

Judith Dimitri

Les bonnes pratiques de gestion des archives personnelles à l'ère numérique

©2014 par Judith Dimitri. Ce travail a été réalisé à l'EBSI, Université de Montréal, dans le cadre du cours SCI6111 – Politique de gestion des archives donné au trimestre d'automne 2014 par Yvon Lemay (remis le 4 décembre 2014).

Table des matières

Introduction.....	1
Pourquoi archiver ses documents numériques?.....	2
Risques et défis de l'archivage numérique.....	2
Les dangers qui guettent les documents personnels numériques	3
Quatre défis de l'archivage numérique selon Catherine C. Marshall.....	3
Bonnes pratiques pour les documents numériques.....	4
Choix des documents à conserver	4
Convention de nommage.....	4
Métadonnées	5
Classification.....	5
Formats de sauvegarde.....	5
Espaces de stockage	6
Bonnes pratiques pour les traces sur le Web.....	8
Courriels.....	8
Blogues	9
Présence sur les médias sociaux.....	10
Facebook.....	10
Twitter.....	11
Conclusion	11
Bibliographie.....	12

Introduction

Les archives personnelles numériques sont un sujet brûlant d'actualité. De nos jours, la grande majorité des gens ont un ordinateur muni d'une connexion internet, ou ont accès à ce service, que ce soit à la bibliothèque ou à l'école. Les individus sont donc amenés à produire ou à amasser une quantité parfois impressionnante de documents numériques qui peuvent être des fichiers textes, des images, des vidéos, etc. De plus, la popularité sans cesse grandissante des courriels, des blogues et surtout des médias sociaux fait en sorte que nous laissons beaucoup plus de traces de nos activités sur le Web. Cet essor de production de fichiers et d'actions sur la toile ne va souvent pas de pair avec une préoccupation quant à leur préservation, ce qui est regrettable. Que ce soit par manque de temps ou d'intérêt, ou simplement par manque de sensibilisation, la plupart des individus néglige le traitement de leurs archives numériques. Cela peut par exemple se caractériser par la conservation d'absolument tous les documents que nous avons produits ou amassés sur notre ordinateur, ou encore par l'impression que notre disque dur est l'emplacement idéal pour la conservation à long terme de nos fichiers.

Le présent travail portera sur les bonnes pratiques à connaître afin d'assurer la gestion de ses archives personnelles numériques. Plus précisément, il s'agira

d'un guide faisant la synthèse des méthodes mises de l'avant par différents experts qui se sont intéressés à ce sujet depuis les années 2000. Ce guide se veut accessible aux néophytes qui n'ont pas de connaissances extensives dans le domaine de l'archivistique et des sciences de l'information.

D'abord, il sera question des raisons qui justifient l'archivage des documents personnels numériques. Les risques et les défis de l'archivage numérique seront ensuite abordés, avant de passer aux bonnes pratiques à suivre pour archiver nos documents personnels numériques. Finalement, le travail traitera des bonnes pratiques pour archiver nos traces sur le Web, que ce soit au niveau des courriels, des blogues et des médias sociaux.

Pourquoi archiver ses documents numériques?

Plusieurs raisons peuvent justifier l'archivage de nos documents numériques, ainsi que le temps et l'énergie que celui-ci requiert. Pour commencer, nous désirons souvent conserver des documents pour les informations qu'ils contiennent. (Williams, John et Rowland, 2009, p. 355) Que celles-ci soient utiles pour notre emploi, nos études ou seulement pour nos intérêts personnels, nous pourrions en avoir besoin à l'avenir, et c'est pour cela que nous conservons les documents concernés. Nos obligations légales sont une autre raison qui nous incite à archiver certains de nos documents numériques, du moins pour un certain temps. Par exemple, nous recevons de plus en plus nos factures par courriel en format électronique plutôt que par la poste. Dans le cas des factures d'électricité, de mazout ou de gaz, il convient de les garder pendant « [...] trois ans pour permettre, d'une part, de comparer les dépenses d'une année à l'autre et, d'autre part, de demander une correction ou une révision en cas d'erreur de facturation. » (Bédard *et al.*, 1999, p. 6) En ce qui concerne les factures de câblodistribution et de téléphone, il suffit de les conserver pendant un an, encore une fois pour pouvoir contester une erreur dans les frais périodiques. (Bédard *et al.*, 1999, p. 6) Au-delà de ces délais, pour tous ces types de factures, nous devons garder un spécimen pour témoigner que nous avons reçu un certain service auprès de ce fournisseur, et au tarif indiqué. Beaucoup de types de documents ayant une valeur légale, tels que les talons de chèque de paie, les preuves d'achat ou les baux, sont encore souvent produits en format papier, mais cela pourrait rapidement changer. Pour connaître la durée de conservation prescrite selon chaque sorte de documents, Jacques Bédard et ses collaborateurs ont produit un guide exhaustif à ce sujet (Bédard *et al.*). Certains documents numériques peuvent être pertinents à archiver, puisqu'ils servent à documenter notre vie personnelle ou professionnelle. Il peut s'agir de correspondance avec des amis, de la famille ou des collègues, des photographies ou des vidéos de ceux-ci, ou encore des documents textuels comme notre curriculum vitæ. Certains documents ont une valeur sentimentale pour ceux qui les ont créés ou accumulés, et il peut s'agir d'une raison valable pour les archiver. (Cox, 2008, p. 143) Les émotions suscitées par ces fichiers peuvent découler de leur contexte de création, de leur provenance ou encore de leur contenu. Enfin, des documents peuvent être conservés pour des fins de création. Par exemple, certaines photographies pourraient être utiles pour faire des montages dans le but de célébrer des événements comme un anniversaire. Nos envies créatives peuvent cependant être difficiles à anticiper, ce qui risque de compliquer la sélection des documents à conserver à cette fin.

Risques et défis de l'archivage numérique

Lorsqu'il est question de l'archivage de leurs documents numériques, les individus ont souvent tendance à être négligents, ou à prendre pour acquis que les technologies sont infaillibles. En y ajoutant la relative nouveauté de la

pratique, cela engendre de nombreux risques et défis dont il convient de tenir compte au moment de se lancer dans un processus d'archivage de nos fichiers.

Les dangers qui guettent les documents personnels numériques

L'archiviste Pierre Couchet (2012) a recensé plusieurs risques qui guettent nos archives numériques :

- Perte ponctuelle de fichiers : Il peut nous arriver d'en effacer suite à une erreur de manipulation. Une mauvaise classification ou un oubli, par exemple du titre du fichier recherché, peut nous faire perdre la trace d'un document. Enfin, si un service en ligne que nous utilisons fait défaut, nous pouvons perdre les fichiers que nous y avons stockés.
- Problèmes logiciels : Une panne ou un bogue d'un logiciel ou du système d'exploitation peuvent mettre en péril nos documents.
- Problèmes matériels : Nos fichiers peuvent être corrompus suite à un crash de disque dur ou d'une altération de leur support.
- Perte matérielle : Si nous égarons le support sur lequel sont stockés nos documents, ceux-ci sont du même coup perdus.
- Sinistres : Nos fichiers peuvent disparaître à la suite d'un vol matériel, d'un incendie, d'une inondation ou d'un tremblement de terre.
- Dégradations volontaires de la part d'un tiers : Un virus informatique, une dégradation intentionnelle ou une interception de notre mot de passe peuvent affecter nos fichiers.
- Obsolescence : Les supports et les formats de fichiers peuvent ne plus être supportés par les technologies futures, ce qui rendrait impossible la lecture de nos documents.
- Défection de services de sauvegarde sur Internet : La fermeture d'un fournisseur de services peut entraîner la perte des fichiers que nous y avons stockés.
- Défaillances de confidentialité de services du Web : Nos documents laissés sur le Web peuvent être espionnés, et les mots de passe les protégeant peuvent être dérobés.
- Changements liés à la santé des personnes : Un accident, une maladie ou un décès peuvent empêcher l'accès à nos fichiers ou nous empêcher de les transmettre à d'autres personnes.

Quatre défis de l'archivage numérique selon Catherine C. Marshall

Selon cette employée de Microsoft Research, il existe quatre défis pour l'archivage numérique qui s'ajoutent aux risques déjà encourus par nos documents. (Marshall, 2008a) Ceux-ci découlent directement de nos conceptions erronées sur la conservation des documents numériques, que nous tenons souvent pour acquises. Le premier défi est la gestion numérique (*digital stewardship*), car celle-ci demande beaucoup de temps et d'énergie que bien peu de gens sont disposés à consacrer à leurs archives. Ils se tournent donc vers des solutions plus faciles comme l'utilisation du « cloud » ou de services Internet, qui s'avèrent risquées pour la préservation de nos archives numériques. Le second défi est la distribution des biens (*distributed assets*), puisque les gens ont tendance à conserver des doublons d'un même document sur différents supports, étant donné que certains ont plusieurs ordinateurs ou plusieurs comptes sur le Web. La raison d'être de ces doublons peut être multiple. Il est possible que ce soit pour éviter de perdre un document, pour le partager, pour l'utiliser en ligne ou sur un service de réseautage. Cependant, ces copies ne sont pas des garanties d'une bonne préservation à long terme, car certains de ces services pourraient

disparaître au fil du temps. Le troisième défi est la valeur et l'accumulation des documents (*value and accumulation*), qui est causé par la tendance des gens à amasser une grande quantité de fichiers, pour des raisons qui leur sont parfois inconnues. Les individus ont également de la difficulté à supprimer certains de leurs documents numériques, ce qui est regrettable, car, comme le souligne Catherine C. Marshall, « [d]eletion is an important way for people to exert their own control over their digital assets [...] ». » (2008a) Il faut donc veiller à conserver seulement les documents ayant de la valeur, même si celle-ci dépend de nombreux facteurs et peut varier avec le temps. Enfin, le quatrième défi est le repérage après le stockage à long terme (*retrieval from long-term storage*). Lorsque les gens archivent leurs documents numériques, ils en disposent en fonction de leurs besoins actuels sans penser au futur. Il peut donc leur arriver d'oublier où ils se trouvent et d'être incapables d'en retrouver la trace, car ils se servent de mauvais descripteurs pour effectuer leur recherche.

Bonnes pratiques pour les documents numériques

Dans le cadre du présent travail, les documents numériques abordés seront les fichiers textuels, les images, les vidéos et les fichiers sonores qu'un individu produit ou amasse, généralement dans son ordinateur personnel. Ces documents requièrent en général les mêmes étapes de traitement afin d'être archivés, que ce soit au niveau du choix des documents à conserver, de la convention de nommage, de l'ajout de métadonnées, de la classification et des espaces de stockage à privilégier. Toutefois, les formats de sauvegarde pour l'archivage diffèrent pour chacun de ces types de documents numériques.

Choix des documents à conserver

Pour commencer l'archivage de nos documents numériques, il convient de déterminer quels sont ceux qui méritent d'être conservés. Dans bien des cas, il peut être tentant de tout garder, dans la mesure où les espaces numériques sont vastes et peu coûteux. (Marshall, 2008b) Cependant, il demeure important de garder un contrôle sur les documents archivés, car des fichiers trop nombreux sont plus longs à traiter et peuvent créer de la confusion pour celui qui les gère. (Touchette, 2012, p. 5) Il vaut donc mieux ne conserver que les documents qui ont de la valeur, laquelle, selon Catherine C. Marshall (2008b), peut être déterminée selon trois critères. Le premier est la source de provenance des documents. Le second est les actions qui ont été posées par l'individu sur ces documents. Enfin, le troisième est la manière dont l'individu a disposé des documents, manière qui révèle bien souvent de leur importance à ses yeux. En outre, Claude Huc (2011, p. 31-32) souligne que nous gardons souvent des doublons, des documents pris sur Internet pour un usage momentané, des brouillons et des documents protégés par un mot de passe qui a été oublié. Après réflexion, il pourrait s'avérer pertinent de les élaguer.

Convention de nommage

Ensuite, il est important de faire en sorte que les documents que nous avons choisis de conserver soient facilement identifiables et repérables. Pour cela, il est utile d'établir une convention de nommage qui nous soit propre. Les noms choisis pour les fichiers et les dossiers doivent permettre de savoir ce qu'ils contiennent au premier coup d'œil, et rester pérennes le plus possible. (Huc, 2011, p. 81) Au niveau des caractères à employer, « [...] on se limitera aux 26 lettres de l'alphabet latin, aux 10 chiffres, au point (.), au tiret bas (_) et au trait d'union (-). » (Huc, 2011, p. 81) Les caractères accentués sont à proscrire, puisque le codage dont ils dépendent peut varier d'un ordinateur à l'autre. Il en va de même pour les caractères spéciaux autres que les trois mentionnés. Les

espaces sont eux aussi à éviter. Les noms doivent être significatifs, tout en ne dépassant pas les 40 caractères pour favoriser leur lisibilité. (Huc, 2011, p. 84) Il ne faut pas utiliser de majuscules, car celles-ci ne sont pas distinguées des minuscules par certains systèmes. Enfin, dans le cas des dates, elles devraient être présentées sous la forme AAAAMMJJ ou AAAA-MM-JJ et placées au début du nom si nous désirons classer nos fichiers en ordre chronologique.

Métadonnées

En plus de la convention de nommage, l'ajout de métadonnées à nos archives personnelles numériques permet d'aider à leur identification et à leur repérage. Il s'agit « [...] des données cachées enregistrées dans le fichier et qui le décrivent [...] ». (Touchette, 2012, p. 4) Lorsque nous lançons une recherche dans les fichiers, la requête repêche les documents à la fois selon leur titre et selon leurs métadonnées. Les métadonnées système donnent des informations comme la date de création ou la taille du fichier. (Touchette, 2012, p. 4) Les métadonnées d'application sont celles qui contiennent par exemple le nom de l'auteur ou encore les mots-clés que nous pouvons ajouter pour décrire le fichier. Cela peut être particulièrement utile dans le cas d'une image ou d'un vidéo, puisque nous pouvons en décrire le sujet à l'aide des mots-clés, ce qui donne un meilleur aperçu de leur contenu qu'un simple titre au nombre de caractères limité.

Classification

Afin de pouvoir repérer les documents que nous avons archivés, surtout après une longue période, il est important de développer un plan de classification. Dans le cas de nos archives personnelles numériques, celui-ci consistera en une arborescence de dossiers, de sous-dossiers et de sous-sous-dossiers dans lesquels nos fichiers seront organisés de manière logique. (Touchette, 2012, p. 5) Ces dossiers doivent être mutuellement exclusifs afin que l'emplacement à donner à chaque fichier soit évident à déterminer. Les fichiers peuvent être classés selon des critères comme leur sujet, leur provenance, leur moment de création ou leur format. Il est cependant recommandé de limiter le nombre de niveaux de dossiers, dans le but de limiter le nombre de clics requis pour accéder au document souhaité.

Formats de sauvegarde

Pour que nos archives personnelles numériques résistent à l'épreuve du temps, il faut qu'elles soient sauvegardées dans des formats qui permettront leur lecture à l'avenir. En effet, les formats peuvent disparaître, rendant la lecture de nos fichiers impossible, même s'ils ont été bien conservés. Il est donc important de privilégier les formats ouverts. Ceux-ci sont « [...] publié[s] et libre[s] de droits, sans restriction d'usage ou de mise en œuvre ; c'est le cas des formats définis par le consortium W3C (par exemple HTML, PNG). » (Huc, 2011, p. 114) S'ils ne peuvent être utilisés, nous pouvons aussi considérer les formats propriétaires ouverts, qui se distinguent des formats ouverts, car son propriétaire privé en a défini les usages permis. (Huc, 2011, p. 114) Les formats propriétaires fermés représentent une menace pour la pérennité de nos documents numériques, puisque seul le propriétaire en détient le codage et ce format dépend donc de la compagnie créatrice. Cela signifie que si celle-ci décide de cesser ses activités, une autre compagnie ne pourrait pas créer de programme pour supporter ce format, puisque son code de programmation n'est pas divulgué. Les documents codés dans ce format deviendraient alors illisibles, et leur contenu serait perdu. Pour chaque type de documents, il convient donc de privilégier les formats d'archivage suivants :

- Fichiers textes : Il est préférable de les archiver en format PDF/A. Il s'agit d'une version épurée du PDF, qui est une norme de facto, ce

qui augmente sa pérennité. (PDF Association, 2009) Le format TIFF a longtemps été la norme pour l'archivage des documents textuels numériques, étant donné qu'il « [...] garantit la reproductibilité à long terme et possède une structure établie. » (PDF Association, 2009). Or, contrairement au format TIFF, le format PDF/A permet facilement la recherche dans le texte. De plus, ils « [...] sont davantage compacts et exigent seulement une fraction de l'espace mémoire des fichiers TIFF respectifs, souvent avec une meilleure qualité. » (PDF Association, 2009) Il est possible de créer un document PDF/A directement à partir de Microsoft Word, d'Open Office ou d'Adobe Acrobat.

- Images : En théorie, le format PNG (*Portable Network Graphics*) est celui à choisir dans le cas d'images numériques. (Huc, 2011, p. 132) Il s'agit d'un format ouvert qui est à la fois un standard du W3C et une norme ISO. De nombreux logiciels disponibles sur Internet permettent de convertir en PNG les images qui ne le sont pas déjà. Il est également possible de les convertir en passant par Microsoft Paint. Cependant, il n'est pas compatible avec certains logiciels ou applications, et il est surtout privilégié pour des images graphiques comme des logos. (Holmes, 2013) En pratique, cependant, le format JPEG standard est le plus répandu et le plus approprié pour les photographies, même s'il est un peu moins fiable que le format PNG. (Huc, 2011, p. 132) Son universalité et sa compatibilité avec presque tous les programmes comptent parmi ses atouts. (Holmes, 2013) Toutefois, la qualité d'un fichier JPEG diminue si des modifications y sont apportées. Le format JPEG2000 devait permettre un archivage plus sécuritaire des images, mais il est généralement boudé hors des milieux professionnels, et il n'est pas encore accepté par tous les logiciels. (Casters, 2011) Enfin, le format TIFF donne un format plus volumineux à un fichier afin de préserver sa qualité, ce qui peut être un avantage dans le cas où il faudrait apporter des modifications à une image dans le futur. (Holmes, 2013)
- Vidéos : En ce qui les concerne, le format MPEG-4 est le plus recommandé. Heureusement, les nouvelles caméras enregistrent souvent les vidéos dans le format MPEG-4 AVC. (Huc, 2011, p. 135) Des logiciels sur le Web peuvent servir à convertir les vidéos qui ne sont pas dans ce format. Fait intéressant, le format MPEG-4 AAC peut également servir à l'archivage de nos fichiers sonores, puisqu'il permet une excellente qualité musicale et qu'il s'agit d'un codage ouvert et normalisé. (Huc, 2011, p. 138)

Espaces de stockage

Pour assurer la pérennité de nos archives personnelles numériques, il faut enfin s'assurer que celles-ci soient stockées sur un support qui soit le plus possible à l'abri de l'obsolescence et de la dégradation. Pour ce faire, certains espaces de stockage sont à privilégier :

- Serveur RAID : Il s'agit là de la meilleure option pour l'archivage de ses données numériques. « Cet acronyme, qui signifie *Redundant Array of Independent Disks*, décrit une technique permettant de répartir des données sur plusieurs disques durs. » (LaSalle, 2010) Il en existe plusieurs types, mais c'est le type 1 qui est le plus pertinent pour les besoins de conservation d'un individu. Ce système copie automatiquement les données sur plusieurs disques en même temps, ce qui en augmente la sécurité, étant donné que le

risque que deux disques durs tombent en panne au même moment est pratiquement inexistant. (LaSalle, 2010) Cependant, un tel serveur peut s'avérer coûteux pour un particulier.

- Mémoire flash : Celle-ci est utilisée dans les clés USB, les cartes mémoires et les disques SSD : ce sont ces deux dernières options qui sont à privilégier. (Couchet, 2012) Ces supports sont résistants au choc, portables et leur lecture n'est pas censée avoir de conséquences sur la pérennité du support, quoique cela reste à vérifier. (Huc, 2011, p. 206) Depuis leur invention, leur capacité de stockage ne cesse de croître et leur coût de diminuer, ce qui est avantageux pour un particulier. Toutefois, il faut leur faire attention, puisqu'ils sont faciles à perdre vu leur petit format. (Huc, 2011, p. 207)
- Disque optique : « Ce support s'avère particulièrement avantageux si des gros volumes de données doivent être gérés, pour les vidéos par exemple. » (Couchet, 2012) Le CD a la plus petite capacité de stockage, avec 700 Mo. (Huc, 2011, p. 196) Le DVD peut contenir jusqu'à 4,7 Go de données, ou 8,5 Go en double couche. Le Blu-ray a une impressionnante capacité de 25 Go, ou de 50 Go en double couche. Le DVD et le Blu-ray ont tous les deux une version enregistrable, qui ne permet qu'un seul enregistrement, et une version réinscriptible, laquelle, comme son nom l'indique, permet d'effacer des données ayant été enregistrées précédemment sur le support afin de permettre un nouvel enregistrement. Dans des conditions de conservation optimales, un disque optique peut durer 10 ans. (Couchet, 2012) Thierry Philippon (2011) privilégie l'emploi du Blu-ray par rapport aux autres types de disques optiques en raison de sa qualité de résolution HD, de sa grande capacité de stockage et de sa « [...] qualité de conservation théoriquement beaucoup plus fiable que le support DVD classique : le traitement n'est pas le même, le taux d'erreur, bien plus faible, en raison d'une meilleure adaptation aux variations thermiques. » Il souligne également que le coût des disques Blu-ray a diminué depuis quelques années. Cependant, un graveur de disques Blu-ray peut représenter un investissement plus important, mais on en retrouve désormais des plus abordables sur le marché. (Pavie, 2013)

En revanche, d'autres sont à éviter, du moins si on souhaite archiver ses documents sur le long terme :

- Disque dur : C'est probablement l'outil de stockage le plus utilisé, puisque chaque ordinateur en est équipé. Il en existe aussi qui sont externes. Cependant, comme le souligne Laurent LaSalle (2010), « [I]a question n'est pas de savoir s'il va briser, mais plutôt quand se brisera-t-il. Une fois l'heure fatidique sonnée, il est très difficile (voire impossible) de récupérer les données. » Il ne supporte pas les chocs, et peut subir des pannes mécaniques ou électriques qui mettent en péril nos données. (Huc, 2011, p. 203) « On constate de grandes disparités entre les fabricants et entre les lots d'un même fabricant » (Huc, 2011, p. 204), mais la durée de vie moyenne d'un disque dur est estimée à 5 ans. Après ce temps, les pannes sont plus fréquentes, surtout si le disque a été utilisé intensivement.
- « Cloud » : S'il peut s'avérer intéressant à court ou à moyen terme, il ne s'agit pas d'une méthode à privilégier pour l'archivage à long terme, puisque l'accès aux données est plus lent et peut s'avérer coûteux en frais de consommation Internet. (LaSalle, 2010) La première sauvegarde des données peut être très longue, et des frais

d'hébergement annuel sont exigés pour un espace pouvant aller de 450 Go à « illimité ». (Pavie, 2013) Dropbox offre cependant une option gratuite à 2 Go d'espace de stockage. Si l'abonnement n'est pas renouvelé, nos données peuvent disparaître (LaSalle, 2010), et leur hébergement est évidemment à la merci du fournisseur.

- Services Internet : Comme le « cloud », ces services nous permettent d'accéder à nos fichiers à partir de n'importe quel ordinateur. Toutefois, ils ne sont pas avantageux pour la pérennité de nos documents numériques, puisque la cessation des activités de l'hébergeur pourrait entraîner la disparition de nos données. La sécurité de nos fichiers et les aspects juridiques et commerciaux sont également à risque. (Couchet, 2012)

Bonnes pratiques pour les traces sur le Web

Ces traces sont de plus en plus volumineuses depuis l'avènement du Web 2.0. Tim O'Reilly (2005) l'a défini par sept critères. Pour notre domaine d'intérêt, nous retiendrons que, selon lui, le Web 2.0 est une plate-forme de création de nouveau contenu, et qu'il sollicite la participation des usagers. Celle-ci peut être aussi bien le partage de leurs connaissances que leur évaluation ou leurs commentaires sur un élément donné. Ces deux critères qui définissent le Web 2.0 encouragent donc les individus à produire du contenu sur le Web.

Ces traces que nous laissons sont importantes à archiver, car, contrairement aux documents numériques, elles ne sont pas stockées directement sur notre ordinateur personnel, ce qui les rend particulièrement volatiles. Nos courriels, nos blogues et nos activités sur les médias sociaux dépendent de leurs fournisseurs respectifs. Si ces derniers venaient à fermer, nous pourrions perdre ces données à tout jamais. Des pratiques existent maintenant pour archiver ces trois types de traces sur le Web, mais, de par leur nouveauté, elles soulèvent leur lot de difficultés et de questionnements.

Courriels

Plusieurs raisons rendent pertinent l'archivage de nos courriels. Richard J. Cox (2008, p. 201) va jusqu'à dire que « [t]here seems to be a common consensus about the fact that e-mail plays not only an important, but crucial, role in the activities of organizations and the lives of individuals. » Tout d'abord, le courriel est un outil de communication prépondérant de nos jours. (Sinn, Syn et Kim, 2011, p. 324) Les usagers s'en servent aussi bien pour leurs communications personnelles que professionnelles. Ils sont aussi portés à consulter d'anciens courriels pour se rappeler de certaines informations qu'ils contiennent. Ceux-ci ne sont pas à l'abri, puisque les services de messagerie peuvent disparaître à tout moment, comme ce fut le cas avec AltaVista, privant ainsi ses utilisateurs de leurs courriels. D'ailleurs, « [m]ajor free e-mail services, such as Yahoo Mail, Hotmail, and Gmail, have policies that grant no responsibility to the companies for inaccessibility or failure to preserve contents in e-mail accounts. » (Sinn, Syn et Kim, 2011, p. 320) Nous constatons également que les gens utilisent souvent les courriels comme moyen d'archiver leurs documents, en se les envoyant en pièce jointe dans un courriel. (Sinn, Syn et Kim, 2011, p. 322) Enfin, il y a plus d'une décennie, Susan Lukesh a souligné que sans l'archivage de nos courriels, nous pourrions perdre des documents historiques, puisque les courriels seront aussi importants pour les historiens que le sont les lettres manuscrites anciennes. (Cox, 2008, p. 223)

Pour prévenir une telle perte, Claude Huc a consacré un chapitre entier de son ouvrage *Préserver son patrimoine numérique* à l'archivage des courriels. Pour un client de messagerie installé sur notre ordinateur, comme Outlook, il est possible de faire glisser les courriels dans un dossier créé dans notre répertoire

pour leur archivage. (Huc, 2011, p. 169) Cette opération engendre des fichiers en format EML, qui est ouvert et approprié pour l'archivage. Le nom du fichier correspond au sujet du courriel, alors il pourrait être pertinent de le modifier selon notre convention de nommage. Pour archiver une quantité importante de messages, cette méthode manuelle n'est pas idéale. Il est cependant possible de modifier l'endroit où les courriels sont sauvegardés sur notre ordinateur par le client de messagerie en modifiant les paramètres de notre compte. En effet, le dossier de sauvegarde par défaut est souvent caché et dans un emplacement obscur pour le protéger des logiciels malveillants. (Huc, 2011, p. 172) Il peut donc être intéressant de rediriger la sauvegarde de nos courriels dans un dossier bien identifié à cet effet. Cependant, pour certains clients de messagerie, cette opération permet seulement d'obtenir des fichiers dans un format propriétaire non recommandé pour l'archivage, comme .dbx dans le cas d'Outlook Express et .pst pour Outlook. Il est alors possible de les lire à l'aide d'un outil sur le Web et de transposer leur contenu dans un format propre à l'archivage, mais, comme le souligne Claude Huc (2011, p. 173), « [...] ceci relève plus ou moins du bricolage. » Pour ce qui est des services de messagerie basés sur le Web, comme Hotmail ou Yahoo!, il n'est habituellement pas possible de transférer un message sur notre ordinateur, sauf en faisant un copier-coller du texte du message, de son objet, de son destinataire et de sa date d'envoi dans un fichier texte dans un format archivable. Or, pour Gmail, il existe Gmail Backup, un logiciel gratuit « [...] qui permet de créer sur l'ordinateur une hiérarchie de répertoires par année, puis par mois et, dans chacun des répertoires terminaux, de stocker les messages un par un au format EML. » (Huc, 2011, p. 177) Les noms des fichiers incluent automatiquement la date du courriel, l'heure de son envoi, l'adresse courriel de son destinataire et son objet. Dans les prochaines années, nous devrions voir l'apparition d'outils similaires pour d'autres services de messagerie du Web. En terminant, dans le cas de l'archivage d'une pièce jointe, Claude Huc (2011, p. 151) recommande de ranger celle-ci avec les autres documents de même nature que nous possédons, et non pas avec le courriel dans lequel elle nous a été envoyée.

Blogues

Un blogue « [...] est constitué d'une succession de billets assez courts, accumulés au fil des jours, les plus récents présentés en premier. » (Huc, 2011, p. 295). Les commentaires des lecteurs y jouent un rôle prépondérant, ce qui confère aux blogues un grand dynamisme qui les distingue des sites web traditionnels. (Huc, 2011, p. 296) Si Steven Ovadia (2008, p. 99) souligne que peu a été écrit sur la valeur des blogues en tant qu'archives, beaucoup a été écrit sur ce que peut représenter un blogue personnel pour son auteur. Il peut par exemple avoir une fonction de journal intime, dans lequel le créateur y exprime ses émotions et partage son quotidien. (Lee, 2011, p. 208) Il peut également s'agir d'un endroit qui redirige principalement le lecteur vers d'autres sources sur le Web. Un blogue peut aussi présenter les intérêts de son auteur, et les résultats de ses réflexions et de ses recherches à leur sujet. Peu importe son contenu, le blogue est important pour son créateur, en raison du temps et de l'énergie qu'il y a consacrés. Beaucoup ont pris conscience de cette importance lorsqu'en juin 2004, Weblogs a cessé ses activités et plus de 3000 usagers ont perdu accès à leur blogue sans avoir eu la chance de l'archiver. (Sinn, Syn et Kim, 2011, p. 320) « My entire *life* is in that blog » (Sinn, Syn et Kim, 2011, p. 320) fut le cri du coeur d'un des utilisateurs concernés, visiblement dévasté par cette perte.

Afin d'éviter qu'une telle situation ne se reproduise, Claude Huc suggère une méthode d'archivage des blogues personnels qui s'inspire de l'archivage des sites web. Comme les blogues sont le plus souvent créés par un outil spécialisé,

son créateur ne dispose souvent pas d'une copie locale. Il faut donc se résoudre à utiliser la technique d'aspiration du blogue. Pour ce faire, il existe de nombreux logiciels libres spécialisés, comme HTTrack ou Wget, qui sont en mesure de produire sur notre ordinateur une copie locale de notre blogue, dans laquelle il est possible de naviguer comme si nous y étions connectés. Par contre, il faut considérer qu'à l'avenir il se peut que certains liens contenus dans notre blogue ne soient plus fonctionnels, puisque les sites auxquels ils renvoient pourraient avoir disparu. Katharine Dunn et Nick Szydowski (2009, p. 18), quant à eux, suggèrent l'utilisation du logiciel libre SiteSucker, mais ils soulignent que s'il peut archiver correctement des blogues créés sur des plateformes telles que WordPress ou Blogger, il ne convient pas à tous les blogues. En effet, SiteSucker ne peut interpréter le format de certains blogues et, dans pareil cas, il ne se contente que d'archiver leur contenu et non leur forme.

Présence sur les médias sociaux

Les bonnes pratiques varient selon le réseau social. Il sera ici question de Facebook et de Twitter, qui sont probablement les deux médias sociaux les plus populaires.

Facebook

Il peut être très intéressant d'archiver nos activités sur Facebook, car, d'après Donghee Sinn et Sue Yeon Syn (2014), « [p]eople indeed present their real selves and document their daily stories on Facebook. Thus, Facebook may serve contemporary people well as a repository of personal identities and histories. » Selon David Saint-Germain et Charles-Édouard Drobecq (2014a), collaborateurs à Convergence, le blogue de l'Association des archivistes du Québec, voici la marche à suivre afin d'archiver son compte Facebook :

- Sur le site de Facebook, il nous faut aller dans les paramètres de l'utilisateur (*General account setting*).
- Un lien « Télécharger mes données Facebook » (*Download my Facebook data*) accompagne chaque catégorie principale. Ce lien mène à « *Start my Archives* ». Une fois cette option sélectionnée, il nous faut composer notre mot de passe.
- Une fois que Facebook aura compilé nos données, nous recevons un message sur notre compte courriel afin de confirmer notre inscription sur le site. Le lien présent dans le courriel nous amène à nous authentifier et, une fois que cela est fait, le téléchargement va débiter.
- Le document qui en résulte est un fichier compressé .zip. Une fois décompressées, les données peuvent être consultables sur un navigateur web en format HTML, mais elles peuvent être sauvegardées en format PDF/A pour une plus grande pérennité.

Les auteurs relèvent cependant qu'il existe plusieurs bémols à l'archivage de nos activités sur Facebook. Par exemple, la procédure d'archivage présentée ne permet pas à l'utilisateur d'obtenir les photos postées sur Facebook par d'autres personnes et sur lesquelles il a été identifié, ce qui est normal quand nous pensons aux questions de droits d'auteur que cela soulèverait. Le processus d'archivage ne semble pas permettre de sauvegarder nos activités sur notre « mur » qui datent de plus de deux ans. De plus, il est incomplet, puisqu'il ne permet pas de sauvegarder les commentaires et les « J'aime » des autres usagers sur nos activités, ni les groupes dont nous faisons partie. Si nous souhaitons archiver notre journal d'activité (*Activity log*), qui contient toutes les opérations que nous avons réalisées depuis l'ouverture de notre compte, la sauvegarde en PDF/A ne donne que des résultats partiels, et l'opération « Copier/coller » entraîne une perte de métadonnées et de la présentation. Étant donnée la relative

nouveauté de Facebook, nous pouvons toutefois espérer que des applications viendront pallier à ces manques dans un avenir rapproché.

Twitter

David Saint-Germain et Charles-Édouard Drobecq (2014b) se sont également penchés sur l'archivage de nos activités sur Twitter. Ils suggèrent de procéder comme suit :

- Sur le site de Twitter, il nous faut d'abord accéder à nos paramètres d'utilisateur.
- Ensuite, il nous suffit de cliquer sur le bouton « *Request your archives* ».
- Twitter nous enverra alors par courriel un lien permettant le téléchargement d'un fichier compressé .zip. Les fichiers qu'il contient sont en format HTML et peuvent être visualisés sur notre navigateur web. Les liens et les hashtags restent alors fonctionnels, un avantage qui est malheureusement perdu si nous choisissons de sauvegarder ces fichiers en format PDF/A pour les archiver.

Cette méthode n'est pas parfaite. En effet, elle n'inclut malheureusement pas nos abonnements ou nos abonnés. Il est toutefois possible de faire une impression PDF/A de ces pages pour contourner ce problème. Twitter, de par la simplicité des traces que nous y produisons, qui se limitent pour la plupart à des tweets de 140 caractères, demeure plus facile à archiver que Facebook.

Conclusion

Pour conclure, il est important de veiller à la bonne gestion de ses archives personnelles numériques en raison des informations qu'elles contiennent et de nos obligations légales, mais aussi pour garder des traces de notre vie personnelle et professionnelle, pour des raisons sentimentales et pour des fonctions de création. Il faut d'autant plus veiller à l'archivage de nos fichiers que ceux-ci sont à risque, notamment en raison de notre négligence ou de la précarité entourant le numérique. Pour bien conserver ses documents personnels numériques, il faut choisir ceux qu'il faut garder et ceux à élaguer, les rendre facilement repérables et identifiables à l'aide d'une convention de nommage et des métadonnées et bien les classer. Il convient également de choisir des formats de fichiers appropriés à la conservation, ainsi que de bons espaces de stockage. Il existe aussi de bonnes pratiques afin de conserver notre présence sur le Web, comme nos courriels, nos blogs et nos activités sur les médias sociaux tels que Facebook et Twitter, mais il s'agit d'une pratique émergente qui peut présenter son lot de défis. Ces traces demeurent cependant importantes à archiver, car elles témoignent de notre existence et de nos activités.

Ce guide présentant une synthèse des bonnes pratiques prônées par des experts dans des sources variées aura permis de démontrer qu'il existe bel et bien des méthodes pour archiver nos documents personnels numériques, et que celles-ci peuvent être à la portée de tous, à condition d'y consacrer du temps et de l'énergie.

Il conviendra toutefois de poursuivre ce guide de bonnes pratiques de préservation des archives personnelles numériques, puisque celles-ci se raffinent et évoluent sans cesse grâce aux nouvelles technologies. Il existe également des documents numériques qui pourraient être souhaitables d'archiver et qui n'ont pas été abordés ici, par exemple les fils RSS ou la liste de nos favoris contenue dans notre navigateur web. Il serait donc pertinent d'étoffer la présente recherche en abordant l'archivage de telles données. Enfin, des mesures de conservation de plus haut niveau ont été passées sous silence, notamment les mesures de sécurité et d'entretien à mettre en place pour la protection de nos documents personnels numériques. Celles-ci couvrent un éventail de sujets

allant des lieux physiques d'entreposage des archives aux mots de passe permettant l'accès à celles-ci (De Gagné, 2011, p. 6), et il serait utile aux particuliers d'en prendre connaissance dans une nouvelle version plus développée de ce guide.

Bibliographie

- Bédard, J., Forcier, S., Lemay, N., Lévesque, M., Minotto, C., Poulin-Thibault, S. et Racine, B. (1999). *Que faire avec vos documents personnels? Lesquels conserver? Pendant combien de temps? Lesquels détruire? Quand?* Repéré à http://www.banq.qc.ca/documents/services/archivistique_ged/ARCHPER.pdf
- Casters, S. (2011, 18 août). Le Jpeg 2000 : peu répandu [Billet de blogue]. Repéré à <http://www.01net.com/editorial/507142/le-jpeg-2000-peu-repandu/>
- Cox, R. J. (2008). *Personal archives and a new archival calling : readings, reflections and ruminations*. Duluth, Minn. : Litwin Books.
- Couchet, P. (2012, 31 mars). L'archivage pérenne des données numériques personnelles [Billet de blogue]. Repéré à <https://archivengines.wordpress.com/2012/03/21/archivage-perenne-donnees-numeriques-personnelles/>
- De Gagné, J.-F. (2011). *Les documents personnels numérisés* (Travail aux cycles supérieurs, Université de Montréal). Repéré à <http://hdl.handle.net/1866/5099>
- Dunn, K. et Szydlowski, N. (2009). Web Archiving for the Rest of Us: How to Collect and Manage Websites Using Free and Easy Software. *Computers in Libraries*, 29(8), 12-18. Repéré à <http://search.proquest.com/docview/57698619?accountid=12543>
- Holmes, N. (2013, 1^{er} juillet). What's the difference between GIF, PNG, JPEG, and TIFF? [Billet de blogue]. Repéré à <https://www.smartimage.com/whats-the-difference-between-gif-png-jpeg-and-tiff/>
- Huc, C. (2011). *Préserver son patrimoine numérique : classer et archiver ses e-mails, photos, vidéos et documents administratifs : guide à l'usage des particuliers et des entrepreneurs individuels*. Paris : Eyrolles.
- LaSalle, L. (2010, 20 décembre). Résolutions 2011 : Soyez efficace dans l'archivage de vos données [Billet de blogue]. Repéré à <http://blogues.radio-canada.ca/triplex/2010/12/20/resolutions-2011-soyez-efficace-dans-l-archivage-de-vos-donnees/>
- Lee, C. A. (2011). Collecting the Externalized Me : Appraisal of Materials in the Social Web. Dans C. A. Lee (dir.), *I, Digital : personal collections in the digital era* (p. 202-238). Chicago : Society of American Archivists.
- Marshall, C. C. (2008a). Rethinking personal digital archiving Part I: four challenges from the field. *D-Lib Magazine*, 14(3-4). Repéré à <http://www.dlib.org/dlib/march08/marshall/03marshall-pt1.html>
- Marshall, C. C. (2008b). Rethinking personal digital archiving Part II: implications for services, *D-Lib Magazine*, 14(3-4). Repéré à <http://www.dlib.org/dlib/march08/marshall/03marshall-pt2.html>
- O'Reilly, T. (2005). What is Web 2.0 : Design Patterns and Business Models for the Next Generation of Software. Repéré à <http://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html>
- Ovadia, S. (2006). The need to archive blog content. *Serials Librarian*, 51(1), 95-102. doi : 10.1300/J123v51n01_06
- Pavie, O. (2013, 14 octobre). Comment sauvegarder les photos pour toujours? [Billet de blogue]. Repéré à <http://blog.ontrack.fr/comment-sauvegarder-photos/>
- PDF Association. (2009, 13 septembre). PDF/A – Une nouvelle norme pour l'archivage à long terme. [Billet de blogue]. Repéré à

<http://www.pdfa.org/2009/09/pdfa-une-nouvelle-norme-pour-larchivage-a-long-terme/?lang=fr>

- Philippon, T. (2011, 9 mai). Archiver ses vidéos : quelles solutions, quelles méthodes? [Billet de blogue]. Repéré à <http://www.magazinevideo.com/test-en-ligne/archiver-ses-vidéos-quelles-solutions-quelles-méthodes/18526.htm>
- Saint-Germain, D. et Drobecq, C.-É. (2014a, 12 avril). Facebook : Archiver les médias sociaux, partie 2 [Billet de blogue]. Repéré à <https://archivistesqc.wordpress.com/2014/04/12/facebook-archiver-les-medias-sociaux-partie-2/>
- Saint-Germain, D. et Drobecq, C.-É. (2014b, 19 mai). Twitter : Archiver les médias sociaux, partie 3 [Billet de blogue]. Repéré à <https://archivistesqc.wordpress.com/2014/05/19/twitter-archiver-les-medias-sociaux-partie-3/>
- Sinn, D. et Syn, S. Y. (2014). Personal documentation on a social network site: Facebook, a collection of moments from your life? *Archival Science*, 14(2), 95-124. doi : 10.1007/s10502-013-9208-7
- Sinn, D., Syn, S. Y. et Kim, S.-M. (2011). Personal records on the web: Who's in charge of archiving, Hotmail or archivists? *Library & Information Science Research*, 33(4), 320-330. doi : 10.1016/j.lisr.2011.02.004
- Touchette, J. (2012). *Le patrimoine numérique, le Web et la mort* (Travail aux cycles supérieurs, Université de Montréal). Repéré à <http://hdl.handle.net/1866/7038>
- Williams, P., John, J. L. et Rowland, I. (2009). The personal curation of digital objects. A lifecycle approach. *Aslib Proceedings*, 61(4), 340-363. doi : 10.1108/00012530910973767