Annexe 19 Principes pour l'élaboration…
InterPARES 2
Annexe 20 Principes directeurs…
Annexe 21 Principes directeurs…

Annexe 20
Principes directeurs à l'usage des producteurs
Produire et maintenir des documents numériques : recommandations à l'usage des particuliers et des petites structures

par Philippe Eppard, University of Albany, State University of New York

Introduction

De nos jours, la plus grande partie de l'information est créée et stockée sous forme numérique. Les avantages du support numérique sont désormais bien connus de tous. Il permet de créer rapidement des documents et de les éditer et les réviser facilement. De les diffuser, grâce à Internet, dans le monde entier en l'espace de quelques secondes. De les manipuler et de les transformer pour les utiliser à diverses fins. Le support numérique apporte également une solution au problème récurrent du stockage de volumineux dossiers de documents papier.

Les bienfaits du numérique, cependant, ont un coût. Ce n'est que très récemment qu'on a commencé à prendre la pleine mesure des problèmes inhérents au numérique. Par exemple, l'information numérique n'est accessible qu'à partir d'un ordinateur. Qui plus est, celui-ci doit être doté des logiciels nécessaires pour lire les chaînes de bits que contient le disque ou la bande magnétique. La facilité avec laquelle les documents peuvent être reproduits et la prolifération des copies rendent plus difficile l'identification de la version complète ou finale d'un document numérique. La protection des droits d'auteur et des droits de propriété intellectuelle en général est mise à mal par la diffusion anarchique de données sur le Web. Enfin, tous les documents numériques sont vulnérables aux virus et aux problèmes techniques, et leur accessibilité est menacée à plus ou moins brève échéance par l'évolution des logiciels et du matériel informatique.

Dès lors, comment s'étonner que certains aient la nostalgie du support papier et de sa rassurante matérialité ? Pourtant, bien que nos systèmes de production et de maintenance de l'information soient appelés pour quelque temps encore à être des systèmes hybrides, contenant à la fois des documents papier et des documents numériques, il n'y a pas de retour en arrière possible : la révolution numérique est là et bien là. Par conséquent, nous devons tous être conscients des risques associés aux documents numériques et connaître les méthodes et pratiques qui s'offrent à nous pour nous en prémunir.

Ces principes directeurs ont été élaborés à l'intention des individus qui sont amenés à produire des documents numériques dans le cadre de leurs activités professionnelles et personnelles, afin de les aider à prendre des décisions éclairées pour tout ce qui touche à la production et la maintenance de ces documents en vue d'assurer leur conservation aussi longtemps qu'il sera nécessaire. Ces principes pourront également être utiles aux petites structures, comme les cabinets médicaux, les groupes d'experts-conseils ou les équipes de recherche notamment.

Ces principes s'appliquent aussi bien aux documents à maintenir pendant une courte période qu'à ceux qui exigent d'être conservés sur le long terme. Le respect de ces principes contribuera à assurer que les archives méritant d'être conservées sur le long terme soient accessibles lorsqu'elles seront remises entre les mains de

l'archiviste.

Définitions

Avant de présenter les recommandations pour la production et la maintenance de documents numériques, il convient de préciser le sens de certains termes utilisés dans les pages qui suivent.

Aux fins des présents principes directeurs, un *document d'archives* se définit comme tout document produit ou reçu par toute personne physique ou morale dans l'exercice de son activité et sélectionné en vue d'une action ultérieure ou à des fins de consultation. Une *publication* désigne tout document destiné à être largement communiqué ou diffusé. Tous les documents d'archives et les publications sont des documents et contiennent des données. Un *document* s'entend de toute information consignée sur un support sous une forme fixe. L'*information* est un ensemble de données destiné à être communiqué dans le temps ou l'espace, et la *donnée* constitue la plus petite unité d'information signifiante et indivisible.

Ces principes directeurs proposent des recommandations en vue de produire et de maintenir des documents numériques, et plus particulièrement des documents d'archives, fiables dont l'exactitude et l'authenticité pourront être maintenues et préservées dans le temps. Qu'entend-on par « fiabilité », « exactitude », « authenticité » et « authentification » ?

Aux fins des présents principes directeurs, la *fiabilité* est définie comme la qualité d'un document auquel on peut accorder foi en tant qu'énoncé des faits. La fiabilité est de la responsabilité de l'auteur des documents, qu'il s'agisse d'un individu ou d'une personne morale au nom de laquelle un individu rédige les documents. Elle est évaluée à partir de la complétude et de l'exactitude des documents, et du degré de contrôle exercé sur leur processus de production.

L'*exactitude* est la mesure dans laquelle les données d'un document sont précises, correctes, fidèles et exemptes d'erreurs ou d'altérations. Pour que l'exactitude d'un document puisse être assurée, il faut en contrôler les processus de production, de transmission, de maintenance et de conservation. Avec le temps, la responsabilité de l'exactitude passe de l'auteur au records manager puis au service d'archives en charge de la conservation définitive (s'il y a lieu).

L'*authenticité* désigne la qualité d'un document d'être ce qu'il prétend être et de n'avoir été ni corrompu ni altéré. Dans le cas des documents d'archives, l'authenticité renvoie donc à la confiance que l'on peut leur accorder en tant que documents d'archives. Pour être en mesure de présumer et de préserver l'authenticité au fil du temps, il faut définir et maintenir l'identité des documents et protéger leur intégrité. L'authenticité est menacée chaque fois qu'un document est transmis dans l'espace et dans le temps. Avec le temps, la responsabilité de l'authenticité passe du records manager au service d'archive en charge de la conservation définitive.

L'*authentification* est une déclaration d'authenticité résultant de l'insertion ou de l'ajout d'éléments ou d'énoncés dans un document donné ; les règles gouvernant le processus d'authentification sont fixées par la loi. C'est donc un moyen de prouver, à un instant T, que le document est bien ce qu'il prétend être. Les mesures d'authentification numérique, comme l'utilisation d'une signature numérique, assurent seulement que les documents sont authentiques lors de leur réception et qu'ils ne peuvent être rejetés, mais ne garantissent en rien qu'ils demeureront authentiques par la suite.

Recommandations

1. Privilégier le matériel informatique, les logiciels et les formats de fichiers qui assureront la meilleure accessibilité aux documents numériques au fil du temps.

Accéder à des documents numériques suppose de disposer des bons logiciels. Un logiciel qui n'est pas compatible avec les versions antérieures (rétrocompatibilité) ou les versions ultérieures (compatibilité ascendante) rend l'accès aux documents d'archives difficile à long terme. Il est également essentiel que les logiciels dédiés à une seule application fonctionnent bien avec ceux des autres applications et systèmes (interopérabilité). Il convient donc d'être particulièrement vigilant quant au choix des logiciels et des matériels, et de privilégier ceux qui permettront un accès continu aux documents.

- A. *Choisir des logiciels qui permettent d'afficher les documents dans leur présentation originale.* Dans l'idéal, pour rester intelligibles et accessibles, les documents doivent conserver la même apparence au fil du temps. Il faut s'assurer que tout nouveau logiciel sera capable de lire les anciens documents dans le format de logiciel sous lequel ils sont conservés et de les afficher à l'écran dans leur forme initiale. En d'autres termes, il faut s'assurer de la rétrocompatibilité des nouveaux logiciels.
- B. *Choisir des logiciels et du matériel informatique qui permettent de partager facilement des documents numériques.* Les logiciels doivent accepter et produire des fichiers dans différents formats. La capacité d'interagir facilement avec d'autres technologies s'appelle l'« interopérabilité ». Elle facilite l'accès aux documents et leur déplacement dans d'autres systèmes.
- C. *Utiliser des logiciels conformes à des normes ou des standards.* C'est un des meilleurs moyens d'assurer la longévité des documents. Il est recommandé d'opter pour des normes ayant été approuvées par des organisations nationales et internationales. Celles-ci sont désignées sous le nom de « normes de droit » [526]. Lorsque les documents n'entrent dans le champ d'aucune norme, il est tout de même possible de favoriser leur longévité en choisissant un logiciel dont l'utilisation est très répandue. On parle alors de « norme de fait » ou « standard »[527]. Il est préférable d'utiliser des logiciels *open source*, c'est-à-dire des logiciels non propriétaires et disponibles sans restriction (voir la sous-section G en page suivante).
- D. *Conserver les spécifications des logiciels.* Ce type de documentation (par exemple, les manuels de l'utilisateur ou toute autre description plus détaillée du logiciel) sera essentiel pour accéder plus tard aux documents ou les migrer vers un nouvel environnement informatique au fil des évolutions technologiques. Il est particulièrement important de documenter de façon exhaustive les logiciels développés en interne.
- E. *En cas de personnalisation d'un logiciel, documenter les modifications apportées.* Il est indispensable de donner des informations détaillées sur les modifications et de décrire précisément les caractéristiques et les propriétés du document qui en découle, ainsi que les résultats attendus de cette personnalisation. Une bonne pratique consiste à inclure ces informations sous la forme de commentaires dans le code du programme. Ainsi, l'information ne sera pas perdue, puisqu'elle fera partie du fichier, et elle sera très utile à ceux qui devront faire des ajustements plus tard, lorsque la technologie aura évolué.
- F. *Documenter la construction de l'ensemble du système.* En vue de favoriser son accessibilité dans la durée, il est conseillé de documenter la structure et les fonctions du système, c'est-à-dire d'identifier ses composants matériels et logiciels, y compris les périphériques, son système d'exploitation et les progiciels. Cette documentation précisera la façon dont les progiciels représentent l'information, la traitent et la communiquent entre eux et aux utilisateurs. Grâce à cette documentation élémentaire, les utilisateurs futurs comprendront le contexte dans lequel le système a été mis en place et était utilisé. Elle fournira les informations nécessaires à la mise à jour du système à mesure que le matériel informatique et les logiciels évolueront.
- G. *Choisir chaque fois que possible des formats logiques décompressés, non propriétaires, largement utilisés et indépendants des plateformes, et dont les spécifications sont disponibles sans restriction.* Ces formats sont souvent désignés sous le nom de « formats ouverts », ce qui signifie que leurs spécifications sont publiées et disponibles sans restriction. Cependant, cela peut également vouloir dire que le format est libre de tout droit de propriété intellectuelle maintenant et qu'il pourra y être soumis à l'avenir et/ou que le format est adopté à grande échelle. Il convient de noter que les formats « ouverts » ne sont pas nécessairement les mêmes formats que ceux qui sont produits par les « logiciels libres », puisque ce second terme désigne les logiciels dont le code source est accessible à tous et peut être modifié. Un logiciel libre ne produit pas nécessairement des formats non propriétaires. Il faut faire la

différence entre les formats de fichier, les formats conteneurs et les formats de balisage, tels que les fichiers XML, et veiller à ce que la version, l'encodage et les autres caractéristiques soient précisément indiqués. Pour les fichiers XML, on s'assurera qu'ils sont bien formés et valides et accompagnés des définitions de type de documents (DTD) ou des schémas correspondants. Si ces mesures se révèlent difficiles ou compliquées à mettre en œuvre, il est recommandé de prendre contact avec un centre d'archives qui accepte des documents numériques et de choisir parmi les formats qu'il recommande pour l'archivage pérenne. Lorsque cela est possible, il est préférable de ne pas compresser les documents numériques, car cela peut entraîner des problèmes pour leur conservation sur le long terme. S'il n'est pas possible de procéder autrement, les techniques de compression sans perte, conformes aux normes internationales reconnues, sont à privilégier.

2. Veiller à ce que le contenu et la forme des documents numériques maintenus en tant que documents d'archives soient stables et fixes.

Un des grands avantages des documents numériques est la facilité avec laquelle il est possible d'éditer, de réviser ou de mettre à jour l'information. Toutefois, cela signifie également que des informations importantes peuvent être modifiées voire perdues, accidentellement ou à dessein. Le problème est particulièrement aigu pour les documents d'archives, puisqu'une de leurs caractéristiques est que leur contenu ne peut être modifié. Il s'ensuit que les informations et les données du document d'archives ne peuvent être ni écrasées, ni modifiées, ni supprimées ou enrichies. Un système qui contient des informations ou des données changeantes, susceptibles de variations, ne contient donc pas réellement des documents d'archives tant que quelqu'un n'a pas décidé de les créer et de les enregistrer sous une *forme fixe*[528] et avec un *contenu stable*[529].

Si la notion de contenu stable ne soulève pas de difficulté particulière, le concept de forme fixe est plus complexe. Il signifie que le message véhiculé par un document d'archives numérique (ou un autre objet numérique) peut s'afficher à l'écran dans la même présentation que celle qu'il avait lors de son élaboration ou de sa réception et de son enregistrement initial. Les trains de bits qui composent le document d'archives numérique et déterminent sa présentation numérique (c'est-à-dire, son format de fichier) peuvent changer, mais sa présentation documentaire doit demeurer identique. Par exemple, lorsqu'un document créé dans Microsoft Word est ensuite enregistré en PDF, la présentation numérique du document est modifiée – en passant du format de fichier .doc de Microsoft Word au format de fichier .pdf d'Adobe –, mais sa présentation documentaire, également appelée *forme documentaire*[530], ne change pas. Il est donc possible d'affirmer que le document a une forme fixe.

Dans certains cas, les documents numériques peuvent être présentés de plusieurs manières. En d'autres termes, l'information qu'ils véhiculent peut prendre différentes formes. Par exemple, il est possible de présenter des données statistiques sous la forme d'un « camembert », d'un diagramme en bâtons ou d'un tableau. Cependant, les variations potentielles de ces modes d'affichage sont généralement limitées par le système. Lorsque c'est le cas, on peut considérer que chaque présentation documentaire a un contenu stable et une forme fixe dans la mesure où l'information est choisie à partir d'un ensemble fixe de données dans le système et que ce sont les règles du système qui gouvernent la forme de leur présentation.

La situation est similaire lorsque la sélection du contenu et de la forme est faite à partir d'un ensemble important d'informations fixes qui ne sont que partiellement sollicitées chaque fois que l'utilisateur effectue une recherche dans le système. Si la même recherche produit toujours les mêmes résultats sur le plan du contenu et de la forme documentaire, on peut considérer que ces résultats ont un contenu stable et une forme fixe. Par conséquent, si l'auteur d'un document établit des règles fixes pour la sélection du contenu et de la forme qui n'autorisent qu'un degré de variabilité déterminé et stable, c'est-à-dire s'il lui confère une *variabilité limitée*[531], on peut affirmer que le document a un contenu stable et une forme fixe.

La question de la présentation des documents numériques est particulièrement importante pour le maintien et

l'évaluation de la fiabilité et de l'exactitude des documents d'archives. Toute mise à niveau, conversion ou migration ultérieure des données peut entraîner une modification de la forme documentaire. Il convient donc de commencer par définir la forme des documents d'archives associés à chaque activité ou procédure, pour ensuite déterminer les caractéristiques essentielles (c'est-à-dire les *caractères internes et externes*[532] essentiels) de chaque présentation ou forme documentaire. Tout changement futur qui pourrait affecter l'identité et l'intégrité du document d'archives n'en sera que plus facile à repérer, surtout dans le domaine de l'art numérique, où la description certifiée de ces caractéristiques par l'artiste peut faciliter la reconnaissance des droits de propriété intellectuelle sur l'œuvre décrite.

3. S'assurer que les documents numériques sont correctement identifiés.

Le nommage des fichiers informatiques est extrêmement important. Le nom d'un fichier doit faciliter l'identification de son contenu et son repérage. L'identification complète des documents d'archives, toutefois, est plus complexe que le simple nommage des fichiers. Elle est essentielle pour distinguer les documents d'archives les uns des autres et les différentes versions d'un même document, ainsi que pour apporter la preuve de l'identité[533] d'un document depuis sa production et tout au long de son cycle de vie, notamment en cas de conservation sur le long terme.

Les informations sur les documents numériques qui contribuent à leur identification et à leur représentation sont communément appelées *métadonnées*[534]. La plupart des applications logicielles balisent automatiquement tous les documents numériques avec certaines données d'identification car ces informations sont nécessaires pour localiser les documents. En l'absence de métadonnées, il serait pratiquement impossible de trouver un document sans ouvrir et lire un dossier ou plusieurs répertoires. Les métadonnées décrivent les propriétés ou attributs des documents numériques. Dans le cas des documents d'archives, cependant, ces propriétés ou attributs sont également indispensables pour maintenir et évaluer leur authenticité ; c'est la raison pour laquelle il est important de s'assurer que toutes les propriétés ou tous les attributs essentiels sont enregistrés et exacts.

Les propriétés ou attributs qui permettent d'identifier les documents numériques sont appelés *métadonnées d'identification*[535].

Les métadonnées d'identification comprennent :

- a. *Le nom des personnes qui participent à la production des documents numériques*, notamment :
  - l'*auteur* – personne(s) physique(s) ou morale(s) responsable(s) de l'émission des documents ;
  - le *rédacteur* – personne(s) physique(s) ou service(s) responsable(s) de l'élaboration du contenu des documents ;
  - l'*expéditeur* – personne physique, fonction ou service responsable du compte électronique ou de l'environnement technique où les documents sont produits et/ou à partir duquel ils sont transmis[536] ;
  - le *destinataire/bénéficiaire* – personne(s) physique(s)ou morale(s) auxquelles les documents sont destinés (bénéficiaire) ou adressés (destinataire) ;
  - le *récepteur* – personne(s) physique(s) ou morale(s) recevant une copie ou une copie cachée des documents.
- b. *Le nom de l'action ou de l'affaire* – c'est-à-dire, le titre ou sujet.
- c. *La forme du document* – c'est-à-dire, s'il s'agit d'un rapport, d'une lettre, d'un contrat, d'un tableau, d'une liste, etc.
- d. *La présentation numérique* – c'est-à-dire, le format, le conteneur, l'encodage, etc.
- e. *La ou les date(s) de production et de transmission*, à savoir :
  - la *date* inscrite sur les documents ou à laquelle les documents ont été élaborés ;
  - les *dates de transmission et/ou de réception* ;

    - la *date d'enregistrement* – c'est-à-dire, la date à laquelle les documents ont été associés à un dossier ou un répertoire informatique, ou à tout autre système de classification ou de classement (voir la recommandation 5).
- f. *La mention du contexte documentaire* – par exemple, le code de classement ou le nom du dossier ou du répertoire informatique, ou d'une unité de classement comparable dans le système de classification ou de classement auquel les documents sont associés, ainsi que le nom de l'ensemble de documents d'archives auquel les documents appartiennent (voir également la recommandation 5).
- g. *La mention de pièces jointes*, s'il y a lieu.
- h. *La mention des droits d'auteur ou de tout autre droit de propriété intellectuelle*, s'il y a lieu.
- i. *La mention de la présence ou du retrait d'une signature numérique*, s'il y a lieu (voir la section Authentification basée sur la technologie de la recommandation 6).
- j. *La mention d'autres moyens d'authentification*, s'il y a lieu. Il peut s'agir, par exemple, d'une corroboration (c'est-à-dire l'énoncé des moyens utilisés pour valider le document d'archives), d'une validation (c'est-à-dire le fait d'authentifier un document d'archives par ceux qui ont participé à son établissement et par les témoins de l'action ou de la signature du document), d'une souscription (c'est-à-dire la mention du nom de l'auteur ou du rédacteur) ou d'une mention de la titulature des signataires (c'est-à-dire la mention du titre, de la qualité et/ou de l'adresse du ou de(s) signataire(s) du document).
- k. *La mention du numéro d'état préparatoire ou de version*, s'il y a lieu.
- l. *L'existence et l'emplacement des documents en double à l'extérieur du système numérique*, s'il y a lieu. S'il existe plusieurs exemplaires d'un document, il faut indiquer lequel est l'exemplaire officiel ou *faisant autorité*[537]. Si l'auteur a certifié le document comme « reproduction approuvée » d'une œuvre (par exemple, une œuvre d'art numérique), il faut mentionner l'existence de cette certification. Dans le cas où plusieurs auteurs détiennent des droits d'auteur sur le document, il convient d'indiquer si ces droits sont levés (ou non) en précisant les dates correspondantes.

4. S'assurer que les documents numériques contiennent les informations qui aideront à vérifier leur intégrité.

Bien que les métadonnées d'identification aident à distinguer des documents numériques les uns des autres, il existe un autre ensemble de métadonnées qui permet aux utilisateurs d'inférer que les documents n'ont pas été modifiés depuis leur production (mais pas de le vérifier ni de le démontrer, puisque cela nécessiterait de comparer les documents avec des copies conservées ailleurs). On appelle ces métadonnées les *métadonnées d'intégrité.* Un document numérique est dit *intègre*[538] lorsqu'il est intact et non altéré, c'est-à-dire si les messages qu'il est censé véhiculer pour accomplir sa fonction sont intacts. Cela signifie que l'intégrité physique d'un document numérique, par exemple le nombre de chaînes de bits, peut être altérée sous réserve que la structuration de son contenu et les éléments requis de sa forme (voir la recommandation 2) soient inchangés. Son contenu et les données qui le composent sont considérés comme intacts si leur valeur informative et leur présentation (c'est-à-dire leur affichage) sont identiques à celles qui étaient les leurs lors de l'enregistrement initial du document. Les attributs relatifs à l'intégrité des documents numériques ont trait à la maintenance des documents, et notamment aux conditions de leur manipulation ; il convient en particulier de veiller à ce qu'ils soient correctement manipulés, par exemple en surveillant et en documentant tout changement technologique ou transfert des documents dans d'autres systèmes. Les métadonnées d'intégrité comprennent :

- a. *Le nom de la personne ou du service métier* – la personne ou le service qui utilise les documents dans le cadre de ses activités ;
- b. *Le nom de la personne ou du service ayant la charge des documents* – il peut s'agir de la personne ou du service métier ;
- c. *La mention des additions faites aux documents*, s'il y a lieu ;
- d. *La mention de toute modification technique apportée aux documents ou aux applications en charge de la gestion et des autorisations d'accès aux documents*, par exemple changement d'encodage, de

conteneur ou de format logique, mise à niveau d'une application, transformation de plusieurs composants numériques en un seul (par exemple, en incorporant directement dans les documents des composants numériques qui leur étaient jusque-là seulement liés, tels que des composants audio, vidéo ou graphiques ou encore des éléments de texte comme les polices de caractères) ;

- e. *Le code de droits d'accès restreints* – indication de la personne, du poste ou du service autorisé à lire les documents, s'il y a lieu ;
- f. *Le code de droits d'accès étendus* – indication de la personne, du poste ou du service autorisé à annoter les documents, les supprimer ou les retirer du système, s'il y a lieu ;
- g. *Le signalement des documents d'archives essentiels* – indication du degré d'importance du document d'archives pour poursuivre l'activité pour laquelle il a été créé ou l'activité de la personne ou du service qui l'a créé, s'il y a lieu[539] ;
- h. Le *sort final* des documents – par exemple, leur retrait du système de production pour les stocker à l'extérieur de celui-ci, le versement à un service d'archives (voir la recommandation 10) ou leur élimination.

5. Organiser les documents numériques en ensembles logiques.

La gestion et l'extraction des documents numériques seront facilitées et améliorées en les traitant par lots, plutôt qu'individuellement. Il est donc important de les regrouper selon une certaine logique. On peut définir les catégories en fonction de la manière de travailler, des activités, des procédures utilisées par l'organisation, par thème ou encore en fonction de la structure de l'organisation. La première étape, particulièrement importante, consiste à séparer les documents d'archives des autres documents numériques. L'organisation des documents d'archives peut être fondée sur le type des documents ou leur durée de conservation. Selon les besoins, il est possible d'opter pour une organisation hiérarchique ou horizontale des groupes de documents entre eux. En règle générale, cette structure devra correspondre à l'organisation des documents d'archives papier (ou sur tout autre support), pour que tous les documents d'archives liés à la même activité ou au même sujet, ou de même type, puissent être facilement localisés et extraits dans l'ensemble auquel ils appartiennent.

Il est conseillé de créer un *plan de classement*[540], document qui détaille le système de classification, en donnant notamment une brève description des différents groupes de documents et des liens qui existent entre eux. Il est également recommandé d'attribuer à chaque ensemble de documents d'archives un code ou un nom, qui doit être lié à chaque document d'archives appartenant au même ensemble, indépendamment de leur support ou de leur emplacement. Ainsi, les documents d'archives d'un même ensemble auront le même code ou nom, suivi d'un chiffre indiquant leur ordre. Cet identifiant devra être enregistré dans les métadonnées d'identification des documents d'archives numériques et être inscrit sur la chemise rassemblant des documents d'archives papier appartenant au même ensemble. Il doit être unique pour chaque document d'archives.

Déterminer la durée de conservation des ensembles de documents d'archives facilitera leur gestion au quotidien et contribuera à garantir que les documents devant être conservés au titre d'archives définitives soient bien repérés comme tels suffisamment tôt et bénéficient de mesures de protection appropriées. L'expérience prouve qu'il est plus facile et plus efficace d'affecter une durée de conservation – la durée pendant laquelle on souhaite ou doit garder les documents – à un ensemble de documents, qu'à chaque document. Tenter de garantir que certaines choses seront conservées aussi longtemps qu'elles sont nécessaires tout en éliminant celles qui ne le sont plus est une tâche beaucoup trop lourde et complexe si l'on procède document par document. Même si certains documents d'un ensemble doivent être conservés plus longtemps que d'autres, conserver tout l'ensemble permet de gagner du temps mais aussi de disposer de renseignements plus complets lorsqu'on a besoin de consulter les documents. Cependant, pour certains types de documents d'archives, il est possible si on le souhaite de créer des sous-groupes à l'intérieur des ensembles en fonction des durées de conservation.

6. Utiliser des techniques d'authentification qui favorisent la maintenance et la conservation des

documents numériques.

L'authenticité des documents numériques est menacée chaque fois qu'ils sont transmis dans l'espace (c'est-à-dire, lorsqu'ils sont envoyés à un destinataire ou transmis d'un système ou d'une application à l'autre) ou dans le temps (c'est-à-dire, lorsqu'ils sont stockés ou lorsque le matériel informatique ou le logiciel utilisé pour leur stockage, leur traitement ou leur communication est mis à jour ou remplacé). Dans la mesure où l'archivage d'un document numérique et son extraction imposent nécessairement de lui faire franchir des frontières technologiques non négligeables (des sous-systèmes d'affichage aux sous-systèmes de stockage et vice-versa), la présomption d'authenticité des documents numériques doit être étayée par des preuves qu'ils ont été maintenus en utilisant des technologies et des procédures administratives qui garantissent la continuité de leur identité et de leur intégrité ou, à tout le moins, minimisent les risques de modification entre le moment où les documents ont été initialement archivés et celui de leur consultation.

Authentification indépendante de la technologie

Présomption d'authenticité. Une présomption d'authenticité est une déduction faite à partir de faits connus concernant la manière dont un document a été produit et maintenu. L'adoption et l'application des recommandations formulées dans le présent document constituent le meilleur moyen d'étayer cette présomption. Les recommandations ont un effet cumulatif : plus le nombre de recommandations satisfaites est grand et plus le degré atteint pour chacune d'elles est élevé, plus forte sera la présomption d'authenticité. Pour retirer les bénéfices escomptés des recommandations proposées dans ce document, il est indispensable de définir et d'appliquer des politiques et des procédures administratives efficaces[541]. Dans l'idéal, il faut utiliser des techniques d'authentification adossées à des politiques et procédures administratives et, autant que faire se peut, indépendantes ou neutres vis-à-vis de la technologie.

Authentification basée sur la technologie

Les techniques d'authentification basées sur la technologie, comme la cryptographie par exemple, permettent de disposer d'un mécanisme technologique garantissant l'authenticité des documents numériques. Une de ces techniques cryptographiques est la signature numérique, qui peut être utilisée pour la transmission de documents entre des personnes, des systèmes ou des applications afin de confirmer leur authenticité à un moment donné. Certains organismes, comme la Commission européenne et la *Securities and Exchange Commission* aux États-Unis, ont reconnu la valeur légale ou réglementaire de ces technologies.

Une mise en garde s'impose toutefois. Les signatures numériques elles-mêmes peuvent devenir obsolètes et, par essence, il est impossible de les faire migrer avec les documents dont elles font partie vers des applications logicielles nouvelles ou mises à jour. De fait, la durée de vie des signatures numériques et des autres technologies d'authentification est parfois beaucoup plus courte que celle requise même pour un document temporaire dont la migration n'est pas nécessaire, à cause de l'évolution rapide des technologies d'authentification. Sauf à ce que de nouvelles technologies de signature numérique ne permettent de conserver durablement ces informations d'authentification avec le document concerné, il est recommandé, lors de la réception d'un document comportant une signature numérique, de la retirer lorsque c'est possible et de compléter les métadonnées d'intégrité pour indiquer que le document comportait une signature numérique lors de sa réception et que celle-ci a été vérifiée, retirée et supprimée.

7. Protéger les documents numériques d'actions non autorisées.

On ne peut présumer l'exactitude et l'authenticité des documents numériques s'il est possible de les modifier sans laisser de trace. Il faut être en mesure d'établir qu'il est impossible à quiconque d'altérer ou de manipuler les documents sans que la personne soit identifiée. Parmi les mesures de sécurité envisageables, on retiendra plus particulièrement la restriction de l'accès physique aux lieux où les ordinateurs se trouvent et la restriction

de l'accès aux documents numériques sur les ordinateurs eux-mêmes. Cette deuxième mesure peut être mise en place par divers moyens, notamment l'utilisation de mots de passe ou l'authentification biométrique pour accéder au système.

Il est également important de définir une hiérarchie des autorisations d'accès (également appelées « droits d'accès » – voir l'exposé sur les *métadonnées d'intégrité* à la recommandation 4) pour tous les utilisateurs du système. Par exemple, certains utilisateurs seront seulement autorisés à lire les documents, alors que d'autres pourront les modifier. Dans tous les cas, il devrait être impossible de modifier un document une fois que celui-ci a été classé selon le *plan de classement* (voir les recommandations 3 et 5), et seule la personne à laquelle a été confiée la responsabilité de l'archivage et de la maintenance des documents d'archives doit pouvoir transférer ou supprimer des documents du système. De plus, le système doit permettre de tenir à jour un historique de l'accès aux documents, afin de contrôler l'administration et l'utilisation des droits d'accès.

Cette recommandation pourra sembler difficile à mettre en œuvre par les personnes qui travaillent à domicile ou même par les petites structures. Toutefois, il est important de se souvenir que si on ne peut démontrer qu'il est impossible d'altérer et de manipuler les documents numériques sans être identifié, toute déclaration quant à leur exactitude et leur intégrité est sans valeur. Il peut donc être utile de conserver hors du système informatique des copies des documents numériques, au moins pour les plus importants d'entre eux, et de mettre en place des mesures systématiques et périodiques pour comparer de manière aléatoire les documents stockés hors du système informatique à leurs équivalents en ligne.

8. Protéger les documents numériques des pertes et altérations accidentelles.

Les ordinateurs sont vulnérables, et de nombreux facteurs peuvent entraîner la corruption des documents d'archives ou des données, ou leur perte accidentelle. Le meilleur moyen de se prémunir contre ces aléas est de faire régulièrement et souvent des copies de sauvegarde des documents. Il est conseillé de conserver ces copies ailleurs que dans les locaux habituels, protection supplémentaire contre l'incendie ou le vol d'équipement. Il existe un grand nombre de techniques, de progiciels et de services de sauvegarde, dont certains génèrent automatiquement des sauvegardes et les transfèrent en lieu sûr, sur un autre site.

- a. *Mettre en place une politique ou une procédure rigoureuse qui assure la sauvegarde quotidienne du système.* Un système ne vaut que par la dernière sauvegarde qu'on en a fait. Il doit donc être fréquemment sauvegardé, au moins une fois par jour, à l'aide de méthodes éprouvées qui permettront de le remettre rapidement en état en cas de défaillance. Ces sauvegardes devront être supprimées à tour de rôle selon la stratégie ou le rythme le mieux adapté au contexte, car elles ne contiennent aucun document d'archives et ne servent qu'à la remise en état du système en cas de défaillance. Nous parlons ici d'une sauvegarde complète du système, englobant le système d'exploitation, les applications logicielles et tous les documents numériques du système. Si, en plus de la sauvegarde du système, il y a besoin d'une copie de sécurité des documents numériques pour le cas où le matériel serait volé ou si certains documents d'archives venaient à être corrompus, il est recommandé de sauvegarder ces documents sur un autre ordinateur, un disque dur externe ou un autre support numérique portable et de stocker ces copies de sécurité ailleurs que dans les locaux où se trouve l'ordinateur qui contient les originaux.
- b. *Choisir et installer la technologie de sauvegarde la plus adaptée.* La comparaison des technologies et services proposés sur le marché permettra de choisir la solution la plus adaptée au contexte. Il existe de nombreux systèmes, des plus simples aux plus sophistiqués, capables de sauvegarder de très gros systèmes. Le système de sauvegarde doit comporter un journal des événements au cas où une défaillance du système surviendrait entre les sauvegardes, et qu'il faille récupérer les documents d'archives ou d'autres documents numériques créés pendant la période qui n'a fait l'objet d'aucune sauvegarde.

9. Prendre des mesures contre l'obsolescence du matériel informatique et des logiciels.

La rapidité avec laquelle le matériel informatique et les logiciels deviennent obsolètes pose de sérieux problèmes pour la maintenance et la conservation sur le long terme des documents numériques. Une manière d'y remédier consiste à éliminer la dépendance à l'égard du matériel informatique en transférant les fonctionnalités de celui-ci à un logiciel (c'est-à-dire, en utilisant une application logicielle qui simule les actions d'une partie du matériel). Ceci constitue une manière plus stable de conserver les fonctions lorsque les matériels deviennent obsolètes.

En raison de l'évolution rapide de l'environnement technologique, individus et services doivent régulièrement mettre à niveau leurs systèmes informatiques ainsi que les documents d'archives présents dans ces systèmes et ceux qui ont été transférés sur un support de stockage, comme un CD, un DVD ou une bande magnétique. En d'autres termes, lorsque certains éléments de l'environnement technologique commencent à être obsolètes, ils doivent être mis à niveau en les faisant passer à la technologie la plus avancée disponible sur le marché en fonction des besoins et des contraintes de l'organisation, et tous les documents numériques à l'intérieur ou à l'extérieur du système doivent être migrés vers la nouvelle technologie. Lors du remplacement du matériel informatique, il est essentiel que le nouveau matériel possède des capacités au moins équivalentes à celui qu'il remplace. Par exemple, un nouveau moniteur doit pouvoir afficher un document d'archives graphique en conservant la forme documentaire du document d'archives original. La planification de mises à niveau technologiques régulières, par rotation, contribuera à prévenir l'obsolescence et à éviter des dépenses importantes et imprévues.

Il arrive parfois que des documents d'archives numériques produits ou maintenus par des systèmes qui deviennent obsolètes doivent être conservés pendant longtemps, même si l'on ne prévoit pas d'y accéder souvent. Si ces documents d'archives sont des documents textuels devant être lus dans un ordre donné et non de manière aléatoire, on peut envisager de les transférer sur microfilm à partir de leur forme numérique. C'est sans conteste le meilleur moyen de les mettre à l'abri des altérations ou pertes accidentelles. La duplication est une autre bonne mesure de protection. On crée une copie des ensembles de documents d'archives essentiels et on les conserve sur un autre ordinateur, un deuxième disque dur, un DVD, en les confiant à une autre personne ou un autre service ou en les stockant à distance du site. Lorsque des documents d'archives ou d'autres objets numériques sont supprimés d'un système informatique opérationnel pour être stockés sur un support magnétique ou optique, par exemple, il est essentiel de retirer également la documentation relative au système et aux documents numériques (par exemple, les métadonnées des documents d'archives) et de les conserver avec les documents d'archives (voir les sous-sections D, E et F de la recommandation 1).

10. Prendre en compte les enjeux de la conservation sur le long terme.

Bien que le présent document s'intéresse à la production de documents numériques en tous genres et à leur maintenance pendant tout le temps où leurs producteurs en auront besoin régulièrement, il est essentiel d'envisager également la conservation sur le long terme des documents numériques importants. En règle générale, seul un petit pourcentage de documents doit être conservé sur le long terme, mais la capacité de prendre en charge l'archivage des documents sur le long terme, et en particulier des documents numériques, excède souvent les moyens des individus ou des petites structures – lorsqu'ils s'en préoccupent, ce qui est loin d'être toujours le cas. Bien que les coûts soient bien réels, tant financiers qu'humains, la conservation sur le long terme des documents est essentielle à la constitution et à la pérennisation de notre patrimoine, à l'exercice des droits et obligations de chacun et à la prise de décisions éclairées.

Pour amorcer ce processus, la première chose à faire est de désigner une personne qui prendra en charge les documents numériques une fois qu'ils ne seront plus requis dans le cours habituel des activités de l'organisation. Cette personne assumera le rôle de *responsable des archives*[542]. Le responsable des archives est un professionnel – ou un groupe de professionnels comme dans un service d'archives ou une société historique

par exemple – formé à l'archivage et à la conservation de documents et qui, idéalement, n'est pas directement concerné par le contenu des documents d'archives et n'a aucun intérêt à ce que d'autres personnes les manipulent ou les détruisent. Pour ce qui est des structures de petite taille, cette personne pourrait être la personne en charge de la gestion des documents d'archives pendant la période où ils sont d'une utilisation courante. Quant aux individus qui gèrent eux-mêmes leurs documents, ils pourront soit prendre eux-mêmes en charge leur conservation soit la confier à un archiviste, un documentaliste ou à toute autre personne compétente en la matière, dans le respect de la législation en vigueur. Dans un cas comme dans l'autre, il convient de définir une politique de conservation le plus tôt possible, car les documents numériques dont la conservation n'a pas été anticipée ne seront pas préservés. Le respect des principes exposés dans le présent document facilitera donc la conservation sur le long terme.

Conclusion

Ce document propose un certain nombre de mesures et de méthodes auxquelles les individus et les petites structures peuvent avoir recours pour produire et maintenir des documents numériques présumés authentiques, exacts et fiables. Pour les individus, la tâche peut sembler bien lourde, mais ne rien faire c'est prendre le risque de perdre des documents d'archives ou d'être confronté à des données corrompues et invérifiables, avec des conséquences bien plus graves à long terme. Les petites structures auront tout intérêt à désigner officiellement la ou les personnes chargées de superviser la maintenance de leurs documents d'archives numériques. Il ne faut cependant pas perdre de vue qu'il n'est pas toujours nécessaire de mettre systématiquement en œuvre toutes les recommandations formulées dans le présent document. Il appartient à chacun de sélectionner et d'adopter les mesures les plus pertinentes en fonction des problèmes rencontrés. Dans certains cas, il est possible que des mesures supplémentaires soient requises en raison des exigences législatives ou réglementaires propres au domaine ou aux caractéristiques des activités et, partant, des documents d'archives qui en sont issus. Dans ces cas, il pourra s'avérer nécessaire de se tourner vers des spécialistes, comme par exemple les services d'archives de sa ville, de sa région ou des archives nationales, ou encore vers des associations professionnelles d'archivistes. D'une manière générale, il est fortement recommandé de prendre contact avec ces spécialistes pour toute question relative à la production et à la maintenance de documents numériques.

Enfin, ces principes directeurs ne sont qu'un des documents élaborés et diffusés par le projet InterPARES, projet de recherche international sur la préservation sur le long terme des documents d'archives numériques authentiques. On trouvera sur le site web d'InterPARES, www.interpares.org, une abondante documentation sur la nature des documents d'archives numériques et l'élaboration de méthodes en vue de produire, maintenir et conserver des documents d'archives fiables, exacts et authentiques.

---

[526] *Norme* : standard adopté par un organisme de normalisation officiel que ce soit au niveau national (par exemple, American National Standards Institute [ANSI], multinational (par exemple, le Comité européen de normalisation [CEN] ou international (par exemple, ISO). Pour les formats de fichier informatique, deux normes ont été adoptées récemment : le PDF/A (norme PDF pour les archives) et l'ODF (format OASIS OpenDocument).

[527] *Standard* : ensemble de recommandations qui n'est pas publié par un organisme de normalisation mais qui est néanmoins largement utilisé et considéré par ses utilisateurs comme une norme. C'est par exemple le cas des formats de fichier PDF, TIFF, DOC et ZIP.

[528] *Forme fixe* : qualité d'un document d'archives dont l'apparence ou la présentation est la même chaque fois qu'il est représenté.

[529] *Contenu stable* : qualité d'un document d'archives qui le rend immuable et dont il découle qu'il n'est possible de le modifier qu'en lui joignant une mise à jour ou en en créant une nouvelle version.

[530] *Forme documentaire* : règles de représentation du contenu du document d'archives, de ses contextes administratif et documentaire et de son autorité. La forme comprend des caractères externes et des caractères internes.

[531] *Variabilité limitée* : qualité d'un document d'archives dont les présentations documentaires sont limitées et contrôlées par des règles fixes et un ensemble stable de données de contenu, de forme et de composition, de sorte que la même

action, interrogation, requête ou interaction de l'utilisateur produira toujours le même résultat.

[532] *Caractères internes* : éléments d'un document d'archives qui communiquent l'action à laquelle il participe et son contexte immédiat. Sont des caractères internes le nom des personnes qui ont participé à sa production, le nom et la description de l'action ou du sujet auquel il se rapporte, la ou les date(s) de production et de transmission, etc. *Caractères externes* : éléments d'un document d'archives qui déterminent son aspect extérieur. Sont des caractères externes la police de caractères, les graphiques, les images, le son, la mise en page, les hyperliens, la résolution des images, etc., ainsi que les signatures, les sceaux numériques, les horodatages ainsi que les signes de validation (filigranes numériques, logos, emblèmes, etc.).

[533] Ici, l'identité est définie comme l'ensemble des attributs d'un document ou d'un document d'archives qui l'identifient et le distinguent de manière unique de tout autre document ou document d'archives. Avec l'intégrité, c'est un des composants de l'authenticité (voir également la recommandation 4).

[534] *Métadonnées* : informations qui qualifient une autre ressource d'information, en particulier en vue de documenter, décrire, conserver ou gérer cette ressource.

[535] *Métadonnées d'identification* : propriétés ou attributs qui permettent d'identifier un objet numérique destiné à être conservé comme document d'archives.

[536] L'identification de l'expéditeur n'est importante que lorsque la personne, la fonction ou le service responsable de la production et/ou de la transmission des documents n'en est ni l'auteur ni le rédacteur, et lorsque la présence du nom de l'expéditeur sur les documents ou en association avec eux, fait peser un doute sur l'identité de leur véritable auteur et/ou rédacteur. Cette situation s'observe le plus souvent avec les courriels où le nom de l'expéditeur est inscrit dans l'en-tête ou avec les pièces jointes d'un courriel dont l'auteur et/ou le rédacteur est quelqu'un d'autre, mais qui est communiqué et/ou transmis matériellement au nom de cette personne par l'expéditeur.

[537] *Exemplaire faisant autorité* : exemplaire d'un document d'archives considéré par le producteur comme son document de référence et généralement assujetti à des contrôles procéduraux qui ne sont pas requis pour les autres exemplaires.

[538] *Intégrité* : qualité de ce qui est complet et inchangé dans ses aspects essentiels. Avec l'identité, un des composants de l'authenticité.

[539] Cela concerne par exemple certaines professions juridiques et médicales, qui doivent identifier les documents d'archives essentiels à la continuité de leurs activités en cas de catastrophe et qui, en conséquence, prendront des mesures de protection spéciales à l'égard de ces documents.

[540] *Plan de classement* : système d'identification et d'organisation des activités métier et des documents d'archives par catégories, selon des conventions, méthodes et règles procédurales logiquement structurées (voir également la recommandation 5).

[541] Voir l'annexe 19, « Principes pour l'élaboration de politiques, de stratégies et de normes pour la conservation sur le long terme des documents d'archives numériques. »

[542] *Responsable archives* : entité ou individu en charge des documents d'archives pouvant démontrer qu'elle/il n'a pas de raison de modifier les documents d'archives conservés ou d'autoriser d'autres à le faire, et qu'elle/il est capable de mettre en œuvre toutes les conditions requises pour la conservation de copies authentiques de ces documents.