <http://www.records.nsw.gov.au/recordkeeping/government-recordkeeping-manual/guidance/guidelines/guideline-20> (Page consultée le: 27 janvier 2009).
- NSW - State Records Authority of New South Wales. (2009b). *Guideline 24 - Records management and web 2.0*. [En ligne]. Adresse Url: <http://www.records.nsw.gov.au/recordkeeping/government-recordkeeping-manual/guidance/guidelines/guideline-24> (Page consultée le: 27 janvier 2009).
- NSW - State Records Authority of New South Wales. ([s.d.]). *Guideline 22 - Managing digital records: 5.4 Managing web records (Guideline 22)*. [En ligne]. Adresse Url: <http://www.records.nsw.gov.au/recordkeeping/government-recordkeeping-manual/guidance/guidelines/guideline-22/5-4->

- managing-web-records (Page consultée le: 27 janvier 2009).
- Öberg, L.-M. et Borglund, E. (2006). « What are the characteristics of records? » *International Journal of Public Information Systems*, **1**, pp. 55-76.
- OQLF - Office québécois de la langue française. (2005). *Extranet*. [En ligne]. Adresse Url: http://gdt.oqlf.gouv.qc.ca/ficheOqlf.aspx?Id_Fiche=8394385 (Page consultée le: 6 mars 2008).
- OQLF - Office québécois de la langue française. (2007a). *Internet*. [En ligne]. Adresse Url: http://gdt.oqlf.gouv.qc.ca/ficheOqlf.aspx?Id_Fiche=8361867 (Page consultée le: 6 mars 2008).
- OQLF - Office québécois de la langue française. (2007b). *Intranet*. [En ligne]. Adresse Url: http://gdt.oqlf.gouv.qc.ca/ficheOqlf.aspx?Id_Fiche=2075213 (Page consultée le: 6 mars 2008).
- Orlikowski, W. et Yates, J. (1994). « Genre repertoire: the structuring of communicative practices in organizations. » *Administrative Sciences Quarterly*, **33**(4), pp. 541-574.
- Oystein, Ø. et Päiväranta, T. (2005). « Autopoietic cybergenres for e-democracy? Genre analysis of a web-based discussion board. » Dans: *Proceedings of the 38th Hawaii international conference on system sciences*, Big Island, Hawaii, 3-6 janvier. Adresse Url: <http://csdl2.computer.org/comp/proceedings/hicss/2005/2268/04/22680098c.pdf> (Page consultée le: 01 janvier 2006).
- PARBICA - Pacific Regional Branch of the International Council on Archives. *Guideline 3: model recordkeeping policy*. Recordkeeping for good governance toolkit. 22 p. Adresse Url: <http://www.parbica.org/Toolkit%20pages/Documents/G3%20Policy/Guideline%203ModelRP.pdf> (Page consultée le: 11 décembre 2012).
- Parent, R. et Boulet, N. (1999a). *Conserver les documents électroniques: comment et pourquoi?* : Conseil du trésor. Sous-secrétariat à l'inforoute gouvernementale et aux ressources informationnelles. Collection en ingénierie documentaire : 4. 31 p. Adresse Url: <http://www.services.gouv.qc.ca/fr/publications/enligne/administration/ingenierie/conservation.pdf> (Page consultée le: 23 septembre 2005).
- Parent, R. et Boulet, N. (1999b). *La protection du document électronique: aspects techniques et juridiques, rapport du groupe de travail sur l'infrastructure juridique du document avec signature numérique*. Conseil du trésor. Sous-secrétariat à l'inforoute gouvernementale et aux ressources informationnelles. Collection en ingénierie documentaire: 5. 23 p. Adresse Url: <http://www.services.gouv.qc.ca/fr/publications/enligne/administration/ingenierie/conservation.pdf> (Page consultée le: 23 septembre 2005).
- Patton, M. Q. (2002). *Qualitative research & evaluation methods*. 3e éd. Thousand Oaks, Calif.: Sage Publications. 598 p.
- Pédauque, R. T. (2003). *Document: forme, signe et medium, les re-formulations du*

- numérique. [En ligne]. Adresse Url: http://archivesic.ccsd.cnrs.fr/documents/archives0/00/00/05/11/sic_00000511_01/sic_00000511.pdf (Page consultée le: 23 septembre 2005).
- Pédauque, R. T. (2006). *Comprendre et maîtriser la redocumentarisation* Adresse Url: http://rtp-doc.enssib.fr/IMG/pdf/CRSH-Projet_30-01-2006.pdf (Page consultée le: 23 avril 2007).
- Pennock, M. et al. (2006). « Archiving the web: what can institutions learn from national and international web archiving initiatives. » Dans: *Institutional Web Management Workshop 2006* University of Bath, 15 juin. Adresse Url: <http://www.ukoln.ac.uk/web-focus/events/workshops/webmaster-2006/sessions/pennock/> (Page consultée le: 23 avril 2007).
- Phillips, J. T. (2003). « The challenge of web site records preservation. » *The Information Management Journal*(janvier/février), pp. 42-48.
- Pipon, B. et Pontier, M.-C. (2004). *Abrégé d'archivistique: principes et pratiques du métier d'archiviste*. Paris: Association des Archivistes Français. 275 p.
- Pires, A. P. (1997). « Échantillonnage et recherche qualitative: essai théorique et méthodologique. » Dans: Poupart, J. et al., éd. *La recherche qualitative: Enjeux épistémologiques et méthodologiques*. Québec: Morin, pp. 113-172.
- Poupart, J. (1997). « L'entretien de type qualitatif: considérations épistémologiques, théoriques et méthodologiques. » Dans: Poupart, J. et al., éd. *La recherche qualitative: Enjeux épistémologiques et méthodologiques*. Québec: Morin, pp. 173-210.
- Premier ministre (France). ([s.d.]). *Le site www.archives.premier-ministre.gouv.fr*. [En ligne]. Adresse Url: http://www.archives.premier-ministre.gouv.fr/home_ie.htm (Page consultée le: 23 avril 2007).
- Public Record Office. (2001). *Managing web resources: management of electronic records on websites and intranets: an ERM toolkit. Version 1.0*. 38 p. Adresse Url: http://www.nationalarchives.gov.uk/electronicrecords/advice/pdf/website_toolkit.pdf (Page consultée le: 23 septembre 2005).
- Public Record Office Victoria. (2007). *Advice 20a: web-generated records, advice to victorian agencies PROA 07/20*. : Public Record Office Victoria 22 p. Adresse Url: http://www.prov.vic.gov.au/records/Web_Advice/PROVRMAAdvice20a.pdf (Page consultée le: 27 janvier 2009).
- Public Record Office Victoria. (2008). *Advice 20b: technical issues for capturing web-generated records. Advice to victorian agencies PROA 07/20*. [En ligne]. Adresse Url: http://www.prov.vic.gov.au/records/Web_Advice/PROVRMAdvice20b.pdf (Page consultée le: 27 janvier 2009).
- Public records office. (2001). *Managing web resources: management of electronic records on websites and intranets: an ERM toolkit*. Surrey. 38 p. Adresse Url: http://wayback.archive.org/web/*/http://www.nationalarchives.gov.uk/docum

- ents/website_toolkit.pdf (Page consultée le: 11 octobre 2007).
- Pymm, B. et Wallis, J. (2009). « Archiving the web: does whole-of-domain archiving information overload? » Dans: *Online information conference: ALIA 14th exhibition & conference*, 20-22 janvier, Sydney, pp. 11. Adresse Url: http://www.information-online.com.au/sb_clients/iog/data/content_item_files/000001/PresentationB8.pdf (Page consultée le: 22 janvier 2009).
- Québec (Province). (1983). *Loi sur les archives. L.R.Q.* [en ligne] http://www2.publicationsduquebec.gouv.qc.ca/dynamicSearch/telecharge.php?type=2&file=/A_21_1/A21_1.html.
- Québec (Province). (2001). *Loi concernant le cadre juridique des technologies de l'information. L.R.Q.* [en ligne] http://www2.publicationsduquebec.gouv.qc.ca/dynamicSearch/telecharge.php?type=2&file=/C_1_1/C1_1.html.
- Queensland Government. Department of Public Works. (2006). *Best practice guide to recordkeeping*. 66 p. Adresse Url: <http://www.archives.qld.gov.au/downloads/IS40BestPracticeGuidetoRKv1.03.00.pdf> (Page consultée le: 23 avril 2007).
- Queensland State Archives. (2006a). *Managing records of online resources and services guideline*. 39 p. Adresse Url: <http://www.archives.qld.gov.au/downloads/Policies/OnlineResourcesGuideline.pdf> (Page consultée le: 23 avril 2007).
- Queensland State Archives. (2006b). *Managing records of online resources and services policy*. 13 p. Adresse Url: <http://www.archives.qld.gov.au/downloads/Policies/OnlineResourcesPolicy.pdf> (Page consultée le: 23 avril 2007).
- Queensland State Archives. (2006c). *Web content management systems and recordkeeping*. [En ligne]. Adresse Url: <http://www.archives.qld.gov.au/Recordkeeping/GRKDownloads/Documents/WCMSRecords.pdf> (Page consultée le: 23 avril 2007).
- Queensland State Archives. (2008). *Identifying a public record in the electronic environment*. [En ligne]. Adresse Url: http://www.archives.qld.gov.au/Recordkeeping/GRKDownloads/Documents/identifying_pub_record_e_envir on_200503.pdf (Page consultée le: 27 janvier 2009).
- Racine, B. (2007). « La conservation des sites Web: le défi à relever. » Dans: *36^e congrès de l'AAQ: D'hier à demain, 40 ans de théories et de pratiques archivistiques. Quelles pistes pour l'avenir?*, 31 mai-02 juin, Sainte Adèle, Québec.
- Ras, M. et Sierman, B. (2006). « Long-term preservation and access of the Dutch Web. » Dans: *IWAW06, 6th international web archiving workshop in conjunction with the 10th european conference on research and advanced technologies for digital libraries*, Alicante, Espagne, 21-22 septembre, pp. 10. Adresse Url: http://www.kb.nl/hrd/dd/dd_projecten/webarchivering/

- documenten/IWAW2006_Ras.pdf (Page consultée le: 22 janvier 2009).
- Région wallonne. (2004). *Les archives des sites web de la Région wallonne*. [En ligne]. Adresse Url: <http://www.wallonie.be/fr/informations-generales/documentation/archives-web/index.html> (Page consultée le: 15 octobre 2005).
- Reilly, B. F. (2002). *Political communications web archiving: a proposal to the Andrew W. Mellon Foundation from the Center for Research Libraries*. 11 p. Adresse Url: <http://www.library.cornell.edu/iris/research/WebPolCom.pdf> (Page consultée le: 15 octobre 2005).
- Reilly, B. F. et al. (2003). « Political communication web archiving: addressing typology and timing for selection, preservation and access. » Dans: *3rd ECDL workshop on web archives in conjunction with the 7th european conference on research and advanced technologies for digital libraries*, Trondheim, Norvège, 21 août, pp. 9. Adresse Url: <http://bibnum.bnf.fr/ecdl/2003> (Page consultée le: 23 avril 2007).
- Rein, G. L. et al. (1997). « A case for document management functions on the web. » *Communications of the ACM*, **40**(9), pp. 81-89.
- Reix, R. (2004). *Systèmes d'information et management des organisations*. 5e éd. Paris: Vuibert. 486 p.
- Rispal Hlady, M. (2002). *La méthode des cas: application à la recherche en gestion*. Bruxelles: De Boeck Université. 250 p.
- Rivard, S. et Talbot, J. (2001). *Le développement de systèmes d'information : une méthode intégrée à la transformation des processus*. 3e éd. Sainte-Foy: Presses de l'Université du Québec. 718 p.
- Roger T. Pédaque (Collectif) et al. (2006). *Le document à la lumière du numérique*. Caen: C & F. 218 p.
- Rousseau, J.-Y. et Couture, C. (1994). *Les fondements de la discipline archivistique*. Sainte-Foy, Québec: Presses de l'Université du Québec. 348 p.
- Roy, S. N. (2003). « L'étude de cas. » Dans: Gauthier, B., éd. *Recherche sociale: de la problématique à la collecte des données*. Sainte-Foy: Presses de l'Université du Québec, pp. 159-184.
- Rygiel, P. et et al. (2005). *Internet et mouvements sociaux: nouvelles pratiques militantes, nouvelles sources pour l'histoire*. Nanterre: Bibliothèque de documentation internationale contemporaine. 112 p. (MATERIAUX pour l'histoire de notre temps).
- Sanderson, R. (2012). « Global web archive integration with memento. » Dans: *JCDL'12 proceedings of the 12th annual international ACM/IEEE-CS joint conference on Digital Libraries*, Washington DC, É-U, 10-14 juin, pp. 379-380. Adresse Url: <http://dl.acm.org/citation.cfm?id=2232900> (Page consultée le: 26 octobre 2012).

- Savoie-Zajc, L. (2003). « L'entrevue semi-dirigée. » Dans: Gauthier, B., éd. *Recherche sociale: de la problématique à la collecte des données*. Sainte-Foy: Presses de l'Université du Québec, pp. 293-316.
- Schamber, L. (1996). « What is a document? Rethinking the concept in uneasy times. » *Journal of American Society of Information Sciences*, **47**(9), pp. 669-671.
- Schneider, S. M. (2004). *September 11 web archive project: final project report*. USA: Library of Congress. 28 p. Adresse Url: <http://webarchivist.org/911-final-report.pdf> (Page consultée le: 15 octobre 2006).
- Schneider, S. M. et al. (2003). « Building thematic web collections: challenges and experiences from the september 11 web archive and the election 2002 web archive. » Dans: *European conference on digital libraries workshop on web archives*, 21 août, Trondheim, Norvège, 21 août, pp. 77-94. Adresse Url: <http://bibnum.bnf.fr/ecdl/2003/> (Page consultée le: 17 octobre 2006).
- Secrétariat du Conseil du Trésor du Canada. (2012). *Gestion et archivage du contenu Web*. [En ligne]. Adresse Url: <http://www.tbs-sct.gc.ca/im-gi/imp-pgi/mawc-gacw-fra.asp> (Page consultée le: 23 mars 2012).
- Sénécal, S. (2005). « The effect of the web on Archives. » *Archivaria*(59), pp. 139-152.
- Senserini, A. et al. (2004). « Archiving and accessing Web pages: the Goddard library Web capture project. » *D-Lib Magazine*, **10**(11).
- SERDA et al. (2003). *Panorama mondial de l'archivage du web: au profit de la Cellule Wall-On-Line. Version : VI*. Ministère de la Région Wallonne. Cellule WALL-ON-LINE. Adresse Url: http://archivesweb.wallonie.be/apps/spip/IMG/pdf/SC0335_09b.pdf (Page consultée le: 26 septembre 2005).
- Shepherd, M. et Watters, C. (1998). « The evolution of cybergenres. » Dans: *Thirty-first annual Hawaii international conference on system sciences*, Kohala Coast, Hawaii, 6-9 janvier. Adresse Url: <http://csdl2.computer.org/comp/proceedings/hicss/1998/8236/02/82360097.pdf> (Page consultée le: 07 janvier 2006).
- Shepherd, M. et Watters, C. (1999). « The functionality attribute of cybergenres. » Dans: *Proceedings of the 32nd Hawaii international conference on system sciences (HICSS '99)*, 1999, Maui, Hawaii, 5-8 janvier. IEEE. Adresse Url: http://web.cs.dal.ca/~shepherd/pubs/functionality_attribute.pdf (Page consultée le: 07 janvier 2006).
- Shepherd, M. et al. (2004). « Cybergenre: automatic identification of home pages on the web. » *Journal of Web Engineering*, **3**(3-4), pp. 236-251.
- Sijtsma, L. (2001). « The Dutch Deposit of Electronic Publications (DNEP) - 1995-2000. » *Liber Quarterly*(11), pp. 186-210.
- Sokvitne, L. (1999). « Our Digital Island: web preservation issues and solutions at the

- State Library of Tasmania. » Dans: *AusWeb99, fifth Australian world wide web conference*, Ballina, NSW, 17-20 avril. Adresse Url: <http://ausweb.scu.edu.au/aw99/papers/sokvitne/paper.html> (Page consultée le: 15 octobre 2006).
- Starr, J. et al. (2012). « A collaborative framework for data management Services: the experience of the University of California. » *Journal of eScience Librarianship*, **1**(2).
- State Library of Tasmania. (2011). *Guidelines for selecting, archiving and preserving websites pertinent to Tasmanian government information and cultural heritage*. [En ligne]. Adresse Url: <http://odi.statelibrary.tas.gov.au/About/selpolicy.asp> (Page consultée le: 27 janvier 2009).
- Stenger, T. et Coutant, A. (2010). « Les réseaux sociaux numériques: des discours de promotion à la définition d'un objet et d'une méthodologie de recherche. » *Journal of language and communication studies*(44), pp. 209-228.
- Stephens, D. O. (2001). « Managing records and information in web environments: policies for multinational companies. » *The Information Management Journal*(Avril), pp. 64-67.
- Stirling, P. et al. (2012). « Web archives for researchers: representations, expectations and potential uses. » *D-Lib Magazine*, **18**(13/4).
- Strodl, S. et al. (2009). « Migrating content in WARC files. » Dans: *The 9th international web archiving workshop (IWA 2009)*, Corfu, Grèce, pp. 7. Adresse Url: http://www.ltu.se/cms_fs/1.83925!/file/Migrating_Content_in_WARC_Files.pdf
- Taniar, D. et Rahayu, J. W. (2004). *Web information systems*. Hershey, Penn.: Idea Group Pub. (<http://www.books24x7.com/marc.asp?isbn=1591402085>).
- Tellis, W. (1997). « Introduction to case study. » *The Qualitative Report*, **3**(2).
- Tesch, R. (1990). *Qualitative research: analysis types and software tools*. Bristol, PA: Falmer Press. 330 p.
- The Central Office of Information. (2009). *Web standards & guidelines: archiving websites (TG105)*. [En ligne]. Adresse Url: <http://coi.gov.uk/documents/guidance/archiving-websites-v1-0.pdf> (Page consultée le: 27 janvier 2009).
- The National Library of Finland. ([s.d.].) *Web harvesting*. [En ligne]. Adresse Url: <http://www.nationallibrary.fi/publishers/deposit/webharvesting.html> (Page consultée le: 22 janvier 2009).
- The University of Melbourne. (2008a). *Overview of the program*. [En ligne]. Adresse Url: <http://www.unimelb.edu.au/records/web-archiving/overview.html> (Page consultée le: 27 janvier 2009).
- The University of Melbourne. (2008b). *Advanced systems access*. [En ligne]. Adresse Url: <http://www.unimelb.edu.au/records/web-archiving/advanced.html> (Page consultée le: 27 janvier 2009).

- The University of Melbourne. (2008c). *Technology*. [En ligne]. Adresse Url: <http://www.unimelb.edu.au/records/web-archiving/technology.html> (Page consultée le: 27 janvier 2009).
- The University of Melbourne. (2008d). *How do I notify the web archiving program about missing pages or want to register a 'soon to be' decommissioned web site?* [En ligne]. Adresse Url: <http://www.unimelb.edu.au/records/web-archiving/notify.html> (Page consultée le: 27 janvier 2009).
- The University of Melbourne. (2008e). *Frequently Asked Questions*. [En ligne]. Adresse Url: <http://www.unimelb.edu.au/records/web-archiving/faqs.html> (Page consultée le: 27 janvier 2009).
- The University of Melbourne. (2008f). *Compliance with PROV Web archiving*. [En ligne]. Adresse Url: <http://www.unimelb.edu.au/records/web-archiving/prov.html> (Page consultée le: 27 janvier 2009).
- The University of Melbourne. (2008g). *Web archiving collection management plan: web archiving @ the University of Melbourne*. [En ligne]. Adresse Url: <http://www.unimelb.edu.au/records/web-archiving/criteria.html> (Page consultée le: 27 janvier 2009).
- The University of Melbourne. (2010). *Web archiving policy*. 4 p. Adresse Url: http://www.unimelb.edu.au/records/pdf/94027_Web_Archiving_Policy_August_2009_%28current%29.pdf (Page consultée le: 27 janvier 2009).
- The University of Melbourne. (2011a). *How do I search the University's web archive?* [En ligne]. Adresse Url: <http://www.unimelb.edu.au/records/web-archiving/search.html> (Page consultée le: 27 janvier 2009).
- The University of Melbourne. (2011b). *Web archiving @ the University of Melbourne*. [En ligne]. Adresse Url: <http://www.unimelb.edu.au/records/web-archiving/> (Page consultée le: 27 janvier 2009).
- The University of Melbourne. (2011c). *How do I search the University's web archive? Access the list*. [En ligne]. Adresse Url: http://www.unimelb.edu.au/records/web-archiving/docs/131470_List_of_archived_University_websites.pdf (Page consultée le: 27 janvier 2009).
- Thelwall, M. A. (2004). *Link analysis: an information science approach*. 1st éd. Amsterdam ; Boston: Elsevier Academic Press. 269 p.
- Thibodeau, K. (1999-2000). « Rupture ou continuité: l'évaluation des archives au seuil de l'époque numérique. » *Archives*, **31**(3), pp. 61-72.
- Thibodeau, K. (2002). *Définition et caractéristiques diplomatiques des archives électroniques*. École des chartes, 4 mars 2002. 18 p.
- TNA - The National Archives. (2006). *Operational selection policy OSP27: the selection of government websites*. 8 p. Adresse Url: <http://www.nationalarchives.gov.uk/documents/osp27.pdf> (Page consultée le: 27 janvier 2009).

- TNA - The National Archives. (2009a). *Archiving websites*. 13 p. Adresse Url: <http://coi.gov.uk/guidance.php?page=240> (Page consultée le: 12 juin 2011).
- TNA - The National Archives. (2009b). *Managing URLs*. [En ligne]. Adresse Url: <http://coi.gov.uk/documents/guidance/managing-urls-v1-0.pdf> (Page consultée le: 27 janvier 2009).
- TNA - The National Archives. (2011). *Operational selection policy 27: UK central government websites*. 8 p. Adresse Url: <http://www.nationalarchives.gov.uk/documents/draft-operational-selection-policy-27.pdf> (Page consultée le: 27 janvier 2009).
- TNA - The National Archives. ([s.d.]a). *Information on web archiving*. [En ligne]. Adresse Url: <http://www.nationalarchives.gov.uk/webarchive/information.htm> (Page consultée le: 27 janvier 2009).
- TNA - The National Archives. ([s.d.]b). *Information for webmasters*. [En ligne]. Adresse Url: <http://www.nationalarchives.gov.uk/webarchive/webmasters.htm> (Page consultée le: 27 janvier 2009).
- TNA - The National Archives. ([s.d.]c). *Themed collections*. [En ligne]. Adresse Url: <http://www.nationalarchives.gov.uk/webarchive/themed-collections.htm> (Page consultée le: 27 octobre 2012).
- Toms, E. (2001). « Recognizing digital genre. » *Bulletin of the American Society for Information Science and Technology*, 27(2), pp. 20-22.
- Turnbaugh, R. C. (2002). « What is an electronic record? » Dans: Dearstyne, B.W., éd. *Effective approaches for managing records and archives*. Maryland: The Scarecrow Press, Inc., pp. 23-34.
- Tyrväinen, P. et Päivärinta, T. (1999). « On rethinking organizational document genres for electronic document management. » Dans: *Proceeding of the 32nd annual Hawaii international conference on system sciences*, Maui, Hawaii, 5-8 janvier. IEEE Computer Society Press. Adresse Url: <http://computer.org/proceedings/hicss/0001/00012/00012011.PDF> (Page consultée le: 07 janvier 2006).
- Umar, A. (2005). « IT infrastructure to enable next generation enterprises. » *Information Systems Frontiers*, 7(3), pp. 217-256.
- UNESCO. (2003). *Charte sur la conservation du patrimoine numérique* [En ligne]. Adresse Url: <http://www.unesco.org/new/fr/communication-and-information/access-to-knowledge/preservation-of-documentary-heritage/digital-heritage/> (Page consultée le: 15 octobre 2006).
- University of London Computer Centre et University of Bath. (2008). *JISC PoWR: the preservation of web resources, digital preservation issues relevant to the UK HE/FE web management community : draft Handbook*. London: UKOLN, The Library, University of Bath. 67 p. Adresse Url: http://jiscpowr.jiscinvolve.org/files/2008/09/handbookdraft_10092008.pdf (Page consultée le: 22 janvier 2009).

- University of North Texas Libraries. (2009). *CyberCemetery*. [En ligne]. Adresse Url: <http://www.library.unt.edu/govinfo/digital-collections/cybercemetery-1/> (Page consultée le: 22 janvier 2009).
- University of Tasmania. (2010a). *Keep your content up to date*. [En ligne]. Adresse Url: <http://www.web-services.utas.edu.au/resources/site-management/keep-your-content-up-to-date> (Page consultée le: 27 janvier 2009).
- University of Tasmania. (2010b). *Archive your website*. [En ligne]. Adresse Url: <http://www.web-services.utas.edu.au/resources/site-management/archive-your-website> (Page consultée le: 27 janvier 2009).
- University of Tasmania. (2010c). *Records management guidelines*. 38 p. Adresse Url: http://www.it.utas.edu.au/__data/assets/pdf_file/0014/32513/Records-Management-Guidelines.pdf (Page consultée le: 27 janvier 2009).
- University of Toronto Libraries. (2009). *Archiving the web*. [En ligne]. Adresse Url: <http://www.archives.gov/records-mgmt/initiatives/web-tech.html> (Page consultée le: 27 janvier 2009).
- University of Toronto. Archives and records management services. (2006). *Electronic records: guidelines for archiving Web sites*. [En ligne]. Adresse Url: <http://utarms.library.utoronto.ca/university-administrators/managing-records/electronic-records> (Page consultée le: 27 janvier 2009).
- University of Toronto. Archives and records management services. (2009). *Web archives*. [En ligne]. Adresse Url: <http://utarms.library.utoronto.ca/researchers/web-archives> (Page consultée le: 27 janvier 2009).
- Upward, F. (2005). « The records continuum. » Dans: McKemmish, S. et al., édés. *Archives: Recordkeeping in society*. Wagga Wagga, N.S.W.: Centre for Information Studies Charles Sturt University, pp. 197-222.
- Verheul, I. et Fédération internationale des associations de bibliothécaires et des bibliothèques. (2006). *Networking for digital preservation: current practice in 15 national libraries*. München: K.G. Saur. 269 p.
- Wagner, R. G. et Gross, J. (2004). *Harvesting the web, preserving chinese voices. The Digital Archive for Chinese Studies (DACHS) in Heidelberg*. Allemagne: University of Heidelberg, Center for Digital Resources in Chinese Studies. 31 p. Adresse Url: <http://www.sino.uni-heidelberg.de/dachs/publ.htm> (Page consultée le: 22 janvier 2009).
- Waller, M. et Sharpe, R. (2006). *Mind the gap: assessing digital preservation needs in the UK*. Digital Preservation Coalition. 40 p. Adresse Url: <http://www.dpconline.org/docs/reports/uknamindthegap.pdf> (Page consultée le: 17 octobre 2006).
- Wang, S. H. (2001). « Toward a general model for web-based information systems. » *International Journal of Information Management*, **21**(5), pp. 385-396.
- Watters, C. et Shepherd, M. A. (1997). « The digital broadsheet: an evolving genre. »

- Dans: *Proceedings of the thirtieth annual Hawaii international conference on system sciences*, Wailea, Hawaii, 07-10 janvier, pp. 22-29. Adresse Url: <http://www.computer.org/comp/proceedings/hicss/1997/7734/06/7734060022.pdf> (Page consultée le: 07 janvier 2006).
- WebArchivist.org. ([s.d.]). [En ligne]. Adresse Url: <http://webarchivist.org/about.htm>
- webarchivists.org. (2009-2012). *webarchivists camp*. [En ligne]. Adresse Url: <http://www.webarchivists.org/category/francais/> (Page consultée le: 23 octobre 2012).
- WebCite Consortium. ([s.d.]). *WebCite*. [En ligne]. Adresse Url: <http://www.webcitation.org/index>
- Wildemuth, B. M. (2009). *Applications of social research methods to questions in information and library science*. London: Libraries Unlimited. 421 p.
- Willer, M. (2008). « Selective archiving of web resources: a study of processing costs. » *Electronic Library and Information Systems*, **42**(4), pp. 341-364.
- Yan, H. et al. (2004). « New data storage and service model of China Web InfoMall. » Dans: *IWAW04, 4th international web archiving workshop in conjunction with the ECDL2004*, Bath, UK, 16 septembre, pp. 10. Adresse Url: <http://wenku.baidu.com/view/719a3af14693daef5ef73d36.html> (Page consultée le: 22 janvier 2009).
- Yates, J. et Orlikowski, W., J. (1992). « Genres of organizational communication: a structurational approach to studying communication and media. » *Academy of Management Review*, **17**(2), pp. 299-326.
- Yates, J. et Summer, T. R. (1997). « Digital genres and the new burden of fixity. » Dans: *Proceedings of the thirtieth annual Hawaii international conference on system sciences*, Wailea, Hawaii, 7-10 janvier. Adresse Url: http://reference.kfupm.edu.sa/content/d/i/digital_genres_and_the_new_burden_of_fix_59538.pdf (Page consultée le: 07 janvier 2006).
- Yin, R. K. (2003). *Applications of case study research*. 2e éd. Thousand Oaks: Sage Publications. 173 p. (Applied social research methods series).
- Yoshioka, T. et al. (2001). « Genre taxonomy: a knowledge repository of communicative actions. » *ACM Transactions on Information Systems*, **19**(4), pp. 431-456.
- Yusof, Z. M. et Chell, R. W. (1998). « The eluding definitions of records and records management: is a universally acceptable definition possible? Part 1. Defining the record. » *Records Management Journal*, **8**(2), pp. 95-112.
- Yusof, Z. M. et Chell, R. W. (2002). « Towards a theoretical construct for records management. » *Records Management Journal*, **12**(2), pp. 55-64.
- Zeldman, J. (2002). « 99.9% of websites are obsolete. » *Digital web Magazine*(4 septembre).
- Zeller, J.-D. (2004). « Documents numériques: à la recherche d'une typologie

perdue. » *Document Numérique*, 8(2), pp. 101-116.

- Zeller, J.-D. (2011). « À qui appartient la responsabilité de conserver les blogues: bibliothécaires, archivistes ou particuliers? » Dans: *GARM 2011: les archives de l'intimité : des journaux intimes aux blogues*, Montréal, 22 et 23 février, pp. 16. Adresse Url: http://ville.montreal.qc.ca/pls/portal/docs/PAGE/ARCHIVES_FR/MEDIA/DOCUMENTS/ARCHIVER_LES_BLOGS_MONTREAL_2011_NOTES.PDF (Page consultée le: 27 février 2011).
- Zwarich, N. et Park, E. G. (2008). « Policies of managing web resources at the canadian government: a records management perspective. » Dans: *Proceedings of the 36th annual confrence of the Canadian Association for Information Science (CAIS)*, University of British Clombia, Vancouver, 5-7 juin. Adresse Url: http://www.cais-acsi.ca/proceedings/2008/zwarich_2008.pdf (Page consultée le: 6 janvier 2006).

Annexe 1 : Principales initiatives d'archivage du Web (dernière mise à jour : novembre 2012)

Il existe des firmes privées spécialisées dans l'archivage du Web (ex. : Hanzo Archives, Page Freezer, Aleph Archives). Nous ne les avons pas recensées dans le tableau puisque nous ne nous intéressons pas à l'étude de solutions commerciales pour l'archivage du Web organisationnel.

Lorsque nous n'avons pas accès aux collections archivées ou à une documentation précise sur un projet d'archivage, nous nous limitons à une description générale des objectifs du projet sans spécifier les contenus archivés, le mode, l'approche et la technique de leur archivage (ex. : projet NWA).

Nous avons identifié les projets qui archivent, volontairement ou involontairement, des contenus à valeur archivistique. Un astérisque précède, dans ce cas, le nom du projet.

PAYS	PROJET	RESPONSABLE DU PROJET	CONTENUS ARCHIVÉS	MODES D'ARCHIVAGE	APPROCHE D'ARCHIVAGE	TECHNIQUES D'ARCHIVAGE	RÉFÉRENCES
Allemagne	deposit. D-NB.DE — 2001	German National Library.	Publications en ligne disponibles également dans un autre format (musique, texte, images, etc.) et documents natifs Web (e-books, sites Web). Sont exclus : les prépublications temporaires, applications logicielles, émissions radiophoniques et télévisées, publication commerciale et la publicité.	Dépôt légal Dans l'avenir : collectes exhaustives et automatiques.	Sélective (collecte de publications individuelles disponibles également dans une version imprimée). D'autres approches seront testées dans un avenir proche.	Pas d'information. Dans l'avenir : développement d'un logiciel libre pour la collecte des sites Web (OAI harvesting).	Aschenbrenner, 2004; DNB, 2009
	DACHS : Digital	Institute of	Ressources Internet reflétant le	Moissonnage,	Collecte sélective,	Archivage à distance	Institute of

PAYS	PROJET	RESPONSABLE DU PROJET	CONTENUS ARCHIVÉS	MODES D'ARCHIVAGE	APPROCHE D'ARCHIVAGE	TECHNIQUES D'ARCHIVAGE	RÉFÉRENCES
	Archive for Chinese Studies — 2001	Chinese Studies, University of Heidelberg.	discours social et politique sur la Chine (Web chinois ou autre : communautés chinoises d'outre-mer, étudiants chinois à l'étranger, chercheurs, institutions ou médias intéressés par la Chine).	achat, don.	préventive (censure, disparition imprévisible) et événementielle semi-automatique. Sites suggérés par des informateurs ou ciblés lors d'événements importants.	selon le modèle OAIS (application de critères de sélection, contrôle de qualité, description des ressources et accès selon certaines conditions).	Chinese Studies, 2008; Nivard Bigourdan, 2005; Sijtsma, 2001; Wagner et Gross, 2004
Australie	PANDORA : Preserving and Accessing Networked Documentary Resources of Australia — 1996	National Library of Australia et bibliothèques partenaires (bibliothèques provinciales et autres institutions patrimoniales).	Publications en ligne, sites Web.	Dépôt volontaire, moissonnage à distance.	Collecte sélective semi-automatique + collecte exhaustive grâce aux services d'un tiers-archiviste : Archive-It. Approche collaborative décentralisée.	Photographies et archivage à la source grâce à un système d'archivage maison PANDAS.	Beagrie, 2003; Day, 2003b; Gatenby, 2004; Koerbin, 2004, 2005; Phillips, 2003; Pymm et Walis, 2009
	* Our Digital Island : A Tasmanian web archive	State Library of Tasmania.	Sites Web gouvernementaux, patrimoine documentaire tasmanien en ligne et les traces de l'évolution de l'usage du Web en Tasmanie.	Principalement dépôt légal + dépôt volontaire.	Différentes approches selon le type de contenu (collecte exhaustive et en profondeur des rapports et textes de lois à partir du Web gouvernemental, capture d'instantanés sélectifs des sites tasmaniens innovateurs).	Archivage à distance grâce au logiciel Anawave WebSnake. Archive accessible en ligne.	Day, 2003b; Sokvitne, 1999; State Library of Tasmania, 2009
Autriche	AOLA : Austrian OnLine Archive — 2001 Ensuite : Webarchive Austria — 2008	Austrian National Library.	Sites Web d'intérêt pour l'Autriche, ensuite, collecte du domaine national (.at) ou collectes thématiques (sites Web ayant des liens spécifiques avec l'Autriche).	Moissonnage exhaustif et sélectif (sélection de sous-domaines).	Approche intégrée : sélection manuelle et moissonnage automatique.	Archivage à distance (Nedlib harvester) : snapshots du domaine Web autrichien.	Aschenbrenner, 2004; Day, 2003b; IIPC, 2012; NLA, 2007; Verheul et al., 2006

PAYS	PROJET	RESPONSABLE DU PROJET	CONTENUS ARCHIVÉS	MODES D'ARCHIVAGE	APPROCHE D'ARCHIVAGE	TECHNIQUES D'ARCHIVAGE	RÉFÉRENCES
Belgique	*Archives des sites Web de la Région wallonne — 2004	Cellule Wall-On-Line, ensuite le commissariat EASI-WAL. Le Commissariat EASI-WAL résulte de la fusion en 2005 du Commissariat à la Simplification Administrative et de Wall-On-Line (placé sous l'autorité directe du Ministre-Président de Wallonie)	Les sites institutionnels belges, le Web régional belge, les sites événementiels et des sites thématiques, notamment les sites menacés de disparition (fin d'un contrat d'hébergement, sites temporaires créés pour une initiative particulière, sites révolus) ou disparus (suite à une mise à jour, à une refonte majeure, à un renouvellement de ministre).	Moissonnage à distance.	Approche élective.	Archivage à distance : capture d'instantanés périodiques.	Commissariat EASI-WAL, [s.d.]; Haettiger, 2003b; Région wallonne, 2009; Serda, 2003
Canada	PPPE : projet pilote des publications électroniques – 1994-1996, ensuite Collection électronique permanente — 1997	Bibliothèque et Archives Canada.	Publications en ligne d'origine canadienne, traitant d'un sujet canadien, ou dont le créateur est canadien; les publications officielles du gouvernement fédéral du Canada. Sont exclus : les publications non officielles ou incomplètes, les sites de promotion, les documents publicitaires, les documents liés par des liens hypertextes ouverts et les sites de services.	Dépôt volontaire, dépôt obligatoire, dépôt négocié avec les éditeurs. Les publications sont reçues par courrier électronique, ou copiées, téléchargées, ou reproduites directement à partir du site de l'éditeur ou du créateur.	Approche sélective selon certains critères (voir contenu archivé). Sélection des publications non canadiennes selon les domaines d'intérêt de la Bibliothèque.	Archivage à distance ou dépôt de document par voie électronique. Le matériel archivé est conservé dans un dépôt numérique fiable basé sur le modèle OAIS. Documents catalogués et cotés (archivés, versés ou liés), accessibles à partir d'Amicus.	BAC, 1998; Chebbi, 2007-2008; Lemaître, 2001; NLC, 1996
	*AW GC : Archive Web du	Bibliothèque et Archives Canada.	Web gouvernemental et des sites Web des administrations	Moissonnage à distance du	Sélective par échantillonnage ou	Archivage à distance : capture	BAC, 2007-2008; Brodie,

PAYS	PROJET	RESPONSABLE DU PROJET	CONTENUS ARCHIVÉS	MODES D'ARCHIVAGE	APPROCHE D'ARCHIVAGE	TECHNIQUES D'ARCHIVAGE	RÉFÉRENCES
	gouvernement du Canada — 2005-2006		provinciales et territoriales, sites Web de partis politiques, etc. Sites intranet et extranet ministériels à la demande.	domaine gouvernemental ou de sites d'intérêt pour les citoyens (événements importants, sujets d'intérêt pour les Canadiens).	Approche thématique.	d'instantanés périodiques (2 fois par an) côté client grâce au robot Heritrix. L'accès et la recherche dans l'archive Web sont possibles grâce au logiciel de dépôt Web Nutchwax et au logiciel visionneuse Wayback.	1997; Cantello et Stegenga, 2008
	WebCite — 2005	WebCite consortium (Toronto).	Membre de l'IIPC, WebCite est un service d'archivage du Web gratuit et sur demande à des fins de référence (pour les chercheurs, les blogueurs, les éditeurs scientifiques).	Moissonnage à distance.	Approche sélective.	Archivage à distance. Le document archivé demeure toujours accessible via la base de données de WebCite tel qu'il a été au moment de sa capture. Le système génère un lien vers la version archivée dans WebCite (une sorte de lien permanent vers la ressource).	WebCite Consortium, [s.d.]
Canada, Québec	Dépôt légal des publications gouvernementales du Québec en ligne — 2001	Bibliothèque et Archives nationales du Québec.	Publications en ligne (entités individuelles) : priorité accordée aux publications gouvernementales.	Dépôt légal (FTP puis formulaire en ligne) et dépôt volontaire.	Approche sélective semi-automatisée.	Capture à la source grâce à : MemoWeb, HTTrak et un utilitaire pour visualiser le code	BAnQ, 2001; Cleroult, 2006; Gharsallah, 2002; Léger, 2003; Lemaître,

PAYS	PROJET	RESPONSABLE DU PROJET	CONTENUS ARCHIVÉS	MODES D'ARCHIVAGE	APPROCHE D'ARCHIVAGE	TECHNIQUES D'ARCHIVAGE	RÉFÉRENCES
						HTML.	2001; Racine, 2007
	*Campagne électorale 2012 — 2012	Bibliothèque et Archives nationales du Québec.	Moissonnage des sites suivants à plusieurs moments pendant la campagne électorale provinciale 2012 : les sites du Parti Québécois, du Parti libéral du Québec, de la Coalition Avenir Québec, de Québec solidaire, d'Option nationale et du Parti vert du Québec. Le site du Directeur général des élections a également fait partie de la collecte.	Dépôt légal par moissonnage à distance après obtention des autorisations d'archivage et de diffusion des sites archivés.	Collecte sélective (thématique) et automatisée.	Capture à distance grâce au robot Heritrix. Accès aux archives se fait aux locaux de BAnQ.	BAnQ, 2012
Catalogne (Espagne)	*PADICAT (Digital Heritage of Catalonia) : The Web Archive of Catalonia — 2005	Biblioteca de Catalunya.	Moissonnage du domaine (.cat) et archivage systématique des sites Web des organisations catalanes. Archivage des contenus relatifs à des événements majeurs (événements politiques) ou des formes particulières de contenus (musée en ligne).	Le mode d'archivage n'est pas spécifié. L'archivage du Web est un projet considéré dans le plan stratégique de 2004.	Collecte exhaustive du domaine (.cat) et collecte sélective d'autres contenus à valeur scientifique et historique. Approche collaborative : plusieurs institutions sont invitées à soumettre des suggestions de sites Web à archiver.	Moissonnage à distance avec Heritrix. Les fichiers sont stockés et compressés dans le format ARC.	Biblioteca de Catalunya, 2011; Llueca et al., 2010
Chine	Web InfoMall — 2001	Xiaoming Li at Peking University + Internet Laboratory of Peking University.	Pages Web uniquement : le Web chinois. Les contenus multimédias sont capturés dans un autre système : CDAL (Chinese Digital Assets Library).	Moissonnage à distance.	Approche exhaustive et automatique.	Archivage à distance : capture d'instantanés périodiques grâce au robot Tianwang. Il est possible de faire une recherche dans la collection en inscrivant l'URL du	Beijing University, 2004; Huang et al., 2008; Yan et al., 2004

PAYS	PROJET	RESPONSABLE DU PROJET	CONTENUS ARCHIVÉS	MODES D'ARCHIVAGE	APPROCHE D'ARCHIVAGE	TECHNIQUES D'ARCHIVAGE	RÉFÉRENCES
						site Web archivé ou par date de capture.	
Croatie	Croatian Web Archive (HAW) — 2003	National and university library in Zagreb.	Web national (portion bien définie et délimitée) : publications statiques et dynamiques.	Dépôt légal, suggestions de sites Web.	Approche sélective semi-automatique.	Archivage à distance grâce à un robot maison : NUL gatherer. Le matériel est accessible grâce à une application de recherche et de visualisation : NUL's WebPAC.	Buzina et Willer, 2007; Milinovic et Topolscak, 2008; National and University Library in Zagreb, 2005; NLA, 2007
Danemark	netarchive.dk — 1998	The State and University Library avec la coopération de la National Library of Denmark.	Le premier niveau du domaine danois (Web public).	Moissonnage à distance (robot de collecte) autorisé par la loi sur le dépôt légal.	Approche exhaustive et automatique avec sélection de sites danois hébergés sur autres domaines.	Archivage à distance : capture d'instantanés périodiques. Accès limité (sur place à la bibliothèque nationale ou à State and University Library). Recherche dans la collection en inscrivant l'URL du site Web archivé ou par date de capture.	Aschenbrenner, 2004; Christensen, 2005; Jacobsen, 2007; Mannerheim, 2000; netarchive.dk, 2008
Égypte	Bibliotheca Alexandrina's	Bibliothèque d'Alexandrie et	Contenus archivés par Internet Archive correspondant aux	Moissonnage à distance grâce au	Approche exhaustive automatique.	Archivage à distance :	Bibliotheca Alexandrina,

PAYS	PROJET	RESPONSABLE DU PROJET	CONTENUS ARCHIVÉS	MODES D'ARCHIVAGE	APPROCHE D'ARCHIVAGE	TECHNIQUES D'ARCHIVAGE	RÉFÉRENCES
	Internet Archive — 1996	Internet Archive.	intérêts de la Bibliothèque d'Alexandrie et aux besoins de ses utilisateurs.	service d'un tiers-archiviste pour les archives capturées entre 1996-2007, ensuite archivage maison.		réalisations d'instantanés du Web égyptien basées sur la technologie d'Internet archive.	2008
États-Unis	Internet Archive — 1996	Internet Archive.org.	La toile mondiale, sites Web liés à des événements ou sujets contemporains et importants.	Moissonnage à distance, services de tiers archiviste (Archive-It).	Intégrale automatique thématique et sélective.	Archivage à distance : capture d'instantanés de sites et domaines Web.	Chaimbault, 2008; Day, 2003b; Gharsallah, 2002
	MINERVA : Mapping the Internet Electronic Resources Virtual Archive — 2007	Library of Congress.	Web américain et international d'intérêt pour le public de la LC.	Moissonnage à distance.	Approche sélective thématique.	Archivage à distance du côté client.	Kenny et al., 2002; Schneider et al., 2003; Schneider, 2004
	K-12 Web Archiving Program — 2008	Internet Archive et Library of Congress.	Internet Archive et la Bibliothèque du Congrès ont lancé le projet K-12 (étude pilote en 2008). Le projet consiste à établir des collections Web du point de vue des étudiants d'une dizaine d'écoles dans 10 états américains. L'objectif est de les sensibiliser à l'importance de la préservation des sources primaires actuelles pour les recherches historiques futures.	Moissonnage à distance.	Approche sélective sur demande.	Archivage à distance : le projet utilise le service Archive-It d'Internet Archive.	Archive-it, [s.d]
	WebArchivist.org — (date non indiquée)	WebArchivist.org est un groupe de développeurs de	Le groupe assiste des chercheurs, bibliothécaires et archivistes intéressés par la préservation et l'analyse de matériel créé pour et diffusé sur le Web dans l'établissement de leur projet d'archivage du Web, le développement de procédures et de politiques d'archivage et la conception de logiciels pour				WebArchivist.org, [s.d.]

PAYS	PROJET	RESPONSABLE DU PROJET	CONTENUS ARCHIVÉS	MODES D'ARCHIVAGE	APPROCHE D'ARCHIVAGE	TECHNIQUES D'ARCHIVAGE	RÉFÉRENCES	
		logiciels basé à l'University of Washington et the SUNY Institute of Technology.	accomplir les projets d'archivage (ex. : WebArchivist Software Suite).					
	CommonCrawl — 2008	CommonCrawl (Californie).	Common Crawl Foundation est un organisme à but non lucratif qui a pour objectif la démocratisation d'accès aux informations diffusées sur le Web en développant un entrepôt de données Web universellement accessible d'où il est possible d'extraire les données Web archivées à des fins d'innovation en recherche ou qui portent sur le domaine des affaires ou de l'éducation. L'entrepôt comporte 3 ensembles de captures réalisées successivement en 2008, 2009 et 2012. Common Crawl a moissonné environ 5 milliards de pages et les métadonnées associées. Les pages capturées en format ARC sont stockées par Amazon's S3 service (à l'adresse suivante : http://aws.amazon.com/datasets/41740) autorisant ainsi le téléchargement des fragments du corpus Web archivé et l'usage du service « Amazon Elastic MapReduce » qui offre aux usagers la possibilité d'extraire, transformer et analyser les données Web qu'ils souhaitent exploiter.					CommonCrawl, [s.d.]
	Memento : Adding Time to the Web — 2009	Memento est une initiative collaborative entre un groupe de développeurs du Los Alamos National Laboratory et du Computer Science Department of Old Dominion University.	Le projet Memento consiste en un cadre technique pour l'archivage des contenus Web avec l'idée d'intégrer les archives Web à la navigation régulière sur le Web. Le projet est fondé sur l'approche de versionnage des ressources Web. Un outil d'archivage a été développé pour opérationnaliser cette approche. Memento est financé par la Library of Congress sous le « National Digital Information Infrastructure and Preservation Program ». Il est possible d'archiver des pages Web grâce à MementoFox plug-in développé pour le navigateur Firefox. Le site Web de Memento donne accès à d'autres outils essentiels pour implanter cette solution d'archivage du Web.					Memento, [s.d.]; Sanderson, 2012
	Web Archiving Service — 2003	California Digital Library.	Initiative collaborative pour l'archivage du Web à des fins de recherche scientifique menée par la California Digital Library. Les spécialistes d'indexation par sujet et les partenaires de la	Moissonnage à distance.	Approche sélective, collaborative.	Archivage à distance : il est possible de contribuer au développement des collections Web	California Digital Library, 2007-2012; Starr et al., 2012	

PAYS	PROJET	RESPONSABLE DU PROJET	CONTENUS ARCHIVÉS	MODES D'ARCHIVAGE	APPROCHE D'ARCHIVAGE	TECHNIQUES D'ARCHIVAGE	RÉFÉRENCES
			bibliothèque ont archivé une sélection de contenus Web « critiques ». Les archives Web publiques sont accessibles en ligne avec la possibilité de faire une recherche par ordre alphabétique du nom de la collection. La recherche en texte intégral, par liste de sites Web capturés, par nom de partenaire, ou par type de documents est possible dans chaque collection.			puisque la California Digital Library autorise l'usage de son logiciel de moissonnage WAS curator par ses partenaires. Les documents capturés ont une URL spécifique appelée « archival URL » qui offre accès aux sites Web archivés.	
	Web Archive Cooperative — 2010	Old Dominion University, Harding University et Stanford University.	Web Archive Cooperative (WAC) est un projet collaboratif ayant pour objet d'avancer la réflexion et la pratique de l'archivage du Web à des fins non commerciales. Le projet permet la collecte des ressources d'intérêt pour la recherche scientifique. Les matériaux collectés ne se limitent pas aux pages Web, mais incluent également des étiquettes, des blogues, des mises à jour Twitter, etc. Des collectes manuelles et automatisées seront entreprises.				WAC, 2012
	*PCWA : Political Communications Web Archiving — 2002	Center for Research Libraries CRL, New York University, Cornell University, Sandford University et University of Texas at Austin.	Web politique (sites Web et autres types de documents en ligne) produit par des groupes politiques et organismes non gouvernementaux en : Amérique latine, Europe de l'Est, Afrique subsaharienne, Asie du Sud-Est et l'Europe de l'Ouest.	Moissonnage à distance.	Approche sélective (thématique et événementielle) automatisée.	Capture à distance de documents Web statiques.	Reilly, 2002; Reilly et al., 2003
	*CyberCemetery — (date non indiquée)	University of North Texas Libraries, the U.S. Government	Les sites Web des anciens gouvernements américains (agences gouvernementales et commissions dissoutes)	Moissonnage à distance.	Approche sélective.	Archivage à distance : réalisation de captures périodiques.	University of North Texas Libraries, 2009

PAYS	PROJET	RESPONSABLE DU PROJET	CONTENUS ARCHIVÉS	MODES D'ARCHIVAGE	APPROCHE D'ARCHIVAGE	TECHNIQUES D'ARCHIVAGE	RÉFÉRENCES
		Printing Office, National Archives and Records Administration.	Ex. : White House Web Sites (différentes versions des sites Web de la Maison-Blanche (1994-2001).				
France	Dépôt légal d'internet — 2002	Bibliothèque nationale de France + Institut National de l'Audiovisuel.	Le Web français. L'INA est responsable de l'archivage des documents audiovisuels en ligne.	Dépôt légal et moissonnage à distance.	Approche combinée ou intégrée (collecte exhaustive ensuite sélective) + des collectes thématiques (ex. : les campagnes électorales).	Archivage à distance.	Beagrie, 2003; BnF, 2008; Day, 2003b
	Mig' Archive — 2005	Groupe d'études sur l'usage des TIC dans les migrations.	Sites Web dans toutes les langues des diasporas sur les migrations.	Non documenté.	Non documenté.	Non documenté.	Groupe d'études sur l'usage des TIC dans les migrations, 2005
	*archives.premier-ministre.gouv.fr — 1996	Service d'archives du cabinet du premier ministre.	Activité gouvernementale véhiculée sur le site du Premier ministre.	Archivage à la source.	Approche sélective (moments politiques importants).	Archivage à distance : captures périodiques du site du premier ministre.	Premier ministre, [s.d.]
	webarchivists.org — 2009	Association à but non lucratif fondée en 2009 (France).	Sensibiliser la communauté aux enjeux de l'archivage du Web et réunir des experts en la matière pour faire avancer la réflexion autour des sujets liés à l'archivage du Web. Le 22 novembre 2012, cette association a organisé le premier Barcamp sur l'archivage du Web (un événement participatif où tout le monde contribue, sous la forme d'ateliers dont les thématiques sont décidées sur le moment).				webarchivists.org, 2009-2012
Finlande	EVA — 1997 Ensuite, Finnish Web Archive — 2006	Finnish University Library.	Capture annuelle du Web finlandais (domaines : .fi et .ax). Sélection de sites Web d'intérêt (collecte thématique).	Dépôt volontaire ou moissonnage à distance.	Collecte exhaustive automatique du Web national, thématique semi-automatisée de sites d'intérêt (événements importants : élections, visites officielles, catastrophes, politiques internationales + sites en	Archivage à distance : collecte en mode FTP ou capture à distance grâce au logiciel Heritrix. Les contenus sont archivés dans le format WARC et	Aschenbrenner, 2004; Day, 2003b; IIPC, 2012; Lounamaa et Salonharji, 1999; Mannerheim, 2000; The National Library

PAYS	PROJET	RESPONSABLE DU PROJET	CONTENUS ARCHIVÉS	MODES D'ARCHIVAGE	APPROCHE D'ARCHIVAGE	TECHNIQUES D'ARCHIVAGE	RÉFÉRENCES
					danger de disparition).	catalogués dans la bibliographie nationale.	of Finland, [s.d.]
Grèce	Greek Web — 2003	Non documenté (premières expérimentations, laboratoires de recherche universitaire).	Web grec (en grec et en anglais).	Moissonnage à distance.	Collecte exhaustive et automatique.	Archivage à distance grâce à un robot de collecte.	Lamos et al., 2004
Japon	WARP : Web Archiving Project — 2002	National Diet Library.	Publications japonaises en ligne (sites Web et périodiques en ligne).	Moissonnage à distance avec autorisations préalables auprès de l'éditeur, ensuite, opérationnalisation de la loi sur le dépôt légal.	Premières expérimentations : collectes sélectives. Programme opérationnel en 2006 : collecte des sites des institutions publiques	Archivage à distance : captures périodiques du Web japonais et archivage à distance de publications individuelles utilisant le logiciel Web curator tool.	NLA, 2007
Lettonie	WWW Archive — 2005, 2008	National Library of Latvia.	Web letton, notamment les publications qui sont pertinentes selon les critères de sélection de la politique de développement des collections de la bibliothèque nationale.	Dépôt légal, dépôt volontaire, moissonnage à distance.	Approche sélective.	Archivage à distance : les logiciels Web Curator Tool, NutchWAX; Wayback Machine sont utilisés,	Gomes et al., 2011; National Library of Latvia, [s.d.]
Lituanie	Archive of Electronic Resources — 2001-2002	National Library of Lithuania.	Web national.	Dépôt légal, autorisations pour les contenus dynamiques Moissonnage à distance pour les contenus statiques.	Approche sélective.	Archivage à distance : captures d'instantanés périodiques du Web lituanien (2 fois par an) grâce au robot NEDLIB. Le	Jodelis, 2003; NLA, 2007

PAYS	PROJET	RESPONSABLE DU PROJET	CONTENUS ARCHIVÉS	MODES D'ARCHIVAGE	APPROCHE D'ARCHIVAGE	TECHNIQUES D'ARCHIVAGE	RÉFÉRENCES
						matériel collecté est catalogué avec les métadonnées du Dublin Core.	
Nouvelle-Zélande	New Zealand Web Archive — 1999-2005; 2008-	Alexander Turnbull Library, ensuite la National Library of New Zealand.	Moissonnage des sites Web sous le domaine (.NZ), mais aussi des sites sur le Pacifique et l'Antarctique ayant une valeur historique et scientifique.	Dépôt légal.	Approche sélective (thématique) et moissonnage exhaustif du domaine (.NZ).	Moissonnage à distance avec Heritrix. Gestion de l'archivage sélectif des sites Web avec Web Curator Tool.	National Library of New Zealand, 2008
Norvège	Paradigma — 2001	National Library of Norway.	Web national .no incluant documents statiques et dynamiques (beaucoup de problèmes avec les contenus dynamiques).	Dépôt légal et moissonnage à distance.	Approche combinée : exhaustive uniquement durant les premières phases.	Archivage à distance : capture d'instantanés fréquents notamment des sites de journaux. Accès au matériel archivé via le Nordic Web Archive (NWA), le portail des archives Web des pays nordiques.	Albertsen, 2003; NLA, 2007
Pays-Bas	eDeposit KB/ Dutch Deposit for Electronic Publications (DNEP) — 1996-2000 Ensuite Web archive of the Netherlands — 2007	Koninklijke Bibliotheek (National library of the Netherlands).	Initialement les publications académiques en ligne ensuite se sont ajoutés les livres électroniques, les livres audio et le Web néerlandais.	Dépôt Web et moissonnage à distance.	Approche sélective.	Archivage à distance grâce au système d'archivage DIAS (Digital Information and Archiving System), logiciel qui fonctionne sur le modèle OAIS.	Beagrie, 2003; Beunen et Schiphol, 2006; Ras et Sierman, 2006; Sijtsma, 2001

PAYS	PROJET	RESPONSABLE DU PROJET	CONTENUS ARCHIVÉS	MODES D'ARCHIVAGE	APPROCHE D'ARCHIVAGE	TECHNIQUES D'ARCHIVAGE	RÉFÉRENCES
	*archiPOL : Archive of web sites of political parties in the Netherlands — 2000	Documentation Centre for Dutch Political Parties et Groningen University Library.	Web politique néerlandais.	Moissonnage à distance.	Approche sélective.	Aspiration fréquente de sites Web sélectionnés.	Documentation centre for Dutch political parties, 2000
Portugal	Portuguese Web Archive (Tombo) — 2002-2006	XLDB research group at the University of Lisbon, ensuite National Foundation for Scientific Computing.	Le domaine Web portugais (.pt).	Moissonnage à distance.	Approche exhaustive automatique.	Archivage à distance grâce à un système d'archivage maison. Le matériel archivé est accessible en inscrivant l'URL du site archivé ou par date de capture.	FCCN, 2000-2009 Gomes et al., 2006; Gomes et al., 2008
République tchèque	WebArchiv — 2000	Czech National Library.	Web tchèque notamment sites Web du domaine biomédical et les biographies en ligne.	En absence d'un cadre légal adéquat, la bibliothèque nationale capture les ressources Web à distance.	Approche combinée : 1- Approche exhaustive automatique pour la collecte du domaine national. 2- Sélective thématique pour compléter la collection initiale (élections, jeux olympiques, etc.).	Archivage à distance grâce aux outils développés par l'IIPC.	Coufal et Žabička, 2007; Czech National Library, 2009; NLA, 2007
Royaume-Uni	UK Web Archive — 2004	British Library et bibliothèques participantes.	Contenus Web ayant une signifiante culturelle, sociale et historique (sites Web, blogues, etc.).	Moissonnage à distance avec autorisation.	Approche coopérative basée sur un archivage sélectif thématique (sujet et format) en accord avec la politique de développement des collections et de la mission des institutions impliquées dans le	Archivage à distance : capture périodique d'instantanés Web (2 fois par an, autres fréquences pour les sites événementiels) grâce au système d'archivage	Beagrie, 2003; British Library, [s.d.]; Day, 2003b; Millar, 2001; NLA, 2007; Waller et Sharpe, 2006

PAYS	PROJET	RESPONSABLE DU PROJET	CONTENUS ARCHIVÉS	MODES D'ARCHIVAGE	APPROCHE D'ARCHIVAGE	TECHNIQUES D'ARCHIVAGE	RÉFÉRENCES	
					projet.	PANDAS ensuite Web Curator Tool.		
	*UK Central Government Web Archive — 2003	The national Archives (les Archives nationales du Royaume-Uni).	Sites Web gouvernementaux britanniques.	Moissonnage à distance.	Sélective, mais exhaustive (tous les sites Web gouvernementaux).	Archivage à distance : capture d'instantanés fréquents. Matériel archivé accessible par URL ou Date.	British Library, [s.d.]; Brown, 2006	
	JISC PoWR : Preservation of Web Resources — 2008	University of Bath et ULCC Digital Archives department.	Projet entrepris afin de sensibiliser à l'importance de la préservation des ressources Web. Outre l'organisation d'ateliers, de conférences et l'expérimentation d'outils d'archivage du Web, le projet a favorisé la publication d'un guide sur la préservation des ressources Web et le maintien d'un blogue sur le sujet. Le blogue n'est plus actif depuis juillet 2010, mais demeure une ressource importante à consulter (version archivée du blogue encore en ligne).					Farell et al., 2010
Scandinavie	NWA : Nordic Web Archive — 2000	Projet coopératif entre plusieurs bibliothèques scandinaves (Danemark, Finlande, Islande, Bibliothèque royale de la Suède) visant le développement d'un prototype d'un système modulaire facilitant l'accès aux archives Web des bibliothèques participantes, et ce, dans le temps et dans l'espace.						Brygfjeld, 2002
Singapour	*Web Archive Singapore — (date non indiquée)	National Library Board.	Sites Web gouvernementaux de la République de Singapour, sites Web portant sur des thématiques populaires (Arts, loisirs, etc.) revêtant une signifiante nationale et historique, documentant la vie sociale et culturelle.	Moissonnage à distance. La technologie utilisée ne supporte pas certains documents multimédias ou formats particuliers (trous dans l'archive).	Approche sélective automatisée.	Archivage à distance grâce à une suite logicielle : le robot de collecte Heritrix, un robot d'indexation et de recherche dans l'archive, NutchWAX et une application de navigation et de consultation de l'archive, WERA.	IIPC, 2004-2008; National library board Singapore, 2009; NLA, 2007	
Slovénie	National repository for	National and University Library	Web slovène et les publications électroniques en ligne.	Dépôt volontaire Autorisations de	Approche sélective, mais exhaustive.	Archivage à distance.	Kavcic-Colic et Grobelnik, 2004;	

PAYS	PROJET	RESPONSABLE DU PROJET	CONTENUS ARCHIVÉS	MODES D'ARCHIVAGE	APPROCHE D'ARCHIVAGE	TECHNIQUES D'ARCHIVAGE	RÉFÉRENCES
	long-term preservation of Slovenian web and electronic resources — 2004 Ensuite Webarchive of Slovenia — 2007	of Slovenia et Jozef Stefan Institute.		collectes.			Kavcic-Colic, 2006
Suède	KulturarW3 — 1996	National Library of Sweden.	Tout le Web suédois (surface).	Moissonnage à distance par voie de dépôt légal.	Approche exhaustive automatisée de tout type de document sur le Web suédois : pages Web, images, sons, etc.	Archivage à distance par aspiration périodique du Web suédois : archivage côté client (.se, .nu et autres domaines spéciaux). Accès limité à l'archive (consultation sur place à la Bibliothèque Nationale).	Aschenbrenner, 2004; Arvidson et al., 2000; Day, 2003b; Gharsallah, 2002; Lemaître, 2001; Mannerheim, 2000; National Library of Sweden, 2008
Suisse	Archives Web suisses — 2008	Bibliothèque nationale suisse avec la collaboration des bibliothèques cantonales suisses et des bibliothèques universitaires.	Domaines Web suisses et documents individuels en ligne.	Moissonnage à distance suite au signalement ou la sélection des sites Web à aspirer et après obtention des autorisations d'archivage.	Approche sélective (thématique et événementielle).	Collecte à distance grâce au robot de collecte Heritrix. Le matériel archivé est catalogué.	Bättig, 2005; Locher, 2008
Taiwan	*NTUWAS : National Taiwan University Web	National Taiwan University Library.	Web universitaire, Web gouvernemental, Web politique, Web des célébrités, événements	Moissonnage à distance grâce à un système maison	Approche sélective et thématique automatique.	Archivage à distance.	Chen et al., 2008

PAYS	PROJET	RESPONSABLE DU PROJET	CONTENUS ARCHIVÉS	MODES D'ARCHIVAGE	APPROCHE D'ARCHIVAGE	TECHNIQUES D'ARCHIVAGE	RÉFÉRENCES
	Archiving System — 2006		signifiants.	d'archivage et HTTrack.			
Privé	AfterLife — (date non indiquée)	AfterLife.org.	Sites Web dont les auteurs sont décédés ou n'ayant plus les moyens pour garder en ligne leurs sites Web.	Dépôt volontaire, legs ou collecte rétrospective.	Approche sélective.	Moissonnage à distance + collecte à la source.	Blatner, [s.d.]a; Blatner, [s.d.]b
	Déjà Vu — (date non indiquée)	Pär Lannerö de Metamatrix (Suède).	Visualiser des anciennes versions de sites Web.	Développement d'émulateurs pour les différents navigateurs Web.			Metamatrix // Interfolio, [s.d.]

Principales sources d'information sur les initiatives d'archivage du Web :

- Aschenbrenner, A. (2004). *Web Archiving - bibliography*. Austrian On-Line Archive. [En ligne]. Adresse Url: <http://www.ifs.tuwien.ac.at/~aola/link/WebArchiving.html> (Page consultée le: 16 octobre 2006, 26 février 2011).
- Harvard University Library. (2008). *Web archiving resources*. [En ligne]. Adresse Url: <http://hul.harvard.edu/ois/systems/wax/resources.html> (Page consultée le: 21 janvier 2009, 26 février 2011).
- IIPC - International Internet Preservation Consortium. (2012). *Member archives*. [En ligne]. Adresse Url: <http://www.netpreserve.org/resources/member-archives> (Page consultée le: 26 février 2011).
- iPRES2011. *International Conference on Preservation of Digital Objects, 1-4 novembre 2011, Singapore*. [En ligne]. Adresse Url: <http://ipres2011.sg/> (Page consultée le: 30 mai 2012).
- NLA - National Library of Australia. (2007). *PADI : Preserving Access to Digital Information*. [En ligne]. Adresse Url: <http://pandora.nla.gov.au/pan/10691/20110824-1153/www.nla.gov.au/padi/topics/92.html> (Page consultée le: 23 avril 2007).

Annexe 2 : Liste des politiques d'archivage du Web et documents complémentaires par pays et par type d'organisme créateur (mise à jour novembre 2011)

CANADA

Bibliothèques et Archives Canada (Archives nationales, bibliothèque nationale)

- 1- BAC - Bibliothèque et Archives Canada. (1999). *Gestion de l'information sur Internet et les intranets pour assurer l'accès à long terme et la reddition de comptes: Guide de mise en œuvre*. Ottawa: Forum sur la gestion de l'information (GI). Groupe de travail (GT) sur les réseaux Internet / intranets. 75 p. Adresse Url: <http://www.lac-bac.gc.ca/gouvernement/produits-services/007002-2032-f.html> (Page consultée le: 15 octobre 2005).
- 2- BAC - Bibliothèque et Archives Canada. Groupe de coordination des collections électroniques. (1998). *Politiques et directives relatives aux publications électroniques diffusées en réseau. Bibliothèque nationale du Canada*. Adresse Url: http://epe.lac-bac.gc.ca/100/200/301/nlc-bnc/networked_epubs-ef/9/8/index-f.html (Page consultée le: 28 décembre 2006).
- 3- IM Forum Internet and Intranet Working Group. (1999). *Approche à la gestion de l'information sur Internet et Intranet pour assurer l'accès à long terme et la reddition de comptes*. Ottawa: Office of Government Records. National Archives of Canada. 75 p. Adresse Url: <http://web.archive.org/web/20091025070625/http://www.imforumgi.gc.ca/archives/groups-groupes/iwgt-rii/implement/implement-eng.pdf> (Page consultée le: 23 novembre 2009).

Document complémentaire

- BAC - Bibliothèque et Archives Canada. (2007-2008). *Government of Canada web archive - FAQ*. [En ligne]. Adresse Url: <http://www.collectionscanada.gc.ca/webarchives/faq/index-e.html> (Page consultée le: 23 avril 2007)

University of Toronto (Service d'archives universitaires)

- 4- University of Toronto. Archives and records management services. (2006). *Electronic records: guidelines for archiving Web sites*. [En ligne]. Adresse Url: <http://utarms.library.utoronto.ca/university-administrators/managing-records/electronic-records> (Page consultée le: 27 janvier 2009).

Documents complémentaires

- University of Toronto Libraries. (2009). *Archiving the web*. [En ligne]. Adresse Url: <http://www.archives.gov/records-mgmt/initiatives/web-tech.html> (Page consultée le: 27 janvier 2009).
- University of Toronto. Archives and records management services. (2009). *Web archives*. [En ligne]. Adresse Url: <http://utarms.library.utoronto.ca/researchers/web-archives> (Page consultée le: 27 janvier 2009).

ÉTATS-UNIS

NARA (Archives nationales)

- 5- NARA - National Archives and Records Administration. (2005). *NARA guidance on managing web records*. 28 p. Adresse Url: <http://www.archives.gov/records-mgmt/pdf/managing-web-records-index.pdf> (Page consultée le: 27 janvier 2009).

Documents complémentaires

- NARA - National Archives and Records Administration. ([s.d.]a). *Implications of recent web technologies for NARA Web guidance*. [En ligne]. Adresse Url: <http://www.archives.gov/records-mgmt/initiatives/web-tech.html> (Page consultée le: 23 avril 2007).
- NARA - National Archives and Records Administration. ([s.d.]b). *Tips for scheduling potentially permanent web content records*. [En ligne]. Adresse Url: <http://www.archives.gov/records-mgmt/publications/web-tips.pdf> (Page consultée le: 23 avril 2007).

AUSTRALIE

National Archives of Australia, Australian government, National Library of Australia (Archives nationales, archives gouvernementales, bibliothèque nationale)

- 6- Australian Government. Information Management Office. (2004). *Archiving web resources. Better practice checklist - 7: practical guide for effective use of new technologies in government*. Archiving web resources. 12 p. Adresse Url: <http://web.archive.org/web/20090614181325/http://www.finance.gov.au/e-government/better-practice-and-collaboration/better-practice-checklists/docs/BPC7.pdf> (Page consultée le: 23 avril 2007).
- 7- NAA - National Archives of Australia. (2001a). *Archiving web resources: a policy for keeping records of web-based activity in the Commonwealth government*. Government recordkeeping standards and policy. 18 p. Adresse Url: http://www.naa.gov.au/Images/archweb_policy_tcm2-902.pdf (Page consultée le: 11 octobre 2005).
- 8- NAA - National Archives of Australia. (2001b). *Archiving Web Resources: guidelines for keeping records of web-based activity in the Commonwealth of Australia*. Government Recordkeeping Standards and Policy. 41 p. Adresse Url: http://www.naa.gov.au/Images/archweb_guide_tcm16-47165.pdf (Page consultée le: 11 octobre 2005).
- 9- NAA - National archives of Australia. ([s.d.]). *Archiving websites: advice and policy statement*. [En ligne]. Adresse Url: http://www.naa.gov.au/Images/Archiving%20websites%20advice%20and%20policy%20statement_tcm2-902.pdf (Page consultée le: 27 janvier 2009).
- 10- NLA - National Library of Australia. (2008). *Online australian publications: a selection guidelines for archiving and preservation by the National Library of Australia*. [En ligne]. Adresse Url: <http://pandora.nla.gov.au/selectionguidelines.html> (Page consultée le: 27 janvier 2009).
- 11- NLA - National Library of Australia et NAA - National Archives of Australia. (2002). *Keeping government publications online: a guide for commonwealth agencies*. 6 p. Adresse Url: <http://web.archive.org/web/20090709110933/http://www.nla.gov.au/guidelines/govpubs.pdf> (Page consultée le: 11 octobre 2005).

Documents complémentaires

- Australian Government. ([s.d.]). *Web guide: archiving websites*. [En ligne]. Adresse Url: <http://webguide.gov.au/recordkeeping/archiving-a-website/> (Page consultée le: 23 avril 2007).
- Australian Government. Department of finance and deregulation. Australian government information management office. (2007). *Guide to minimum Website standards - electronic record keeping and archiving*. [En ligne]. Adresse Url: <http://www.agimo.gov.au/archive/mws/archiving.html> (Page consultée le: 23 avril 2007).

Australie (Melbourne) - University of Melbourne (Service d'archives universitaires)

- 12- The University of Melbourne. (2010). *Web archiving policy*. 4 p. Adresse Url: http://www.unimelb.edu.au/records/pdf/94027_Web_Archiving_Policy_August_2009_%28current%29.pdf (Page consultée le: 27 janvier 2009).
- 13- The University of Melbourne. (2008g). *Web archiving collection management plan: web archiving @ the University of Melbourne*. [En ligne]. Adresse Url: <http://www.unimelb.edu.au/records/web->

archiving/criteria.html (Page consultée le: 27 janvier 2009).

Documents complémentaires

- The University of Melbourne. (2008a). *Overview of the program*. [En ligne]. Adresse Url: <http://www.unimelb.edu.au/records/web-archiving/overview.html> (Page consultée le: 27 janvier 2009).
- The University of Melbourne. (2008b). *Advanced systems access*. [En ligne]. Adresse Url: <http://www.unimelb.edu.au/records/web-archiving/advanced.html> (Page consultée le: 27 janvier 2009).
- The University of Melbourne. (2008c). *Technology*. [En ligne]. Adresse Url: <http://www.unimelb.edu.au/records/web-archiving/technology.html> (Page consultée le: 27 janvier 2009).
- The University of Melbourne. (2008d). *How do I notify the web archiving program about missing pages or want to register a 'soon to be' decommissioned web site? URL: .* [En ligne]. Adresse Url: <http://www.unimelb.edu.au/records/web-archiving/notify.html> (Page consultée le: 27 janvier 2009).
- The University of Melbourne. (2008e). *Frequently Asked Questions*. [En ligne]. Adresse Url: <http://www.unimelb.edu.au/records/web-archiving/faqs.html> (Page consultée le: 27 janvier 2009).
- The University of Melbourne. (2008f). *Compliance with PROV Web archiving*. [En ligne]. Adresse Url: <http://www.unimelb.edu.au/records/web-archiving/prov.html> (Page consultée le: 27 janvier 2009).
- The University of Melbourne. (2011a). *How do I search the University's web archive?* [En ligne]. Adresse Url: <http://www.unimelb.edu.au/records/web-archiving/search.html> (Page consultée le: 27 janvier 2009).
- The University of Melbourne. (2011b). *Web archiving @ the University of Melbourne*. [En ligne]. Adresse Url: <http://www.unimelb.edu.au/records/web-archiving/> (Page consultée le: 27 janvier 2009).
- The University of Melbourne. (2011c). *How do I search the University's web archive? Access the list*. [En ligne]. Adresse Url: http://www.unimelb.edu.au/records/web-archiving/docs/131470_List_of_archived_University_websites.pdf (Page consultée le: 27 janvier 2009).

Australie (Tasmanie) - University of Tasmania (Service d'archives universitaires), et **Archives Office of Tasmania** (Archives provinciales)

- 14- Archives Office of Tasmania. (2009). *State records guideline No. 15: recordkeeping strategies for websites and web pages*. [En ligne]. Adresse Url: http://www.linc.tas.gov.au/_data/assets/pdf_file/0007/334978/State_Records_Guideline_No_15_Revised.pdf (Page consultée le: 27 janvier 2009).
- 15- State Library of Tasmania. (2011). *Guidelines for selecting, archiving and preserving websites pertinent to Tasmanian government information and cultural heritage*. [En ligne]. Adresse Url: <http://odi.statelibrary.tas.gov.au/About/selpolicy.asp> (Page consultée le: 27 janvier 2009).
- 16- University of Tasmania. (2010c). *Records management guidelines*. 38 p. Adresse Url: http://www.it.utas.edu.au/_data/assets/pdf_file/0014/32513/Records-Management-Guidelines.pdf (Page consultée le: 27 janvier 2009).

Documents complémentaires

- University of Tasmania. (2010a). *Keep your content up to date*. [En ligne]. Adresse Url: <http://www.web-services.utas.edu.au/resources/site-management/keep-your-content-up-to-date> (Page consultée le: 27 janvier 2009).
- University of Tasmania. (2010b). *Archive your website*. [En ligne]. Adresse Url: <http://www.web-services.utas.edu.au/resources/site-management/archive-your-website> (Page consultée le: 27 janvier 2009).

Australie (Queensland) - Queensland State Archives (Archives provinciales)

- 17- Queensland State Archives. (2006a). *Managing records of online resources and services guideline*. 39 p. Adresse Url: <http://www.archives.qld.gov.au/downloads/Policies/OnlineResourcesGuideline.pdf> (Page consultée le: 23 avril 2007).
- 18- Queensland State Archives. (2006b). *Managing records of online resources and services policy*. 13 p. Adresse Url: <http://www.archives.qld.gov.au/downloads/Policies/OnlineResourcesPolicy.pdf> (Page consultée le: 23 avril 2007).

- 19- Queensland State Archives. (2008). *Identifying a public record in the electronic environment*. [En ligne]. Adresse Url: http://www.archives.qld.gov.au/Recordkeeping/GRKDownloads/Documents/identifying_pub_record_e_e_nviron_200503.pdf (Page consultée le: 27 janvier 2009).

Documents complémentaires

- Queensland Government. Department of Public Works. (2006). *Best practice guide to recordkeeping*. 66 p. Adresse Url: <http://www.archives.qld.gov.au/downloads/IS40BestPracticeGuidetoRKv1.03.00.pdf> (Page consultée le: 23 avril 2007).
- Queensland State Archives. (2006c). *Web content management systems and recordkeeping*. [En ligne]. Adresse Url: <http://www.archives.qld.gov.au/Recordkeeping/GRKDownloads/Documents/WCMSRecords.pdf> (Page consultée le: 23 avril 2007).

Australie (Victoria) - Public records office of Victoria (Archives provinciales)

- 20- Public Record Office Victoria. (2007). *Advice 20a: web-generated records, advice to victorian agencies PROA 07/20*. : Public Record Office Victoria 22 p. Adresse Url: http://www.prov.vic.gov.au/records/Web_Advice/PROVRMAdvice20a.pdf (Page consultée le: 27 janvier 2009).
- 21- Public Record Office Victoria. (2008). *Advice 20b: technical issues for capturing web-generated records. Advice to victorian agencies PROA 07/20*. [En ligne]. Adresse Url: http://www.prov.vic.gov.au/records/Web_Advice/PROVRMAdvice20b.pdf (Page consultée le: 27 janvier 2009).

Australie (New South Wales) - Government New South Wales (Archives provinciales)

- 22- NSW - State Records Authority of New South Wales. (2009a). *Guideline 20 - Keeping web records*. [En ligne]. Adresse Url: <http://www.records.nsw.gov.au/recordkeeping/government-recordkeeping-manual/guidance/guidelines/guideline-20> (Page consultée le: 27 janvier 2009).
- 23- NSW - State Records Authority of New South Wales. (2009b). *Guideline 24 - Records management and web 2.0*. [En ligne]. Adresse Url: <http://www.records.nsw.gov.au/recordkeeping/government-recordkeeping-manual/guidance/guidelines/guideline-24> (Page consultée le: 27 janvier 2009).
- 24- NSW - State Records Authority of New South Wales. ([s.d.]). *Guideline 22 - Managing digital records: 5.4 Managing web records (Guideline 22)*. [En ligne]. Adresse Url: <http://www.records.nsw.gov.au/recordkeeping/government-recordkeeping-manual/guidance/guidelines/guideline-22/5-4-managing-web-records> (Page consultée le: 27 janvier 2009).

Nouvelle-Zélande - Archives New Zealand (Archives nationales)

- 25- Archives New Zealand. (2009). *Guide to managing web records*. [En ligne]. Adresse Url: <http://archives.govt.nz/advice/continuum-resource-kit/continuum-publications-html/g20-guide-managing-web-records> (Page consultée le: 27 janvier 2009).
- 26- Archives New Zealand. ([s.d.]). *Decommissioning websites: factsheet F23*. [En ligne]. Adresse Url: http://archives.govt.nz/sites/default/files/F23_-_Final_Version_of_Decommissioning_Websites_factsheet.pdf (Page consultée le: 27 janvier 2009).
- 27- New Zealand Government. ([s.d.]). *Web standards: Managing web information as records and archiving websites*. [En ligne]. Adresse Url: <http://www.webstandards.govt.nz/managing-web-information-as-records-and-archiving-websites-2/> (Page consultée le: 27 janvier 2009).

ROYAUME-UNI

The National Archives (Archives nationales)

- 28- Public records office. (2001). *Managing web resources: management of electronic records on websites and intranets: an ERM toolkit*. Surrey. 38 p. Adresse Url: http://wayback.archive.org/web/*/http://www.nationalarchives.gov.uk/documents/website_toolkit.pdf (Page consultée le: 11 octobre 2007).
- 29- TNA - The National Archives. (2006). *Operational selection policy OSP27: the selection of government websites*. 8 p. Adresse Url: <http://www.nationalarchives.gov.uk/documents/osp27.pdf> (Page consultée le: 27 janvier 2009).
- 30- TNA - The National Archives. (2009a). *Web standards & guidelines: Archiving websites*. 13 p. Adresse Url: <http://coi.gov.uk/guidance.php?page=240> (Page consultée le: 12 juin 2011).
- 31- TNA - The National Archives. (2011). *Operational selection policy 27: UK central government websites*. 8 p. Adresse Url: <http://www.nationalarchives.gov.uk/documents/draft-operational-selection-policy-27.pdf> (Page consultée le: 27 janvier 2009).
- 32- TNA - The National Archives. ([s.d.]a). *Information on web archiving*. [En ligne]. Adresse Url: <http://www.nationalarchives.gov.uk/webarchive/information.htm> (Page consultée le: 27 janvier 2009).

Document complémentaire

- TNA - The National Archives. (2009b). *Managing URLs*. [En ligne]. Adresse Url: <http://coi.gov.uk/documents/guidance/managing-urls-v1-0.pdf> (Page consultée le: 27 janvier 2009).
- TNA - The National Archives. ([s.d.]b). *Information for webmasters*. [En ligne]. Adresse Url: <http://www.nationalarchives.gov.uk/webarchive/webmasters.htm> (Page consultée le: 27 janvier 2009).

Annexe 3 : Grille d'analyse des politiques d'archivage du Web organisationnel

IDENTIFICATION DE LA FICHE D'ANALYSE	
Type d'analyse	Analyse – Politique AW
Titre de la fiche	AW – Politique – Titre abrégé du document
Date de création de la fiche	Date de création de la fiche d'analyse.
I. DESCRIPTION DU DOCUMENT	
I.1 Titre du document analysé	Indiquer le titre du document tel qu'affiché sur la page de couverture du document ou sur la première page Web qui le constitue.
I.2 Type du document	Indiquer s'il s'agit d'une ligne directrice, un guide de mise en œuvre, une recommandation technique, une procédure, une note de service, etc.
I.3 Date de création ou de MAJ	Indiquer la date de création ou de mise à jour du document.
I.4 Auteur(s)	Indiquer l(es) auteur(s) du document.
I.5 Catégorie d'organisme	Identifier la catégorie de l'organisme responsable du contenu du document analysé.
I.6 Source	Spécifier s'il s'agit d'un document disponible en ligne, d'un document recueilli lors des entrevues, d'un document communiqué par un professionnel de l'information ou autre.
I.7 Lien avec autre(s) document(s)	Indiquer les renvois signalés dans le document analysé vers d'autres documents et spécifier la nature de ces derniers (ex. : lois, autres politiques, standards, normes ou annexes à considérer). S'il s'agit d'un document complémentaire, décrire son contenu dans le champ Commentaires et intégrer les éléments analysés dans les champs de la fiche.
II. ANALYSE DU DOCUMENT	
II.1 Portée	Spécifier s'il s'agit d'une politique organisationnelle (politique mise en œuvre dans un contexte précis) ou d'un cadre de référence archivistique. Ce champ complète les informations indiquées dans le champ I.5 Catégorie d'organisme. Il nous aide à distinguer entre les politiques mises en œuvre dans le cadre de projets d'archivage initiés par des archives nationales (ex. : BAC, 2007-2008), des politiques que ces institutions publient afin d'orienter les organismes à établir leur propre politique

	d'archivage (ex. : BAC, 1999).
II.2 Lectorat visé	Spécifier le lectorat visé par le document analysé (ex. : archiviste, webmestre). Ce champ nous aide à identifier à qui s'adresse la politique et les distinguer des intervenants qui doivent être impliqués dans le processus d'archivage. Cette donnée nous sert principalement à décrire le document analysé.
II.3 Contexte	Indiquer le contexte réglementaire ou normatif dans lequel s'inscrit le document analysé (ex. : mettre en application une loi, respecter un standard documentaire).
II.4 Système(s) visé(s)	Indiquer la liste des systèmes Web visés par le document analysé (ex. : intranet, extranet, site Web public).
II.5 Définitions	Transcrire toute définition utile à la compréhension des concepts clés de cette recherche (ex. : archives Web, site Web, portrait d'un site Web).
II.6 Intervenants clés	Indiquer la liste des acteurs devant être impliqués dans la gestion et la conservation des archives Web tel que décrit dans le document analysé.
II.7 Rôles et responsabilités	Spécifier les rôles et responsabilités de chacun des intervenants clés identifiés.
II.8 Motifs	Identifier les raisons pour lesquelles la capture et la conservation des documents/sites Web ont été justifiées (ex. : continuité des affaires de l'organisation, obligations légales).
II.9 Unité(s)/contenu(s) d'archivage	Indiquer les composantes Web à capturer (ex. : l'ensemble d'un site Web, une section du site, des documents dans un format spécifique, de la documentation des processus liés à la gestion du site ou de la technologie qui le soutient).
II. 10 Statut des contenus capturés	Spécifier le statut des contenus capturés/archivés : archives définitives ou archives intermédiaires, exemplaire principal ou exemplaire secondaire.
II.11 Critères de sélection	Relever l'ensemble des critères de sélections des contenus archivés (ex. : caractère unique du document, niveau de profondeur, lien avec une activité, fréquence de mise à jour).
II.12 Mode(s) d'archivage	Spécifier si l'archivage fait/recommandé est de nature manuelle, automatique, réalisé via un contrat avec un tiers-archiviste, par dépôt volontaire, par versement, ou autre.
II.13 Approche(s) d'archivage	Spécifier l(es) approche(s) recommandée(s)/appliquée(s) : intégrale, événementielle, sélective, collaborative, centrée sur les objets, centrée sur les événements/changements, intégrée, ou autre.
II.14 Technique(s)	Spécifier la(es) technique(s) recommandée(s)/appliquée(s) : archivage à la source, archivage à distance, archivage des transactions, archivage en ligne, documentation des changements, etc.

II.15 Outil(s) d'archivage	Indiquer la liste des outils utilisés/recommandés pour réaliser la capture des contenus à archiver.
II.16 Outil(s) archivistique(s)	Indiquer la liste des outils archivistiques à développer ou à utiliser dans le cadre de la gestion et l'archivage des systèmes Web.
II.17 Principes directeurs	Indiquer les principes qui régissent le processus d'archivage des systèmes Web qu'il s'agisse de principes généraux ou principes archivistiques.
II.17.1 Principes généraux	Les principes généraux concernent la gestion du processus d'archivage ou sont liés à la conception et à la gestion des systèmes Web (ex. : adopter des formats standards et des technologies ouvertes).
II.17.2 Principes archivistiques	Les principes archivistiques concernent la gestion du cycle de vie des archives Web (ex. : bien décrire les documents diffusés en leur associant des métadonnées pertinentes, conserver les sites Web retirés en ligne dans un dépositaire de site local).
II.18 Facteurs à considérer	Identifier la liste des facteurs à prendre en compte lors de la définition ou la mise en œuvre du processus d'archivage (ex. : objectif du système Web, complexité du système Web, type du système Web, fréquence de mise à jour).
II.19 Stratégie(s) de préservation à long terme	Indiquer la(es) stratégie(s) de préservation appliquée(s) ou recommandée(s) (ex. : émulation, migration).
II.20 Support(s) d'archivage	Indiquer les supports utilisés ou recommandés pour l'archivage des systèmes Web.
COMMENTAIRES	
Indiquer toute autre observation utile. Décrire le contenu des documents complémentaires associés au document analysé.	

Annexe 4 : Grille d'observation des SIW

IDENTIFICATION DE LA FICHE D'ANALYSE			
Titre de la fiche	Analyse SIW - Titre attribué au niveau de l'arborescence Scrapbook (voir Annexe 3, p. xxii)		
Code du SIW	Indiquer le code du SIW observé.		
Date (début)	Indiquer la date de début de l'observation.		
Date (fin)	Indique la date de la fin de l'observation.		
CARACTÉRISTIQUES DU SIW			
Critères	Description	Réponse à inscrire	Critères d'évaluation
I. INFORMATION GÉNÉRALE			
I.1 Catégorie du SIW	Est-ce qu'il s'agit d'un site gouvernemental, universitaire, associatif, etc.?	Décrire	
I.2 Taille du site	Est-ce qu'il s'agit d'un petit site, d'un site de taille moyenne ou d'un site volumineux? La taille du site sera déterminée en fonction du nombre de pages accessibles depuis la page d'accueil : < 15 pages = petit site; entre 15-25 pages = site de taille moyenne; > 25 pages = site volumineux.	Liste	S'il s'agit d'un petit site : examiner toutes les pages. S'il s'agit d'un site moyen ou volumineux : il faut aussi examiner dans une première étape tous les liens (pages) pointés à partir de la page d'accueil, de même que les pages du deuxième niveau de profondeur. Ensuite, il faut sélectionner les pages les plus représentatives (diversité du contenu, variété des genres, caractère original du contenu, format particulier, fonctionnalités, etc.).
I.3 Date de création	Quelle est la date de création du site Web?	Date	

I.4 Date de mise à jour	Date(s) de mise(s) à jour du SIW.	Date	
I.5 Usages/objectifs	Quels sont les objectifs du site Web : diffusion d'informations, dépôt de données, offre de services en ligne, autre.	Liste	
I.6 Créateur(s) des contenus	Qui sont les personnes responsables de la production du contenu du site Web (contenu en HTML, documents diffusés, etc.). Est-ce que les internautes peuvent enrichir ou modifier le site Web?	Si l'information est disponible sur le site, décrire	Information à vérifier et/ou compléter lors des entrevues.
I.7 Responsable(s) du SIW	Qui sont les personnes responsables de la gestion du site Web : webmestre(s), gestionnaire(s) d'information (bibliothécaire, archiviste), un développeur privé.	Si l'information est disponible sur le site, décrire	Information à vérifier et/ou compléter lors des entrevues.
I.8 Gestionnaire de l'infrastructure technologique du SIW	La personne responsable de la gestion du site Web est la même personne responsable de la création de son contenu?	oui-non- N/A	Information à vérifier et/ou compléter lors des entrevues.
I.9 Clientèle visée	À quelle(s) clientèle(s) s'adresse le site Web : grand public, abonné, partenaires, employés?	Décrire	
II. TECHNOLOGIE			
II.1 Type du SIW	Est-ce que le site est statique ou dynamique?	Liste	
II.2 Nom(S) du domaine (DNS)	Quelles sont les principales adresses des serveurs qui hébergent le SIW?	Décrire	Vérifier dans la barre des adresses.
II.3 Protocole(s) de communication	Quels sont les protocoles de communication qu'utilise le SIW? HTTP, HTTPS, FTP, SOAP, etc.	Décrire	Vérifier dans la barre des adresses.

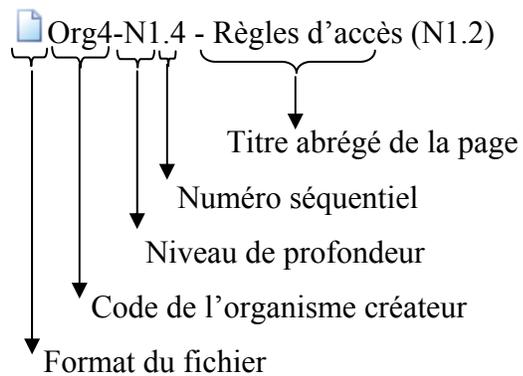
II.4 Navigateur(s)	Le SIW est optimisé pour quelle(s) version(s) de navigateur(s)?	Décrire	
II.5 Plugiciels	Est-ce que la visualisation de certains fichiers nécessite le téléchargement de plugiciels?	oui-non	
II.6 Langage de balisage	Quel langage de balisage est utilisé pour organiser le contenu du site? HTML, XHTML, XML, RDF, technologies émergentes (XUL, Xforms, STML, SVG, services Web)	Liste	Voir le code source.
II.7 Feuille de style	Quel langage est utilisé pour gérer la mise en forme du contenu du site? CSS, XSL, XSL-T, XSL-FO, autre	Liste	Voir le code source.
III. STRUCTURE			
III.1 Liens	Quels sont les types de liens présents sur le SIW? liens internes, liens externes, liens internes et externes	Liste	Mêmes critères que ceux appliqués pour la taille du site
III.1.1 Liens internes	Proportion (nombre de liens internes présents sur quatre pages différentes du SIW : pages représentatives des variétés des genres d'information disponibles à partir du SIW)	Décrire	
III.1.2 Liens externes	Proportion (nombre de liens externes présents sur quatre pages différentes du site offrant une certaine variété des genres d'information disponibles à partir du SIW, les mêmes testées pour les liens externes)	Décrire	
III.1.3 Caractéristiques des liens	Est-ce qu'il a une distinction claire entre liens internes et liens externes?	Décrire	

	Vers quels types d'information pointent les liens? pages HTML, images, documents téléchargeables, etc.?	Décrire	
	Les liens sont-ils absolus ou relatifs?	Liste	
III.4 Navigation	Quels sont les moyens de navigation disponibles sur le site? barres de navigation, onglets, carte du site, index, moteur de recherche, liens, etc.	Décrire	
III.5 Multifenêtrage	Est-ce que le site est structuré en cadres (<i>frames</i>)?	oui-non	Voir code source.
Si oui :			
III.5.1 Version Noframe	Est-ce qu'il existe une version sans cadres (<i>no-frame</i>) du site?	oui-non	Voir code source.
III.6 Version imprimable	Est-ce qu'il existe une version imprimable du site?	oui-non	
IV. CONTENU			
IV.1 Langue	Le site est-il unilingue, bilingue ou multilingue?	Liste	
IV.2. Composantes			
IV.2.1 Ressources constituant le site Web	Quelles sont les principales composantes des différentes pages Web du site? du HTML, des images, du son, etc.	Décrire	
IV.2.2 Ressources accessibles depuis le SIW	Y a-t-il des ressources accessibles depuis le site Web mais consultables avec une interface particulière (flux audio, vidéo)?	oui-non	
IV.2.3 Ressources consultées séparément	Y a-t-il des ressources accessibles sur le site mais conçues pour être consultées séparément (Word, PDF, etc.)?	oui-non	

IV.3 Formats	Quels sont les principaux formats de fichiers composant le site Web?	Décrire	
IV.4 Normalisation	Est-ce que les composantes du site sont présentées dans des formats standards?	Décrire	Se référer aux standards développés par le W3C et les services gouvernementaux (QC).
IV.5 Mise à jour	La mise à jour porte sur l'ensemble du site, quelques pages, certains éléments du contenu?	Décrire	Si l'information n'est pas disponible sur le site observé, comparer avec quelques versions archivées du site sur Internet Archive.
IV.6 Fréquence de mise à jour	Est-ce que la mise à jour du site est fréquente?	oui-non	Si l'information n'est pas disponible sur le site observé, comparer avec quelques versions archivées du site sur Internet Archive.
IV.7 Volume	Longueur de pages (écran d'impression) : > 2 pages; = 2 pages; < 2 pages	Liste	
	Niveau de profondeur (nombre de cliques pour consulter une information) : > 3 clics, = 3 clics < 3 clics	Liste	Tester en cliquant sur différents liens à partir de la page d'accueil.
	Est-ce que le niveau de profondeur est constant?	oui-non	
V. INTERACTIVITÉ			
V.1 Personnalisation	Est-ce qu'il est possible de personnaliser la présentation du site Web par profil d'utilisateurs?	oui-non	
V.2 Formulaire	Est-ce que le site comprend des formulaires en ligne?	oui-non	
V.3 Moteur de recherche	Est-ce que le site comprend un moteur de recherche?	oui-non	

V.4 Interaction	Est-ce que le site comprend une rubrique d'aide en ligne? Un forum de discussion?	oui-non	
V.5 Flux d'information	Est-ce que le site Web offre un babillard d'information, un fil RSS?	oui-non	
V.6 Accès et consultation d'une BDD	Est-ce que le site permet la consultation et l'interrogation d'une BDD (catalogue, dossier en ligne, etc.)?	oui-non	
COMMENTAIRES			
Indiquer toute autre observation utile.			

Annexe 5 : Arborescence Scrapbook – règles de nommage des pages capturées



Exemple :

 Org4-N2.2 - Politique Acquisition (N1.5)

⇒ Il s'agit d'un document PDF publié sur le site de O4 (code du quatrième site d'étude). Le document est situé au deuxième niveau de profondeur à partir de la page d'accueil (N2). La page a été repérée en naviguant au niveau de la page N1.5 (titre complet de la page : Org4-N1.5 – Politiques).

Annexe 6 : Grille d'observation des archives Web

IDENTIFICATION DE LA FICHE D'ANALYSE	
Type d'analyse	Observation des archives Web – Code de l'organisation.
Titre de la fiche	Analyse AW - Titre du document analysé.
Date de création de la fiche	Mentionner la date à laquelle la fiche a été créée.
I. ANALYSE DU DOCUMENT	
I.1 Titre du document analysé	Indiquer le titre du document observé tel qu'indiqué sur la page Web. Si cette information n'est pas disponible dans le corps même du document observé, la rechercher au niveau des menus de navigation, de l'URL ou dans le code source de la page observée.
I.2 Date de création du document	Indiquer la date de création du document observé. Si elle n'est pas indiquée, le mentionner.
I.3 Date de MAJ du document	Indiquer la date de mise à jour du document observé. Si elle n'est pas indiquée, le mentionner.
I.4 URL du document	Saisir l'adresse URL du document observé. Cette information va déterminer si une des règles d'édition Web est respectée. Elle facilite aussi la détermination du type ou de la catégorie du document au cas où le titre de ce dernier n'est pas indiqué.
I.5 Créateur du document	Indiquer l'auteur de la page Web observée. Cette information permet de déterminer la provenance du document (document créé, document reçu ou document de référence). Si cette information n'est pas indiquée, le mentionner.
I.6 Liens existants sur le document	Décrire la nature des liens présents sur le document observé (liens internes, liens externes) et les quantifier.
I.7 Format du document	Indiquer le format du document observé. Cette information sert entre autres à déterminer le degré d'emploi des formats standards vs les formats propriétaires.
I.8 Composantes du document	Identification de la liste des objets documentaires qui composent le document observé et leur format.

I.9 Contenu du document	Description du contenu du document observé. Cette description servira à identifier la catégorie et le type du document (pouvoir établir un lien avec une activité ou un service de l'organisme).
I.10 Catégories du document	Identifier la catégorie du document observé : - Document de gestion : document produit ou reçu dans le cadre des activités administratives d'une organisation (Couture, 1999: 91). - Document de fonction : document produit ou reçu dans le cadre de la mission propre d'une organisation (Couture, 1999: 92).
I.11 Type du document	Identifier le type du document observé (ex. : rapport annuel, plan stratégique, formulaire, communiqué, etc.).
I.12 Références consultées	Indiquer la liste des outils et manuels archivistiques consultés pour déterminer la catégorie ou le type du document observé.
COMMENTAIRES	Indiquer toute autre observation utile.

Annexe 7 : Guide d'entrevue

GUIDE D'ENTREVUE

INTRODUCTION

Description de la recherche

Remerciements.

Objectifs de la recherche : (1) développer une meilleure compréhension de la typologie des archives Web et (2) examiner les pratiques organisationnelles en matière de gestion des archives produites, reçues ou diffusées grâce aux systèmes d'information Web (SIW) d'une organisation.

Archives Web : documents issus ou relatifs à un système d'information Web, produits dans le cadre des activités d'un organisme donné y compris le site Web, les documents diffusés sur ce site Web ainsi que la documentation relative à leur gestion.

Subventionnée par FQRSC

Présentation du déroulement de l'entrevue

Durée de l'entrevue.

Signature du formulaire de consentement.

Permission d'enregistrement de l'entrevue.

QUESTIONNAIRE

1. Information sur le répondant

1.a Quel est le titre du poste que vous occupez actuellement?

1.b Pourriez-vous décrire vos activités et responsabilités?

2. Caractéristiques du site Web public

2.a En quelle année a(ont) été créé(s) le(s) site(s) Web public(s) de votre organisation?

Depuis, a-t-il subi des refontes majeures (contenu, structure, technologies)?

2.b Pouvez-vous me décrire les caractéristiques techniques du (ou des) site(s) Web public(s) de votre organisation?

2.c Pourriez-vous me décrire les objectifs et les clientèles du (ou des) site(s) Web public(s) de votre organisation?

2.d Le SIW est sous la responsabilité de quelle unité, quels groupes de professionnels?

- Qui est responsable de la gestion globale du (ou des) site(s) Web public(s) de votre organisation?
- Qui est responsable de la gestion de l'infrastructure technologique du (ou des) site(s) Web public(s) de votre organisation?
- Qui est responsable de la création et de la mise à jour du contenu du (ou des) site(s) Web public(s) de votre organisation?
- Qui est responsable de la publication d'informations sur le(ou les) site(s) Web public(s) de votre organisation?
- Qui est responsable de l'archivage du (ou des) site(s) Web public(s) de votre organisation?

2.e Pourriez-vous, en quelques points, me décrire le processus de publication de l'information sur le SIW?

3. Actif documentaire du site Web public de l'organisation

3.a Quels sont les principaux types de documents que nous pourrions retrouver sur le (ou les) site(s) Web public(s) de votre organisation?

3.b Pourriez-vous m'indiquer la provenance des documents disponibles à partir du (ou des) site(s) Web public(s) de votre organisation?

3.c Sur quels critères ou caractéristiques vous vous basez pour identifier les archives Web?

3.d Quels documents accessibles à partir du SIW peuvent être localisés (enregistrés ou archivés) dans d'autres systèmes documentaires (papier, BDD, serveur local)? Pourriez-vous m'en donner quelques exemples?

3.e Quels documents sont uniquement disponibles à partir du SIW? Pourriez-vous m'en donner quelques exemples?

4. Démarche d'archivage des SIW de l'organisation

4.a Pourriez-vous me décrire la stratégie d'archivage du site Web public de votre organisation?

- Quelles composantes du site Web public sont archivées?
- À quelle fréquence?
- Dans quel(s) format(s)?

4.b Quels outils et technologies sont utilisés pour implanter cette stratégie?

4.c Pour quelles raisons a été implanté cette stratégie?

4.d Les choix actuels vous semblent adéquats?

4.e Les archives Web capturées sont-elles exploitées? De quelle manière et dans quel format?

4.f La gestion et la conservation des archives Web diffèrent-elles de la gestion et de la conservation des autres documents numériques? Si oui, pourriez-vous préciser en quoi?

5. Outils de gestion utilisés ou développés

5.a Est-ce que la démarche d'archivage de votre site Web public est documentée? Si oui, quels outils sont développés et/ou utilisés pour guider ou soutenir la démarche d'archivage du Web dans votre organisation?

5.b Comment cette démarche est intégrée dans la stratégie globale de la gestion de l'information dans votre organisation?

5.c Quelle est la portée de ces outils? (SIW et archives Web visés)

6. Difficultés rencontrées

6.a Quelles sont les principales difficultés rencontrées lors de la mise en œuvre de la démarche d'archivage du Web dans votre organisation?

- Niveau technologique et technique
- Niveau gestion documentaire

6.b Les ressources actuellement mobilisées sont-elles suffisantes?

7. Travaux futurs

7. Quelles sont les prochaines étapes de développement des activités d'archivage du Web dans votre organisation?

8. Autres SIW

8.a Quels autres SIW retrouve-t-on dans votre organisation?

8.b Quels sont les principaux types de documents accessibles à partir de ces SIW?

8.c Parmi ceux-ci (SIW et documents), lesquels sont archivés (en tout ou en partie)?

8.d Est-ce que les méthodes, approches et techniques utilisées diffèrent de celles utilisées pour les sites Web publics de votre organisation? Si oui, en quoi sont-elles différentes?

9. Information démographique

9.a Quelle est votre formation quant au niveau d'études postsecondaires, au programme d'étude et au domaine d'étude?

9.b Pourriez-vous me donner un aperçu de votre cheminement de carrière (postes occupés, responsabilités).

9.c Quel est le nombre d'années d'expérience de travail ?

- a) Dans le poste actuel
- b) Avec les systèmes d'information Web
- c) Dans la gestion et l'archivage du Web
- d) En tout

9.d Avez-vous reçu une formation particulière pour acquérir les compétences nécessaires à l'archivage du Web? Si oui, dans quel cadre? Si non, comment avez-vous développé vos compétences?

REMERCIEMENTS ET FIN DE L'ENTREVUE

Démonstration, si possible.

Copie de la documentation pertinente.

Coordonnées d'autres intervenants clés.

Sommaire des résultats.

Remerciements et fin de l'entrevue.

Annexe 8 : Formulaire de consentement

FORMULAIRE DE CONSENTEMENT

[Entrevues en personne]

Titre de la recherche :

Archivage du Web organisationnel dans une perspective archivistique.

Responsable de la recherche :

Aïda Chebbi, candidate au doctorat
École de bibliothéconomie et des sciences de l'information
Université de Montréal
C.P. 6128, succursale Centre-ville
Montréal QC H3C 3J7
Téléphone : (514) 887 2432

Nom de la directrice de recherche :

Louise Gagnon-Arguin, professeure associée, Université de Montréal

Nom de la co-directrice de recherche :

Christine Dufour, professeure adjointe, Université de Montréal

A) RENSEIGNEMENTS AUX PARTICIPANTS

1. Objectifs de la recherche

Notre recherche a pour but de développer une meilleure compréhension de la nature des archives Web et de documenter les politiques et les pratiques actuelles d'archivage du Web organisationnel. Plus précisément, cette recherche vise à (1) dresser un portrait des archives Web diffusées, créées ou reçues grâce aux sites Web publics d'organismes québécois, (2) analyser leurs politiques d'archivage et (3) décrire les pratiques d'archivage en usage dans ces organismes.

2. Participation à la recherche

La participation à cette recherche consiste à répondre oralement à une série de questions relatives aux types, caractéristiques et pratiques de gestion et de conservation des archives Web diffusées, créées ou reçues grâce au site Web public d'une organisation. Répondre à ces questions devrait requérir environ 90 minutes de votre temps.

En participant à cette recherche, vous ne courez pas de risques ou d'inconvénients particuliers. Le projet permettra aux participants d'apporter une contribution réelle à la compréhension des pratiques organisationnelles en matière de gestion des archives Web.

3. Confidentialité

Les renseignements que vous nous donnerez demeureront confidentiels. Chaque participant à la recherche se verra attribuer un numéro de code et seule la chercheuse aura accès à la liste des participants et des numéros de codes qui leur auront été attribués. De plus, les renseignements et les informations recueillies seront conservés dans un classeur sous clé situé dans un bureau fermé. Aucune information permettant de vous identifier d'une façon ou d'une autre ne sera publiée. Ces renseignements personnels seront détruits 7 ans après la fin du projet. Seules les données ne permettant pas de vous identifier seront conservées après cette date, le temps nécessaire à leur utilisation.

4. Droit de retrait

Votre participation est entièrement volontaire. Vous êtes donc libre d'accepter ou de refuser d'y participer. De même, vous pouvez décider de ne pas répondre à une question. Vous êtes aussi libre de vous retirer en tout temps, sans préjudice et sans devoir justifier votre décision. Si vous vous retirez de la recherche, les renseignements qui auront été recueillis au moment de votre retrait seront détruits.

5. Diffusion des résultats

Sur demande des participants, les résultats de cette recherche seront transmis une fois que l'analyse des données aura été effectuée.

Les conclusions générales de cette recherche pourraient être utilisées dans le cadre de publications ou de communications.

B) CONSENTEMENT

Je déclare avoir pris connaissance des informations ci-dessus, avoir obtenu les réponses à mes questions sur ma participation à la recherche et comprendre le but de cette recherche.

Après réflexion, je consens librement à prendre part à cette recherche. Je sais que je peux me retirer en tout temps sans préjudice et sans devoir justifier ma décision.

Signature : _____ Date : _____

Nom du participant : _____

Je déclare avoir expliqué le but, la nature, les avantages et les inconvénients de l'étude et avoir répondu aux questions posées.

Signature du chercheur : _____ Date : _____

Nom du chercheur : _____

Toute plainte relative à votre participation à cette recherche peut être adressée à l'ombudsman de l'Université de Montréal, au numéro de téléphone (514) 343-2100.

Annexe 9 : Certificat d'éthique



COMITÉ D'ÉTHIQUE DE LA RECHERCHE DE LA FACULTÉ DES ARTS ET DES SCIENCES (CÉRFAAS)

CERTIFICAT D'ÉTHIQUE

Le Comité d'éthique de la recherche de la Faculté des arts et des sciences, selon les procédures en vigueur, a examiné le projet de recherche suivant :

« Archivage du Web organisationnel dans une perspective archivistique. »

Requérant : *Aïda Chebbi, étudiante au doctorat (Ph.D.),
École de bibliothéconomie et des sciences de l'information.*

Sous la direction de :

*Louise Gagnon-Arguin, Professeure associée,
et
Christine Dufour, Professeur adjointe,*

École de bibliothéconomie et des sciences de l'information

Le Comité a conclu que la recherche proposée respecte les règles d'éthique énoncées dans la « Politique sur la recherche avec des êtres humains » de l'Université de Montréal.

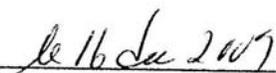
Tout changement anticipé au protocole de recherche doit être communiqué au CÉRFAAS qui devra en évaluer l'impact au chapitre de l'éthique.

Toute interruption prématurée du projet ou tout incident grave devra être immédiatement signalé au CÉRFAAS.

Un suivi annuel est exigé afin de maintenir la validité de ce certificat.



Deirdre Meintel, professeure titulaire
Présidente, Comité d'évaluation accélérée



Date de délivrance

Annexe 10 : Feuille de réponse

Entrevue # _____

Répondant :

Date de l'entrevue

Début : Fin :

GUIDE D'ENTREVUE – FEUILLE DE RÉPONSE

INTRODUCTION

Code du participant
Code de l'organisme
Date de l'entrevue
Heure de début
Heure de fin
Sexe du participant
Nom de l'intervieweur

Homme Femme

QUESTIONNAIRE

1. Information sur le répondant

1.a Quel est le titre du poste que vous occupez actuellement?

1.b Pourriez-vous décrire vos activités et responsabilités?

2. Caractéristiques du site Web public

2.a En quelle année a été créé le site Web public de votre organisation?

Depuis, a-t-il subi des refontes majeures (contenu, structure, technologies)?

2.b Pouvez-vous me décrire les caractéristiques techniques du (ou des) site(s) Web public(s) de votre organisation?

- Technologie de développement :

- Environnements virtuels :

- Interactivité :

- Fréquence de mise à jour :

2.c Pourriez-vous me décrire les objectifs et les clientèles du (ou des) site(s) Web public(s) de votre organisation? (si plusieurs sites Web publics, remplir fiches supplémentaires)

	Objectifs (à cocher)	Clientèle(s)
—	Diffusion de l'information sur l'organisation	
—	Interaction avec les internautes	
—	Offre d'applications et de services en ligne	
—	Autre :	

2.d Le SIW est sous la responsabilité de quelle unité, quels groupes de professionnels?

Responsabilités	Unité(s)	Groupe(s) de professionnel(s)
Gestion du SIW		
Gestion de l'infrastructure technologique du SIW		
Création et mise à jour du contenu du SIW		
Publication d'informations sur le SIW		
Archivage du SIW		

2.e Pourriez-vous, en quelques points, me décrire le processus de publication de l'information sur le SIW?

3. Actif documentaire du site Web public de l'organisation

3.a Quels sont les principaux types de documents que nous pourrions retrouver sur le (ou les) site(s) Web public(s) de votre organisation?

3.b Pourriez-vous m'indiquer la provenance des documents disponibles à partir du (ou des) site(s) Web public(s) de votre organisation?

3.c Sur quels critères ou caractéristiques vous vous basez pour identifier les archives Web?

3.d Quels documents accessibles à partir du SIW peuvent être localisés (enregistrés ou archivés) dans d'autres systèmes documentaires (papier, BDD, serveur local)? Pourriez-vous m'en donner quelques exemples?

3.e Quels documents sont uniquement disponibles à partir du SIW? Pourriez-vous m'en donner quelques exemples?

4. Démarche d'archivage des SIW de l'organisation

4.a Pourriez-vous me décrire la stratégie d'archivage du site Web public de votre organisation? (si plusieurs sites Web publics, remplir fiches supplémentaires)

Composantes archivées	Méthode	Fréquence	Format
Tout le site			
Sections particulières du site :			
Documents de gestion du site Web			
Documents publiés sur le site Web			
Documents créés grâce au site Web			
BDD qui alimente le contenu du site Web			
Formats particuliers de documents			
Applications et logiciels			
Autre			

4.b Quels outils et technologies sont utilisés pour implanter cette stratégie?

4.c Pour quelles raisons a été implanté cette stratégie?

4.d Les choix actuels vous semblent adéquats?

4.e Les archives Web capturées sont-elles exploitées? De quelle manière et dans quel format?

4.f La gestion et la conservation des archives Web diffèrent-elles de la gestion et de la conservation des autres documents numériques? Si oui, pourriez-vous préciser en quoi?

5. Outils de gestion utilisés ou développés

5.a Est-ce que la démarche d'archivage de votre site Web public est documentée? Si oui, quels outils sont développés et/ou utilisés pour guider ou soutenir la démarche d'archivage du Web dans votre organisation?

5.b Comment cette démarche est intégrée dans la stratégie globale de la gestion de l'information dans votre organisation?

5.c Quelle est la portée de ces outils? (SIW et archives Web visés)

6. Difficultés rencontrées

6.a Quelles sont les principales difficultés rencontrées lors de la mise en œuvre de la démarche d'archivage du Web dans votre organisation?

- Niveau technologique et technique

- Niveau gestion documentaire

6.b Les ressources actuellement mobilisées sont-elles suffisantes?

7. Travaux futurs

7. Quelles sont les prochaines étapes de développement des activités d'archivage du Web dans votre organisation?

8. Autres SIW

8.a Quels autres SIW retrouve-t-on dans votre organisation?

Intranet	Extranet	Autre

8.b Quels sont les principaux types de documents accessibles à partir de ces SIW?

Intranet	Extranet	Autre

8.c Parmi ceux-ci (SIW et documents), lesquels sont archivés (en tout ou en partie)?

8.d Est-ce que les méthodes, approches et techniques utilisées diffèrent de celles utilisées pour les sites Web publics de votre organisation? Si oui, en quoi sont-elles différentes?

9. Information démographique

9.a Quelle est votre formation quant au niveau d'études postsecondaires, au programme d'étude et au domaine d'étude?

Niveau d'étude	Programme d'étude	Domaine d'étude
Cégep général		
Cégep technique		
Universitaire 1 ^{er} cycle		
Universitaire 2 ^e cycle		
Universitaire 3 ^e cycle		
Autres		

9.b Pourriez-vous me donner un aperçu de votre cheminement de carrière (postes occupés, responsabilités).

9.c Quel est le nombre d'années d'expérience de travail ?

Nombre d'années d'expérience	Moins de 1 an	1 à 5 ans	6 à 10 ans	Plus que 10 ans
Dans le poste actuel				
Avec les systèmes d'information Web				
Dans la gestion et l'archivage du Web				
En tout				

9.d Avez-vous reçu une formation particulière pour acquérir les compétences nécessaires à l'archivage du Web? Si oui, dans quel cadre? Si non, comment avez-vous développé vos compétences?

FIN DE L'ENTREVUE

Annexe 11 : Protocole d'entrevue

PROTOCOLE D'ENTREVUE

[Entrevues en personne]

Ce guide a été conçu afin de rappeler à la chercheuse quelles tâches elle doit être amenée à faire et quelles questions doivent être posées à un moment ou l'autre au cours de l'entrevue.

1. Documents et matériel à apporter

- 1 protocole d'entrevue
- 2 formulaires de consentement (photocopiés recto/verso)
- Enregistreur numérique + (chargeur et piles)
- Clé USB en cas de communication de documents sous format numérique
- Document pour prise de notes

2. Description du projet et signature du formulaire de consentement

- Remerciements
- Décrire le projet et le déroulement de l'entrevue
- Lire ensemble le formulaire de consentement
- Questions ?
- Signer les deux exemplaires du formulaire (l'employé en conserve une copie)

3. Installation du matériel

- Installer l'enregistreur numérique (mettre le volume suffisamment fort, rapprocher le micro du répondant)
- Tester l'enregistreur numérique
- Mettre en marche l'enregistreur

4. Description de tâches

- Inviter l'employé à répondre aux différentes questions et à faire une démonstration de ses tâches, des systèmes et des produits si possible
- Demander à l'employé s'il possède de la documentation pertinente : règle de conservation détaillée, procédures d'archivage, etc. et s'il peut en laisser une copie

5. Fin de l'entrevue

- Remercier le répondant
- Faire une copie de sécurité des fichiers obtenus lors de l'entrevue, le cas échéant

6. Après l'entrevue

- Prise de notes « post- entrevue »
- Copie de sécurité de l'enregistrement
- Envoi de la lettre de remerciement

Annexe 12 : Lettre de sollicitation

Montréal, [date]
[Nom du participant], [Poste]
[Adresse]

Objet : Recherche doctorale sur les pratiques d'archivage du Web organisationnel dans une perspective archivistique

[Madame, Monsieur] [Nom du participant],

Malgré un usage étendu des technologies Web dans les organisations, les pratiques de capture et de conservation des archives produites dans un environnement Web, restent peu documentées et rarement étudiées dans un contexte organisationnel.

Par la présente, je souhaite solliciter votre participation à une recherche doctorale qui vise à mieux comprendre la nature des archives Web et à explorer les pratiques d'archivage des sites Web publics dans des établissements universitaires au Québec. Cette recherche est réalisée sous la supervision de Louise Gagnon-Arguin et Christine Dufour, professeures à l'École de bibliothéconomie et des sciences de l'information de l'université de Montréal. Elle bénéficie du soutien financier du Fonds Québécois de Recherche sur la Culture – Programme de bourse de doctorat de recherche.

Votre participation consistera en une entrevue d'une durée d'environ une heure et demie qui se déroulera à votre lieu de travail au moment qui vous convient le mieux. L'objectif est de colliger des informations sur les principaux types de documents diffusés sur le site Web public de votre institution, les stratégies et les outils mis en œuvre pour l'archivage de ces documents et les principaux intervenants dans ce processus. Les données recueillies demeureront confidentielles et anonymes. Un formulaire de consentement vous sera soumis avant de commencer l'entrevue. Vous serez libre de ne pas répondre aux questions et de cesser l'entrevue en tout temps.

Votre participation est essentielle à la réalisation de ce projet de recherche. Je communiquerai avec vous sous peu pour connaître votre disponibilité et, je l'espère, prendre rendez-vous.

Je vous remercie à l'avance de l'attention que vous porterez à cette demande. Je vous prie d'agréer, [Madame, Monsieur], l'expression de nos salutations les plus distinguées.

Aïda Chebbi
Étudiante au doctorat
[Coordonnées]

Annexe 13 : Lettre de remerciements

Montréal, [date]
[Nom du participant], [Poste]
[Adresse]

Objet : Étude sur les pratiques d’archivage du Web organisationnel dans une perspective archivistique

[Madame, Monsieur] [Nom],

Nous vous remercions vivement pour l’entrevue accordée la [date], réalisée dans le cadre d’une recherche visant à mieux comprendre le processus de capture, gestion et conservation des archives Web dans un contexte organisationnel. Votre participation a permis de recueillir des données essentielles pour notre recherche.

Nous vous prions d’agréer, [Madame, Monsieur], l’expression de nos salutations les plus distinguées.

Aïda Chebbi
Étudiante au doctorat
[Coordonnées]

Annexe 14 : Guide d'entrevue (BANQ)

ARCHIVAGE DES SITES WEB DANS UNE PERSPECTIVE ARCHIVISTIQUE ÉTAT DES LIEUX ET TRAVAUX À BANQ

Guide d'entrevue

Organisme : Bibliothèque et Archives nationales du Québec

Personnes rencontrées :

Date : 25 octobre 2012

Durée : 2h

1- Le programme d'archivage du Web à BANQ est :

- à ses débuts
- à un stade expérimental
- bien établi

2- À quand remontent les travaux sur l'archivage du Web à BANQ?

3- Nature de ces travaux :

- Expérimentations
- Service de consultation auprès des organismes (quel type?) au Québec
- Conception d'outils pour l'archivage du Web (lesquels?)
- Autre

4- Unité(s)/service(s) responsable(s)?

5- Collaboration avec d'autres institutions

- à l'échelle nationale?
- à l'échelle internationale?

* Quelles sont les autres institutions responsables de l'archivage du Web dans votre région/pays?

6- Quelles sont les approches et techniques adoptées (par BANQ) ou recommandées (aux autres organismes)?

* Contenus devant être archivés et critères de sélection/évaluation :

a- Archivage du Web québécois

- Tout le Web québécois
- Web gouvernemental
- Web académique
- Publications gouvernementales en ligne
- Web social
- Sites Web statiques/sites Web dynamiques
- Formats particuliers (PDF, animations multimédias, etc.)
- Pas d'archivage

b- Archivage du Web organisationnel au Québec

- Site Web entier (portrait du site Web)
- Composantes particulières du site Web (journaux historiques, base de données, etc.)
- Pas d'archivage

* Fréquence d'archivage?

* Archivage à distance ou archivage à la source?

* Archivage à l'interne ou faire appel aux services d'un tiers-archivageur?

* Formats d'archivage?

* Outils d'archivage (logiciels, équipements informatiques)?

* Avez-vous besoin de permission particulière pour collecter les sites Web?

* Avez-vous un texte légal/réglementaire relatif à l'archivage des sites Web?

* Difficultés rencontrées?

7- Statut des contenus archivés

- Publication?
- Document d'archives?
 - Dans ce cas : comment gérer le cycle de vie d'un site Web? (site/document Web en ligne – versions archivées du site/document Web – archives définitives du site/document Web)

8- Avez-vous développé un cadre de référence pour gérer les projets d'archivage du Web (modèles de politiques ou de règles de conservation relatives aux sites Web)?

9- Quelles options d'exploitation de ces ressources offrez-vous?

* Recherche en ligne?

* Collections thématiques?

10- Imposez-vous des conditions d'accès aux collections Web constituées?

11- Avez-vous défini une stratégie de préservation des contenus archivés?

12- Quelles sont les priorités actuelles de votre institution en matière d'archivage des sites Web?

13- À combien s'élève le budget alloué au programme d'archivage du Web?

14- Autres informations/commentaires

Annexe 15 : Guide de codage

Archives Web

Code qui s'applique à tous les documents Web analysés.

Archives Web\Auteur

Auteur/Unité créatrice du document analysé.

Archives Web\Auteur\Indiqué

Le nom de l'auteur ou de l'unité créatrice est indiqué.

Archives Web\Auteur\Non indiqué

Le nom de l'auteur ou de l'unité créatrice n'est pas indiqué.

Archives Web\Catégorie du document

Catégorie archivistique du document analysé : document de fonction ou document de gestion.

Archives Web\Catégorie du document\Document de fonction

Document de fonction : document produit ou reçu dans le cadre de la mission propre d'une organisation (ex. : formulaire d'inscription à un programme d'étude).

Archives Web\Catégorie du document\Document de gestion

Document de gestion : document produit ou reçu dans le cadre des activités administratives d'une organisation (ex. : rapport annuel).

Archives Web\Composantes du document

Objets documentaires qui composent le document analysé (ex. : texte, image, son, etc.).

Archives Web\Composantes du document\Texte

Le document analysé comporte du texte.

Archives Web\Composantes du document\Image

Le document analysé comporte des images.

Archives Web\Composantes du document\Animation multimédia

Le document analysé comporte une (des) animation(s) multimédia(s).

Archives Web\Composantes du document\Son

Le document analysé comporte des fichiers sonores.

Archives Web\Composantes du document\Vidéo

Le document analysé comporte des fichiers vidéo.

Archives Web\Composantes du document\Base de données

Le contenu du document analysé est généré par une base de données.

Archives Web\Composantes du document\Liens

Le document analysé comporte des liens internes ou externes.

Archives Web\Composantes du document\Liens\Internes

Le document analysé comporte des liens internes.

Archives Web\Composantes du document\Liens\Externes

Le document analysé comporte des liens externes.

Archives Web\Contenu

Description du contenu du document analysé afin de renseigner sur son objet.

Archives Web>Date de création ou de mise à jour

Date de création ou de mise à jour du document analysé.

Archives Web>Date de création ou de mise à jour\indiquée

Date de création ou de mise à jour du document analysé est indiquée.

Archives Web>Date de création ou de mise à jour\non indiquée

Date de création ou de mise à jour du document analysé n'est pas indiquée.

Archives Web\Fonctionnalités du document

Fonctionnalités disponibles sur le document analysé.

Archives Web\Fonctionnalités du document\Navigation

Fonctionnalités disponibles sur le document analysé : navigation dans la même page observée ou vers d'autres pages.

Archives Web\Fonctionnalités du document\Personnalisation

Des options de personnalisation du contenu sont offertes (ex. : changement de la taille des caractères, des couleurs, etc.).

Archives Web\Fonctionnalités du document\Recherche

Il est possible de faire une recherche dans le document (moteur de recherche).

Archives Web\Fonctionnalités du document\Interaction

Il est possible pour un internaute d'interagir avec le webmestre (ou autre acteur organisationnel) via la page Web (ex. : formulaire pour envoyer une question ou des commentaires).

Archives Web\Format du document

Format du document analysé (ex. : PDF, HTML).

Archives Web\Modes de création

Mode de création du document analysé.

Archives Web\Modes de création\Numérisation

Document produit en format papier, ensuite, numérisé et diffusé en ligne.

Archives Web\Modes de création\Suite bureautique

Document produit par une suite bureautique (ex. : document produit par Word).

Archives Web\Modes de création\Suite multimédia

Document produit par une suite multimédia.

Archives Web\Modes de création\Système de gestion de contenus

Document produit par un système de gestion de contenus.

Archives Web\Type

Type du document analysé (ex. : rapport, lettre, formulaire).

Archives Web\Valeurs archivistiques

Valeurs archivistiques du document analysé.

Archives Web\Valeurs archivistiques\Valeur administrative

Les archives sont une preuve, une trace d'une transaction, d'une action, d'une décision.

Archives Web\Valeurs archivistiques\Valeur historique

Les archives sont une source d'information, un témoignage sur des faits, des actions et des décisions.

Entrevues\Actif documentaire du site Web

Description de l'actif documentaire du site Web (volume et nature des documents diffusés/produits sur le site Web) selon le participant à l'étude.

Entrevues\Actif documentaire du site Web\Catégories de documents

Catégories des documents diffusés/produits sur le site Web selon le participant à l'étude.

Entrevues\Actif documentaire du site Web\Catégories de documents\Documents de gestion

Documents liés à la gestion des affaires communes entre les organisations : gestion des ressources humaines, financières, matérielles, etc.

Entrevues\Actif documentaire du site Web\Catégories de documents\Documents d'exploitation

Documents liés à la mission propre d'une organisation (ex.: pour un milieu académique = plans de cours, rapports de recherche, guides étudiant).

Entrevues\Actif documentaire du site Web\Formats

Formats des documents diffusés sur le site Web selon le participant à l'étude.

Entrevues\Actif documentaire du site Web\Formats\Format(s) propriétaire(s)

Les documents diffusés sur le site Web se présentent dans un format propriétaire (ex. : PPT).

Entrevues\Actif documentaire du site Web\Formats\Format(s) standard(s)

Les documents diffusés sur le site Web se présentent dans un format standard ou normalisé (ex. : XML, PDF).

Entrevues\Actif documentaire du site Web\Formats\Format(s) Web

Les documents sont produits dans un format Web (ex. : HTML, ASP, FLASH).

Entrevues\Actif documentaire du site Web\Provenance

Provenance ou environnement de production des documents Web.

Entrevues\Actif documentaire du site Web\Provenance\Documents déposés sur le site Web

Documents produits dans un autre environnement et publiés/diffusés sur le site Web.

Entrevues\Actif documentaire du site Web\Provenance\Documents déposés sur le site Web\Documents externes

Documents produits dans un autre environnement par une unité externe et publiés/diffusés sur le site Web de l'organisme à l'étude.

Entrevues\Actif documentaire du site Web\Provenance\Documents déposés sur le site Web\Documents internes

Documents produits par l'organisation et publiés/diffusés sur son site Web.

Entrevues\Actif documentaire du site Web\Provenance\Documents liés au site Web

Documents produits par une application liée au site Web de l'organisme à l'étude (ex. : BDD, applications Web 2.0).

Entrevues\Actif documentaire du site Web\Provenance\Documents propres au site Web

Documents produits dans un environnement Web par l'organisme à l'étude (ex. : Page HTML).

Entrevues\Actif documentaire du site Web\Volume de l'information

Volume de l'information gérée sur le site Web de l'organisation et volume de l'information archivée, si communiqué par le participant à l'étude.

Entrevues\Actif documentaire du site Web\Volume de l'information\Volume des archives Web

Volume de la masse Web archivée, si communiqué par le participant à l'étude.

Entrevues\Actif documentaire du site Web\Volume de l'information\Volume du site Web

Volume du site Web public de l'organisation, si communiqué par le participant à l'étude.

Entrevues\Archivage – autres systèmes

Autres systèmes Web archivés dans l'organisation.

Entrevues\Archivage – autres systèmes\Oui

D'autres systèmes Web dans l'organisation sont archivés.

Entrevues\Archivage – autres systèmes\Non

Aucun autre système Web dans l'organisation n'est archivé.

Entrevues\Caractéristiques du site Web

Caractéristiques du site Web public de l'organisme à l'étude, selon le participant.

Entrevues\Caractéristiques du site Web\Composantes du site Web

Liste des composantes du site Web public, selon le participant.

Entrevues\Caractéristiques du site Web\Objectif(s) du site

Objectif(s) du site Web public, selon le participant.

Entrevues\Caractéristiques du site Web\Clientèle du site

Clientèle visée par le site Web public, selon le participant.

Entrevues\Caractéristiques du site Web\Technologies de développement

Technologies de développements du site Web public, selon le participant.

Entrevues\Caractéristiques du site Web\Type du SIW

Caractéristiques technologiques du site Web analysé, selon le participant.

Entrevues\Caractéristiques du site Web\Type du SIW(s)\Site Web statique

Le site Web est de nature statique, selon le participant.

Entrevues\Caractéristiques du site Web\Type du SIW(s)\Site Web dynamique

Site Web de nature dynamique, selon le participant.

Entrevues\Caractéristiques du site Web\Type du SIW(s)\Site Web à composantes dynamiques

Le site Web comporte des composantes de nature dynamique, selon le participant.

Entrevues\Développements futurs

Projets futurs en matière de gestion et d'archivage du site Web de l'organisation : développement de nouveaux outils, rédaction de procédures, acquisition de matériels, embauche d'une personne compétente, exploitation des archives Web, etc.

Entrevues\Développements futurs\Amélioration des pratiques

Seulement une révision et un contrôle des pratiques de gestion et d'archivage du site Web public de l'organisation sont prévus.

Entrevues\Développements futurs\Non

Aucun développement n'est prévu en matière de gestion et d'archivage du site Web public de l'organisation.

Entrevues\Développements futurs\Oui

De nouveaux développements sont prévus concernant la gestion et l'archivage du site Web public de l'organisation.

Entrevues\Difficultés rencontrées

Difficultés rencontrées dans le développement, la mise en œuvre ou l'application de la stratégie d'archivage du site Web de l'organisation, selon le participant.

Entrevues\Difficultés rencontrées\Difficultés financières

Des difficultés d'ordre financier (ex. : coupure budgétaire, prix élevé des équipements) ont été rencontrées, selon le participant.

Entrevues\Difficultés rencontrées\Difficultés matérielles

Des difficultés liées aux ressources matérielles (ex. : besoin d'un serveur dédié à l'archivage du site Web, aspirateur Web adopté ne permet pas la capture des documents de nature dynamique) ont été rencontrées, selon le participant.

Entrevues\Difficultés rencontrées\Manque de communication

Le participant à l'étude identifie le manque de communication entre les équipes de travail comme une difficulté rencontrée lors de la mise en œuvre de la démarche d'archivage du site Web.

Entrevues\Difficultés rencontrées\Manque de ressources humaines qualifiées

Le participant ne dispose pas, lui ou son équipe de travail, des connaissances et des compétences essentielles pour la gestion et l'archivage du site Web public de son organisation.

Entrevues\Difficultés rencontrées\Manque de temps

Le participant est débordé dans ses tâches quotidiennes et ne trouve pas le temps nécessaire pour s'occuper du volet archivage du site Web public de son organisation.

Entrevues\Documentation de la démarche d'archivage

L'organisation (ou l'unité) a documenté sa démarche d'archivage du site Web : qui fait quoi, comment, outils utilisés, etc.

Entrevues\Documentation de la démarche d'archivage\Non

La démarche d'archivage n'est pas documentée dans l'organisation.

Entrevues\Documentation de la démarche d'archivage\Oui

La démarche d'archivage est documentée dans l'organisation.

Entrevues\Documentation de la démarche d'archivage\Partiellement documentée

La démarche d'archivage est partiellement documentée dans l'organisation : description minimale des procédures, documentation à mettre à jour, documentation de la démarche auprès d'un groupe particulier d'intervenants dans le processus d'archivage, etc.

Entrevues\Évaluation des pratiques d'archivage du SIW

Évaluation des pratiques de gestion et d'archivage du site Web par le participant à l'étude.

Entrevues\Évaluation des pratiques d'archivage du SIW\Bonnes pratiques

Le participant juge que le site Web de l'organisation est bien géré et que les pratiques d'archivage sont adéquates.

Entrevues\Évaluation des pratiques d'archivage du SIW\Mauvaises pratiques

Le participant juge que le site Web de l'organisation est mal géré et que les pratiques d'archivage ne sont pas appropriées.

Entrevues\Évaluation des pratiques d'archivage du SIW\Pratiques à améliorer

Le participant juge que les pratiques de gestion et d'archivage du site Web devraient être améliorées.

Entrevues\Exploitation des contenus archivés

Moyens et outils prévus ou établis pour exploiter les archives Web à des fins d'information et de recherche.

Entrevues\Exploitation des contenus archivés\En réflexion

La façon dont les archives Web peuvent être exploitées (ex. : intégration dans le fonds documentaire, recherche dans les archives Web) est discutée.

Entrevues\Exploitation des contenus archivés\Non

Les archives capturées ne sont pas exploitées.

Entrevues\Exploitation des contenus archivés\Oui

Les documents archivés sont exploités (ex.: développement d'une section archives Web sur le site Web public).

Entrevues\Facteurs

Facteurs indiqués par le participant à l'étude qui peuvent avoir un impact sur le choix ou la mise en œuvre d'une stratégie d'archivage.

Entrevues\Facteurs\Complexité du site Web

Facteur à considérer : complexité du site Web.

Entrevues\Facteurs\Fréquence de mise à jour du site Web

Facteur à considérer : fréquence de mise à jour du site Web.

Entrevues\Facteurs\Collaboration

Facteur à considérer : niveau de collaboration entre différents intervenants impliqués dans le processus de gestion et d'archivage d'un site Web.

Entrevues\Facteurs\Hébergement du site Web

Facteur à considérer : organisme responsable de l'hébergement du site Web.

Entrevues\Facteurs\Ressources

Facteur à considérer : ressources disponibles pour mettre en œuvre la stratégie d'archivage du site Web.

Entrevues\Informations démographiques

Informations démographiques recueillies lors des entrevues sur les participants à l'étude.

Entrevues\Informations démographiques\Sexe

Sexe du participant.

Entrevues\Informations démographiques\Sexe\Masculin

Le participant est un homme.

Entrevues\Informations démographiques\Sexe\Féminin

Le participant est une femme.

Entrevues\Informations démographiques\Domaines d'étude

Domaines d'étude du participant à la recherche.

Entrevues\Informations démographiques\Niveau d'étude

Niveau d'études du participant à la recherche.

Entrevues\Informations démographiques\Catégorie du participant

Catégorie du participant : webmestre, archiviste, autre.

Entrevues\Informations démographiques\Catégorie du participant\Archiviste

Le participant à la recherche est un archiviste.

Entrevues\Informations démographiques\Catégorie du participant\Autre

Le participant à la recherche appartient à une catégorie autre que webmestre ou archiviste.

Entrevues\Informations démographiques\Catégorie du participant\Webmestre

Le participant à la recherche est un webmestre/édimestre.

Entrevues\Informations démographiques\Poste du participant

Description de la fonction (titre du poste) du participant à l'étude.

Entrevues\Informations démographiques\Expérience - Gestion et archivage des SIW

Nombre d'années d'expérience du participant à l'étude dans la gestion et l'archivage des sites Web.

Entrevues\Informations démographiques\Expérience - Poste actuel

Nombre d'années d'expériences dans le poste qu'occupe actuellement le participant à l'étude.

Entrevues\Informations démographiques\Expérience - SI

Nombre d'années d'expérience du participant à l'étude dans la gestion ou le développement de systèmes d'information (ou systèmes d'information Web).

Entrevues\Informations démographiques\Expérience - Total des années

Nombre total des années d'expérience du participant à la recherche sur le marché du travail.

Entrevues\Informations démographiques\Formations - Archivage du Web

Formations reçues par le participant à l'étude pour améliorer ses compétences en matière d'archivage des sites Web.

Entrevues\Informations démographiques\Formations - Archivage du Web\Oui

Le participant à l'étude a suivi une ou plusieurs formations en matière d'archivage des sites Web.

Entrevues\Informations démographiques\Formations - Archivage du Web\Non

Le participant à l'étude n'a pas suivi de formation(s) en matière d'archivage des sites Web.

Entrevues\Outils d'archivage

Outils utilisés ou développés pour la gestion et l'archivage du site Web de l'organisation.

Entrevues\Outils d'archivage\Documentation technique

L'organisation a développé de la documentation technique relative à la gestion et à l'archivage de son site Web public : programmation, scripts, etc.

Entrevues\Outils d'archivage\Politique d'archivage

L'organisation a développé (ou en cours de développement) d'une politique pour l'archivage de son site Web public.

Entrevues\Outils d'archivage\Procédures

L'organisation a développé (ou en cours de développement) de procédures pour la gestion et l'archivage de son site Web public : procédure de capture des contenus, procédure de prise de copies de sauvegarde, procédure de préservation, etc.

Entrevues\Outils d'archivage\Règles de conservation

L'organisation a développé (ou en cours de développement) des règles de conservation pour la gestion et l'archivage de ses archives Web.

Entrevues\Outils d'archivage\Règles de conservation\Règle(s) approuvée(s) par BAnQ

BAnQ a approuvé la ou les règle(s) de conservation liée(s) à la gestion et à l'archivage du site Web public de l'organisation.

Entrevues\Outils d'archivage\Règles de conservation\Règle(s) non approuvée(s) par BAnQ

BAnQ n'a pas encore approuvé la ou les règle(s) de conservation liée(s) à la gestion et à l'archivage du site Web public de l'organisation.

Entrevues\Outils d'archivage\Règles de conservation\Pas de règle(s) de conservation

L'organisation n'a pas encore développé une règle de conservation pour l'archivage de son site Web public.

Entrevues\Rôles et responsabilités

Rôles et responsabilités relatifs à la gestion et à l'archivage du site Web public de l'organisation.

Entrevues\Rôles et responsabilités\Responsabilités\Archivage du site

Rôles et responsabilités associés à l'archivage du site Web (ex. : capture des contenus, gestion des archives Web).

Entrevues\Rôles et responsabilités\Responsabilités\Création du contenu

Rôles et responsabilités associés à la fourniture du contenu ou des documents à publier sur le site Web.

Entrevues\Rôles et responsabilités\Responsabilités\Diffusion de contenu

Rôles et responsabilités associés à la mise en ligne des contenus (ex. : révision linguistique, traitement HTML, diffusion en ligne, mise à jour du contenu).

Entrevues\Rôles et responsabilités\Responsabilités\Responsabilité de gestion du site Web

Rôles et responsabilités associés à la gestion du site Web (d'un point de vue technique et technologique).

Entrevues\Stratégie d'archivage

Mode, approche et technique d'archivage.

Entrevues\Stratégie d'archivages\Non archivage

L'organisme n'archive pas son site Web public.

Entrevues\Stratégie d'archivages\Ressources

Les ressources disponibles et utilisées pour mettre en œuvre la stratégie d'archivage du site Web de l'organisation.

Entrevues\Stratégie d'archivage\Approche d'archivage

Manière dont le processus d'archivage sera opérationnalisé.

Entrevues\Stratégie d'archivage\Approche d'archivage\Macro-archivage

Tout le site Web public de l'organisation est archivé.

Entrevues\Stratégie d'archivage\Approche sélective\Micro-archivage

Une section particulière du site Web public de l'organisation et/ou des documents particuliers qui y sont diffusés (format, objet, catégorie) sont archivés.

Entrevues\Stratégie d'archivage\Approche sélective\Meta--archivage

Les documents relatifs aux opérations de gestion et de maintenance du site Web public sont archivés.

Entrevues\Stratégie d'archivage\Critères de sélection

Critères de sélection des contenus Web à archiver.

Entrevues\Stratégie d'archivage\Finalité de l'archivage

Raisons pour lesquelles le site Web ou certains contenus Web sont archivés.

Entrevues\Stratégie d'archivage\Mode d'archivage

Façon dont on va procéder pour capturer les documents Web.

Entrevues\Stratégie d'archivage\Mode d'archivage\Manuel

Le site/document Web est archivé manuellement (ex. : copie du répertoire du site Web à la source).

Entrevues\Stratégie d'archivage\Mode d'archivage\Automatique

Le site/document Web est aspiré grâce à un logiciel.

Entrevues\Stratégie d'archivage\Mode d'archivage\Tiers-archivage

L'archivage du site Web est réalisé par un tiers-archivageur (ex. : hébergeur du site Web).

Entrevues\Stratégie d'archivage\Technique d'archivage

Procédé technique qui permet de capturer/acquérir des documents Web.

Entrevues\Stratégie d'archivage\Technique d'archivage\Archivage à distance

Le site Web est archivé à distance, à partir d'un poste client.

Entrevues\Stratégie d'archivage\Technique d'archivage\Archivage à la source

Le site Web est archivé à la source du côté serveur.

Entrevues\Stratégie d'archivage\Temps requis

Temps requis pour une opération d'archivage (copies de sauvegarde ou archivage du site selon les règles archivistiques), si communiqué par le participant à l'étude.

Entrevues\Stratégie d'archivage\Unité d'archivage

Archives Web et composantes Web capturées : tout le site, une partie du site, documents postés sur le site Web, documentation Web, journal historique, documents codés en HTML, logiciels et applications, etc.

Entrevues\Stratégie(s) de préservation

Stratégie(s) de préservation envisagée(s) pour pérenniser les documents archivés.

Organisation

Code de l'organisation.

Organisation\O1

Organisme O1.

Organisation\O2

Organisme O2.

Organisation\O3

Organisme O3.

Organisation\O4

Organisme O4.

Organisation\O5

Organisme O5.

Organisation\O6

Organisme O6.

Organisation\O7

Organisme O7.

Organisation\O8

Organisme O8.

Organisation\O9

Organisme O9.

Organisation\O10

Organisme O10.

Organisation\O11

Organisme O11.

Organisation\Milieu

Milieu d'étude.

Organisation\Milieu\Milieu gouvernemental

Organisation appartenant au milieu gouvernemental.

Organisation\Milieu\Milieu professionnel

Organisation appartenant au milieu professionnel.

Organisation\Milieu\Milieu universitaire

Organisation appartenant au milieu universitaire.

Politiques\Archivage-Principes archivistiques

Principes archivistiques qui guident le processus d'archivage d'un site Web.

Politiques\Archivage-Principes Web

Principes Web qui guident le processus d'archivage d'un site Web.

Politiques\Cadre réglementaire

Cadres réglementaires indiqués dans la politique analysée (ex. : référence à des textes législatifs, à des normes).

Politiques\Facteurs à considérer

Facteurs à considérer lors de la mise en œuvre d'une stratégie d'archivage.

Politiques\Facteurs à considérer\Complexité du site

Facteurs à considérer : complexité du site Web.

Politiques\Facteurs à considérer\Type du système

Facteurs à considérer : type du système.

Politiques\Facteurs à considérer\Objectifs du site

Facteurs à considérer : objectifs du site Web.

Politiques\Facteurs à considérer\Fréquence de mise à jour

Facteurs à considérer : fréquence de mise à jour du site Web.

Politiques\Facteurs à considérer\Visibilité du site

Facteurs à considérer : visibilité du site Web.

Politiques\Mesures de gestion

Mesures de gestion documentaire à mettre en œuvre tout au long du cycle de vie du site Web.

Politiques\Mesures de gestion\Conception et mise en ligne du site Web

Mesures de gestion à considérer lors de la conception et la mise en ligne du site Web.

Politiques\Mesures de gestion\Mise à jour, modification et retrait

Mesures de gestion à considérer lors de la mise à jour, la modification ou le retrait du site Web (ou d'un contenu sur le site Web).

Politiques\Mesures de gestion\Conservation

Mesures de gestion à considérer lors de l'archivage du site Web (gestion à des fins administrative et patrimoniale).

Politiques\Objectif(s)

Objectif(s) de la politique (ex. : fournir un cadre de référence pour la gestion et la conservation des archives Web, proposer des recommandations techniques).

Politiques\Organisme créateur

Politiques d'archivage selon l'organisme créateur.

Politiques\Organisme créateur\Archives nationales

Politiques d'archivage produites par des archives nationales.

Politiques\Organisme créateur\Archives provinciales

Politiques d'archivage produites par des archives provinciales.

Politiques\Organisme créateur\Archives universitaires

Politiques d'archivage produites par des institutions universitaires.

Politiques\Organisme créateur\Association professionnelle

Politiques d'archivage produites par des associations professionnelles.

Politiques\Organisme créateur\Bibliothèque nationale

Politiques d'archivage produites par des bibliothèques nationales.

Politiques\Organisme(s) visé(s)

Organismes auxquels s'adresse la politique.

Politiques\Portée

Portée de la politique analysée : politique organisationnelle (politique mise en œuvre dans un contexte précis) ou politique référence (politique élaborée par une institution archivistique).

Politiques\Pays

Politiques d'archivage selon le pays.

Politiques\Pays\Australie

Politiques d'archivage produites en Australie.

Politiques\Pays\Canada

Politiques d'archivage produites au Canada.

Politiques\Pays\États-Unis

Politiques d'archivages produites aux États-Unis.

Politiques\Pays\Grande-Bretagne

Politiques d'archivage produites en Grande-Bretagne.

Politiques\Pays\Nouvelle Zélande

Politiques d'archivage produites en Nouvelle-Zélande.

Politiques\Stratégie d'archivage

Description du mode, de l'approche et de la technique d'archivage recommandées.

Politiques\Stratégie d'archivage\Approche d'archivage

Manière dont le processus d'archivage sera opérationnalisé.

Politiques\Stratégie d'archivage\Approche d'archivage\Archivage orienté objet

Archivage de l'ensemble du site Web (portraits) et/ou de composantes bien particulières de ce dernier.

Politiques\Stratégie d'archivage\Approche d'archivage\Archivage orienté événement

Archivage des traces des transactions réalisées par le biais du site Web.

Politiques\Stratégie d'archivage\Approche d'archivage\Archivage en ligne

Mise en ligne, sur une section particulière du site Web actif, d'une version archivée du site Web (ou de quelques-unes de ses composantes).

Politiques\Stratégie d'archivage\Critères de sélection

Critères de sélection des contenus Web à archiver.

Politiques\Stratégie d'archivage\Mode d'archivage

Façon dont on va procéder pour capturer les documents Web.

Politiques\Stratégie d'archivage\Mode d'archivage\Archivage manuel

Archivage manuel du site Web.

Politiques\Stratégie d'archivage\Mode d'archivage\Archivage automatique

Archivage automatique d'un site Web grâce à un robot de collecte (aspirateur Web).

Politiques\Stratégie d'archivage\Mode d'archivage\Archivage mixte

Archivage à la fois manuel et automatique.

Politiques\Stratégie d'archivage\Mode d'archivage\Externalisation

Archivage du site Web par un tiers-archivageur.

Politiques\Stratégie d'archivage\Technique d'archivage

Le procédé technique qui permet de capturer/acquérir des documents.

Politiques\Stratégie d'archivage\Technique d'archivage\Archivage à distance

Archivage à distance sur un poste client.

Politiques\Stratégie d'archivage\Technique d'archivage\Archivage à la source

Archivage à la source à partir du serveur sur lequel est hébergé le site Web.

Politiques\Stratégie d'archivage\Technique d'archivage\Archivage des transactions

Archivage des transactions réalisées par le biais du site Web.

Politiques\Stratégie d'archivage\Technique d'archivage\Documentation des changements

Documentation des changements ayant lieu sur le site Web.

Politiques\Stratégie d'archivage\Support(s) d'archivage

Supports de conservation des documents Web archivés.

Politiques\Stratégie d'archivage\Support(s) d'archivage\CD-ROM

Support d'archivage : CD-ROM.

Politiques\Stratégie d'archivage\Support(s) d'archivage\DVD

Support d'archivage : DVD.

Politiques\Stratégie d'archivage\Support(s) d'archivage\Autre

Autre support d'archivage.

Politiques\Stratégie d'archivage\Unité d'archivage

Archives Web et composantes Web capturées.

Politiques\Stratégie d'archivage\Unité d'archivage\Site Web entier

Archives Web et composantes Web capturées : tout le site.

Politiques\Stratégie d'archivage\Unité d'archivage\Une section du site Web

Archives Web et composantes Web capturées : une partie du site.

Politiques\Stratégie d'archivage\Unité d'archivage\Documents Web

Archives Web et composantes Web capturées : des documents Web bien particuliers.

Politiques\Stratégie d'archivage\Unité d'archivage\Transactions en ligne

Archives Web et composantes Web capturées : documents Web issus des transactions en ligne.

Politiques\Stratégie d'archivage\Unité d'archivage\Web gouvernemental

Archives Web et composantes Web capturées : un territoire Web bien défini, le Web gouvernemental.

Politiques\Stratégie d'archivage\Unité d'archivage\Documents de gestion du site Web

Archives Web et composantes Web capturées : documents relatifs aux opérations de gestion du site Web.

Politiques\Stratégie d'archivage\Statut des contenus à archiver

Statut des contenus Web à archiver.

Politiques\Stratégie d'archivage\Statut des contenus à archiver\Archives intermédiaires

Statut des contenus Web à archiver : archives intermédiaires destinées à une conservation à moyen terme.

Politiques\Stratégie d'archivage\Statut des contenus à archiver\Archives définitives

Statut des contenus Web à archiver : archives définitives destinées à une conservation permanente.

Politiques\Stratégie d'archivage\Statut des contenus à archiver\Documents à évaluer

Statut des contenus Web à archiver : document à évaluer (pas de statut particulier).

Politiques\Stratégie d'archivage\Stratégies de préservation

Recommandations pour garantir l'exploitabilité et l'accès à long terme aux contenus Web archivés : émulation, migration, copies de sauvegarde, etc.

Politiques\Stratégie d'archivage\Stratégies de préservation\Émulation

Recommandations pour garantir l'exploitabilité et l'accès à long terme aux contenus Web archivés : émulation.

Politiques\Stratégie d'archivage\Stratégies de préservation\Migration

Recommandations pour garantir l'exploitabilité et l'accès à long terme aux contenus Web archivés : migration.

Politiques\Stratégie d'archivage\Stratégies de préservation\copies de sauvegarde

Recommandations pour garantir l'exploitabilité et l'accès à long terme aux contenus Web archivés : copies de sauvegarde.

Politiques\Stratégie d'archivage\Stratégies de préservation\autre

Recommandations pour garantir l'exploitabilité et l'accès à long terme aux contenus Web archivés : autre.

Politiques\Stratégie d'archivage\Outils archivistiques

Outils archivistiques à mettre en œuvre ou recommandés dans la politique (plan de classification, calendrier de conservation, procédure, etc.) pour la gestion et la conservation d'un site Web.

Politiques\Stratégie d'archivage\Outils d'archivage

Outils d'archivage à mettre en œuvre ou recommandés dans la politique (ex. : aspirateur Web, scripts).

Politiques\Stratégie d'archivage\Finalités de l'archivage

Raisons pour lesquelles le site Web ou certains contenus Web doivent être archivés.

Politiques\Stratégie d'archivage\Finalités de l'archivage\Finalités administratives

Le site Web ou certains contenus Web doivent être archivés pour des raisons administratives (preuve).

Politiques\Stratégie d'archivage\Finalités de l'archivage\Finalités du témoignage

Le site Web ou certains contenus Web doivent être archivés pour des raisons patrimoniales (témoignage et information).

Politiques\Stratégie d'archivage\Intervenants

Intervenants-clés devant être impliqués dans le processus d'archivage du Web tel que décrit dans le document analysé.

Politiques\Stratégie d'archivage\Intervenants\Archivistes

Intervenants-clés devant être impliqués dans le processus d'archivage du Web tel que décrit dans le document analysé : archivistes.

Politiques\Stratégie d'archivage\Intervenants\Webmestres

Intervenants-clés devant être impliqués dans le processus d'archivage du Web tel que décrit dans le document analysé : webmestres.

Politiques\Stratégie d'archivage\Intervenants\Service judiciaire

Intervenants-clés devant être impliqués dans le processus d'archivage du Web tel que décrit dans le document analysé : service judiciaire de l'organisation.

Politiques\Stratégie d'archivage\Intervenants\Spécialistes des technologies de l'information/communication

Intervenants-clés devant être impliqués dans le processus d'archivage du Web tel que décrit dans le document analysé : spécialistes des technologies de l'information/communication dans l'organisation.

Politiques\Stratégie d'archivage\Intervenants\Archives nationales

Intervenants-clés devant être impliqués dans le processus d'archivage du Web tel que décrit dans le document analysé : les archives nationales.

Politiques\Stratégie d'archivage\Intervenants\Gestionnaires de contenus

Intervenants-clés devant être impliqués dans le processus d'archivage du Web tel que décrit dans le document analysé : créateurs de contenus Web.

Politiques\Stratégie d'archivage\Intervenants\Rôles et responsabilités

Rôles et responsabilités des différents intervenants devant être impliqués dans la gestion et l'archivage du site Web public de l'organisation.

Politiques\Stratégie d'archivage\Définition(s)

Définitions extraites du document analysé (ex. : page Web, site Web, page d'accueil, technique d'archivage, approche d'archivage).

Politiques\Type de politique

Type de la politique : lignes directrices, guide technique, procédure, etc.

Politiques\SIW visé(s)

Catégorie des systèmes d'information Web visés par la politique analysée : intranet, extranet, site Web public, etc.

Politiques\SIW visé(s)\Site Web public

Site Web visé par la politique : site Web public de l'organisation.

Politiques\SIW visé(s)\Intranet

Site Web visé par la politique : intranet.

Politiques\SIW visé(s)\Extranet

Site Web visé par la politique : extranet.

Politiques\SIW visé(s)\Sites de commerce électronique

Site Web visé par la politique : site de commerce électronique.

Politiques\SIW visé(s)\Web 2.0

Site Web visé par la politique : application Web 2.0.

Politiques\SIW visé(s)\Autres

Site Web visé par la politique : autre.

SIW\Catégorie

Catégorie du site Web analysé.

SIW\Catégorie\Gouvernemental

Catégorie du site Web analysé : site Web gouvernemental.

SIW\Catégorie\Universitaire

Catégorie du site Web analysé : site Web universitaire.

SIW\Catégorie\services d'archives

Catégorie du site Web analysé : site Web d'un service d'archives.

SIW\Catégorie\organisme professionnel

Catégorie du site Web analysé : site Web d'organisme professionnel.

SIW\Date de création

Indication de la date de création du site Web.

SIW\Date de création\Indiquée

Date de création du site Web est indiquée.

SIW\Date de création\Non indiquée

Date de création du site Web n'est pas indiquée.

SIW\Date de mise à jour

Indication de la date de mise à jour du site Web.

SIW\Date de mise à jour\Indiquée

Date de mise à jour du site Web est indiquée.

SIW\Date de mise à jour\non indiquée

Date de mise à jour du site Web n'est pas indiquée.

SIW\Clientèle visée

Clientèle visée du site Web.

SIW\Composantes

Composantes documentaires et technologiques du site Web analysé.

SIW\Formats

Indication des formats de documents présents sur le site Web.

SIW\Fréquence de mise jour

Fréquence de mise à jour du site Web analysé.

SIW\Fréquence de mise jour\Fréquente

Mise à jour fréquente (date de mise à jour similaire sur la majorité des pages Web observées).

SIW\Fréquence de mise jour\Irrégulière

Mise à jour irrégulière (date de mise à jour différente d'une page Web à l'autre).

SIW\Fréquence de mise jour\Fréquente\Non indiquée.

Il n'est pas possible de déterminer la fréquence de mise à jour du site Web analysé

SIW\Interactivité

Possibilités offertes à l'utilisateur d'interagir avec le site Web.

SIW\Interactivité\Formulaire Web

Possibilité de saisir des informations en ligne.

SIW\Interactivité\Personnalisation

Possibilité de personnalisation de l’affichage de l’information sur le site Web analysé.

SIW\Interactivité\Moteur de recherche

Possibilité d’effectuer des recherches dans les pages du site Web analysé.

SIW\Interactivité\Flux d’information

Possibilité de s’abonner à un flux d’information.

SIW\Interactivité\Accès à une base de données

Possibilité d’accéder et d’interroger en ligne une base de données.

SIW\Langage de balisage

Langage de balisage utilisé pour organiser le contenu du site Web analysé.

SIW\Langage de feuille de style

Langage utilisé pour gérer la mise en forme du contenu du site Web analysé.

SIW\Liens

Types de liens présents sur le site Web analysé.

SIW\Liens\Liens internes

Liens qui permettent la navigation à l’intérieur du site Web analysé.

SIW\Liens\Liens externes

Liens qui permettent la navigation à l’extérieur du site Web analysé.

SIW\Multifenêtrage

Structure en cadres du site Web.

SIW\Nom du domaine

Nom(s) du(es) domaine(s) du site Web analysé.

SIW\Normalisation

Le contenu du site Web analysé est affiché dans des formats standards.

SIW\Normalisation\Oui

La majorité des documents présents sur le site Web analysé sont affichés dans un format standard.

SIW\Normalisation\Non

La majorité des documents présents sur le site Web analysé sont affichés dans un format propriétaire.

SIW\Normalisation\Partielle

Coexistence de formats standards et de formats propriétaires sur le site Web observé (proportion plus ou moins égale).

SIW\Objectifs

Objectifs du site Web.

SIW\Plugiciels

Plugiciels essentiels pour la lecture de certains documents Web.

SIW\Plugiciels\nécessaires

Plugiciels nécessaires pour accéder au contenu d'une page/document Web.

SIW\Plugiciels\non nécessaires

Les plugiciels ne sont pas nécessaires pour accéder au contenu d'une page/document Web.

SIW\Protocole de communication

Protocole(s) de communication du site Web analysé.

SIW\Responsables du site

Personne responsable du site Web (personne à contacter pour poser des questions, faire des commentaires, etc.).

SIW\Taille du site

Description de la taille du site Web analysé.

SIW\Taille du site\Petit site

Taille du site Web analysé : petit site.

SIW\Taille du site\Site de taille moyenne

Taille du site Web analysé : site de taille moyenne.

SIW\Taille du site\Grand site

Taille du site Web analysé : site volumineux.

SIW\Type

Type du site Web analysé.

SIW\Type\Statique

Le site Web observé est de nature statique.

SIW\Type\À composantes dynamiques

Le site Web observé comporte quelques composantes de nature dynamiques.

SIW\Type\Dynamiques

Le site Web observé est de nature dynamique.

Manque d'intérêt - Free node

Latent - Le participant ne ressent pas le besoin d'archiver le site Web.

Rôle - Communauté archivistique - Free node

Latent - Rôle de la communauté archivistique (ex.: BAnQ, BAC, ICA).