

Préservation des courriers électroniques

Collection de notes d'orientation sur les types de données

Artefactual Systems et Digital
Preservation Coalition



**Veille technologique de la
DPC – Note d'orientation**

Juillet 2021



Digital Preservation Coalition

Informations sur la traduction française

La traduction française de cette note d'orientation a été réalisée dans le cadre des travaux de la Cellule nationale de veille sur les formats entre novembre 2022 et avril 2023. Lancée en 2019, la Cellule nationale de veille sur les formats (CNVF), sous l'égide de l'association française Aristote et de son groupe de travail sur la Pérennisation de l'information numérique (PIN), regroupe à ce jour une douzaine de partenaires. Ses principaux objectifs sont la mutualisation des activités de veille sur les formats, la sensibilisation des professionnels sur le sujet, la contribution ou l'influence sur les outils associés. Elle a pour ambition d'être un interlocuteur francophone reconnu dans les travaux internationaux sur ces sujets.

La note d'orientation a été traduite en français par : Édouard Vasseur (École nationale des Chartes). Une première traduction automatique a été réalisée avec le logiciel DeepL puis a été revue et corrigée par le groupe de travail.

Collection de notes d'orientation sur les types de données

Chaque note d'orientation de la collection « Types de données » est conçue pour fournir une introduction à l'état actuel des connaissances de la communauté sur les types de données couramment rencontrés par ceux qui cherchent à préserver les collections numériques. La préservation numérique consiste à faire en sorte que l'information reste trouvable, utilisable et fiable à long terme. La meilleure approche en matière de dépôt numérique varie en fonction du périmètre et du contenu de ses collections, des ressources dont il dispose et des attentes de ses financeurs et utilisateurs. Il existe cependant de bonnes pratiques largement applicables qui ont été définies à la suite de nombreuses années de recherche, de mise en œuvre concrète et d'élaboration de consensus. Elles sont présentées ici comme un point de départ, et comprennent des ressources complémentaires pour approfondir les recherches.

Cette collection de notes d'orientation sur les types de données a été rédigée par le personnel d'Artefactual Systems en collaboration avec Digital Preservation Coalition. Ces notes ont été élaborées en collaboration avec l'organisme britannique Nuclear Decommissioning Authority.

La préservation numérique est un domaine en pleine évolution qui se développe en réponse à des facteurs externes et des nouveaux défis. De nouveaux formats, normes et exemples de bonnes pratiques apparaîtront au fil du temps ; ainsi, les informations contenues dans ce rapport devront être mises à jour. Nous vous invitons à nous faire part de vos commentaires et de vos réactions à l'adresse suivante : info@dpconline.org.

1 Présentation des types de données

La messagerie électronique est un système permettant l'échange asynchrone de messages entre ordinateurs. Un courrier électronique est un document textuel structuré qui comprend généralement un en-tête et un corps, peut intégrer des images ou d'autres objets, et peut inclure des pièces jointes. Les courriers électroniques sont généralement structurés à l'aide des normes *Internet Message Format* (IMF) ([RFC 5322](#), 2008) et *Multipurpose Internet Mail Extensions* (MIME) ([RFC 2405](#), 1996).

Un seul courrier électronique peut exister dans de nombreuses instances et les courriers électroniques peuvent avoir des relations complexes les uns avec les autres, car ils font souvent partie de fils de discussion (groupes de messages contenant des réponses et des transmissions à d'autres destinataires). Les en-têtes des courriers électroniques comprennent des informations visibles par l'utilisateur, telles que l'expéditeur, le(s) destinataire(s) et l'objet du message, ainsi que des informations généralement cachées à l'utilisateur, telles que les informations de routage et d'authentification, et les horodatages. Dès qu'un courrier électronique arrive dans la boîte aux lettres d'un destinataire, des métadonnées supplémentaires peuvent être générées automatiquement, comme des drapeaux indiquant le statut « Lu » ou « Non lu », le niveau d'importance ou la nécessité d'un suivi. Les utilisateurs peuvent ajouter des étiquettes à certains groupes de messages et organiser les messages ou les discussions dans des dossiers. Ces facteurs, ainsi que la possibilité de conserver des copies des messages envoyés et d'ajouter des pièces jointes aux messages, font des comptes de messagerie des systèmes d'archivage (*recordkeeping systems*) de facto. Pour de nombreuses organisations, les comptes de messagerie peuvent constituer l'une des plus importantes sources documentaires de preuves des activités et des transactions, et ont donc une valeur archivistique durable ([InterPARES](#), 2009, p.5).

2 Défis en matière de préservation

Le courrier électronique est un type de données complexe constitué de multiples discussions impliquant de nombreuses personnes différentes. Il peut inclure des pièces jointes (qui peuvent consister en tout autre type de données ou format de fichier) ; il peut être de grande taille et il peut être difficile à capturer efficacement. Ces difficultés sont décrites plus en détail ci-dessous :

2.1 Pièces jointes

L'introduction du standard MIME dans les années 1990 a normalisé la manière dont les fichiers numériques peuvent être joints aux courriers électroniques. Bien qu'en théorie, des fichiers de n'importe quelle taille et de n'importe quel format puissent être joints aux messages, certains programmes de messagerie appliquent des limites à la taille et au nombre de fichiers pouvant être transmis, et excluent certains types de fichiers tels que le code exécutable (Tech Terms, 2015). Il peut être nécessaire de préserver les pièces jointes aux messages en même temps que les messages auxquels elles ont été jointes. La préservation d'un compte de messagerie peut donc nécessiter la préservation de volumes potentiellement importants d'autres fichiers numériques aux formats variés. Les courriers électroniques et leurs pièces jointes doivent être préservés de manière à maintenir les relations entre eux.

2.2 Extraction et transformation

Dans le rapport de veille technologique sur les courriers électroniques de la *Digital Preservation Coalition*, Chris Prom mentionne que « le premier problème, et à certains égards le plus pressant pour ceux qui veulent préserver les courriers électroniques, est de savoir comment les collecter en premier lieu » ([Prom](#), 2019, p.19). Les stratégies de capture peuvent varier en fonction de la façon dont les organisations, ou même des individus particuliers au sein des organisations, gèrent et stockent les

courriers électroniques. La plupart des systèmes de messagerie électronique intègrent des options d'export. Cependant, dans certains cas, les archivistes n'ont pas d'accès direct à un compte de messagerie, et il est alors nécessaire d'utiliser des outils spécialisés (tels que [OfflineIMAP](#)) pour récupérer les courriers électroniques. Tout effort de préservation des courriers électroniques nécessite une certaine forme de transformation lors de leur extraction du système d'origine ; il n'existe pas de format de préservation « natif » ([groupe de travail EA-PDF, 2021¹, p.7](#)).

2.3 Questions juridiques

Le corps d'un courrier électronique est un texte non structuré qui peut contenir des données permettant d'identifier personnellement l'utilisateur, définies au sens large comme « toute information relative à une personne physique identifiée ou identifiable » ([RGPD, Définitions, 2016](#)). Les pièces jointes aux messages peuvent également contenir des données d'identification ou d'autres données personnelles sensibles. Le droit d'auteur peut également être source de complexité, car le propriétaire de la messagerie reçoit des messages dont il n'est pas l'auteur. La réutilisation et la conservation peuvent être affectées par différents cadres juridiques. La réutilisation peut par exemple être régie par le Règlement général sur la protection des données ([RGPD](#)), la loi allemande Bundesdatenschutzgesetz ([BDSG](#)), la loi britannique sur la protection des données ([DPA](#)) ou la loi américaine Federal Trade Commission Act ([FTCA](#)). Les comptes de messagerie peuvent contenir des volumes importants de données d'identification et de données sensibles selon la nature de l'organisation (banques, cabinets médicaux ou certains types d'organismes gouvernementaux) ou la fonction de la personne (personnel des ressources humaines, cadres ou responsables syndicaux). Les données d'identification et les données personnelles sensibles peuvent également apparaître accidentellement lorsque l'utilisateur enfreint les politiques organisationnelles régissant l'utilisation de la messagerie électronique, parce qu'il n'existe pas de telles politiques ou parce que l'utilisateur utilise un compte de messagerie électronique à la fois à des fins professionnelles et personnelles. Les particuliers qui font don de leurs comptes de messagerie aux archives peuvent avoir utilisé leur messagerie pour correspondre avec un certain nombre d'autres particuliers qui ne savent pas que leurs échanges sont conservés.

2.4 Volume et volatilité

Des centaines de milliards de courriers électroniques sont envoyés et reçus chaque jour dans le monde ([Radicati, 2021 ; Statista, 2021](#)). L'utilisation du courrier électronique est massive dans de nombreuses organisations, et plusieurs personnes d'une même organisation peuvent être mises en copie sur des fils de discussion, si bien que de nombreux doublons d'un courrier électronique (et de ses pièces jointes) peuvent se trouver dans plusieurs comptes de messagerie. De nouveaux fils de discussion peuvent également se détacher des fils originaux, avec des destinataires ajoutés et supprimés au cours de la durée de vie du fil. Selon les politiques institutionnelles de gestion des documents d'activité (*records management*), les utilisateurs peuvent être libres de supprimer les courriers électroniques et les fils de discussion, même si certains suppriment des courriers électroniques de leur boîte de réception sans supprimer les messages ou fils de discussion correspondants des boîtes d'envoi. Les messages « supprimés » peuvent ne pas l'être en réalité, même s'ils semblent l'être du point de vue de l'utilisateur, et peuvent être récupérés dans le dossier « Corbeille » par l'utilisateur pendant une certaine période. Ce type de volatilité peut rendre l'évaluation et la collecte des courriers électroniques exceptionnellement difficiles. « [Qu'est]-ce qui est considéré comme la copie du document engageant (*record copy*) ? Le message qui a été composé et envoyé ? Celui qui a été reçu et stocké dans le dossier « Éléments envoyés » de l'application client de l'expéditeur ? La copie stockée

¹ Le lien du document d'origine était obsolète et a été remplacé par une redirection fonctionnelle vers le site de la bibliothèque de l'Université de l'Illinois.

sur le système de serveur de messagerie électronique du destinataire ? Une copie téléchargée ? Une copie stockée localement dans un fichier PST ? Une copie sur un iPhone ? », telles sont les questions d'un récent rapport du groupe de travail sur la messagerie électronique ([Council on Library and Information Resources \(CLIR\)](#), 2018, p.30).

3 Formats de messages typiques

Il n'existe pas de format unique parfait pour la préservation et l'utilisation future des courriers électroniques. Les décisions prises concernant les formats de fichiers doivent dépendre des caractéristiques et des fonctionnalités à préserver et des cas d'utilisation future à prendre en charge. Notez que le tableau ci-dessous décrit des formats de transfert typiques pour les archives de courrier électronique qui peuvent également être des formats de préservation et/ou d'accès appropriés, mais ne doit pas être considéré comme une liste exhaustive de formats. Il est recommandé d'effectuer des recherches et des analyses approfondies avant d'extraire des messages ou de les migrer vers un nouveau format.

Formats de fichier	Extensions	Brève description
Internet Message Format	.eml	<p>EML est un format texte conçu pour se conformer à la spécification originale RFC du courrier électronique (RFC822, 1982) et à ses successeurs. Les fichiers EML sont généralement des courriers électroniques uniques avec des pièces jointes intégrées au fichier sous forme de contenu encodé en base64 ou référencées à partir du fichier. Les messages (mais ce n'est généralement pas le cas des pièces jointes) peuvent être ouverts et lus dans n'importe quel éditeur de texte. Le format EML peut être exporté et restitué par de nombreux programmes de messagerie, notamment Microsoft Outlook, Zimbra, Eudora, Mozilla Thunderbird, Apple Mail et Gmail. Il existe également une bonne prise en charge des outils permettant de convertir les messages exportés dans d'autres formats en format EML. Une liste utile d'outils utilisés pour capturer, convertir et préserver les messages sous le format EML et d'autres formats est fournie par le groupe de travail Task Force on Technical Approaches to Email Archives (Email Archiving Tools, 2018²).</p> <p>EML est basé sur les standards RFC, mais il n'existe pas de spécification pour le format lui-même, et sa capacité à gérer les fils de discussion, les pièces jointes et d'autres caractéristiques relatives à la structure interne des comptes de messagerie est limitée (EA-PDF Working Group, p.26³). Comme le format EML est basé sur du texte et qu'il est omniprésent, il pourrait être facile à restituer à moyen terme. Certains grands projets de préservation de courriers électroniques, comme la plate-forme EASi de</p>

² Le lien du document d'origine est obsolète. Voir le rapport final du groupe de travail à l'adresse suivante : [The Future of Email Archives • CLIR](#) (consulté le 23 juin 2023).

³ Le lien du document d'origine était obsolète et a été remplacé par une redirection fonctionnelle vers le site de la bibliothèque de l'Université de l'Illinois.

Formats de fichier	Extensions	Brève description
		<p>l'université de Harvard, font migrer d'autres formats vers EML pour les préserver.</p> <p>Les messages EML peuvent être extraits des comptes de messagerie et stockés sous forme Maildir, un format de répertoire structuré pour représenter l'état de livraison des messages. Le format Maildir peut être exporté et importé dans un certain nombre d'applications de messagerie ; cependant, il manque de spécifications détaillées et présente certains problèmes de compatibilité avec les systèmes de fichiers qui peuvent nécessiter une correction avant que les fichiers puissent être lus par certains programmes de messagerie (Wikipedia, Maildir, 2020). L'outil Open Source OfflineIMAP peut être utilisé pour extraire des comptes de messagerie à distance et les stocker sous Maildir.</p> <p>Les deux organismes, Bibliothèque et Archives Canada et National Archives and Records Administration (agence américaine), considèrent les fichiers EML comme le format de transfert préféré pour le courrier électronique (LAC, 2015 ; NARA, 2020).</p>
MBOX	.mbox	<p>Comme EML, MBOX est un format texte conçu pour mettre en œuvre les normes de messagerie RFC. Contrairement à EML, un seul fichier MBOX peut contenir un agrégat de messages, généralement le contenu d'un dossier. MBOX est bien pris en charge par de nombreux clients de messagerie ; un utilisateur peut exporter et stocker un compte sous la forme d'une série de fichiers MBOX, chacun représentant un dossier de messages avec des pièces jointes intégrées sous forme de texte encodé en base64. Ces fichiers peuvent être importés dans d'autres clients de messagerie ou lus dans ePADD, un outil Open Source conçu par l'université de Stanford pour prendre en charge « l'évaluation, le traitement, la préservation, la découverte et l'accès à des messageries constituant des archives historiques » (Stanford Libraries, 2021). Le projet RATOM (Review, Appraisal, and Triage of Mail) de l'Université de Caroline du Nord a également développé des outils pour travailler avec les fichiers MBOX.</p> <p>Un seul fichier MBOX peut encapsuler un dossier entier de courriers électroniques, ce qui donne lieu à des relations plus solides entre les courriers électroniques et les fils de discussion, et facilite le chargement du dossier dans un client de messagerie ou un outil comme ePADD. Cependant, en raison du volume important de messages qui peuvent résider dans un seul dossier, les fichiers MBOX peuvent souvent devenir très volumineux et peu maniables ; cela signifie également que, dans certains cas, la corruption d'un seul message peut empêcher le chargement ou l'ouverture de l'ensemble du fichier (EA-PDF Working Group⁴, p. 26 ;</p>

⁴ *Ibid.*

Formats de fichier	Extensions	Brève description
		<p>DPC et Prom, 2019, p.17). MBOX en tant que format de préservation souffre également du fait qu'il s'agit d'une famille de formats comportant quatre variantes (MBOXO, MBOXRD, MBOXCL et MBOXCL2) qui ne sont pas entièrement compatibles entre elles (Library of Congress (LC), 2016). Malgré ces inconvénients, les fichiers MBOX sont très prometteurs en tant que format de préservation et d'accès, étant donné l'omniprésence du format, l'existence d'outils de prise en charge, la capacité de restituer des dossiers de messages entiers et leur structure textuelle sous-jacente.</p> <p>Les deux organismes, Bibliothèque et Archives Canada et National Archives and Records Administration (agence américaine) considèrent MBOX comme un format de transfert de messages privilégié (LAC, 2015 ; NARA, 2020).</p>
Dossiers personnels Microsoft Outlook	.pst .olm	<p>PST est un format propriétaire développé par Microsoft pour prendre en charge le téléchargement, le stockage et l'utilisation hors ligne des courriers électroniques. Il existe deux versions, ANSI (la plus ancienne des deux) et Unicode, qui « se différencient principalement par les versions de mise en œuvre des logiciels, les jeux de caractères, les contraintes de taille maximale des fichiers et les valeurs binaires » (LC, 2013). Un fichier PST peut correspondre à un compte de messagerie entier ou à une partie de celui-ci, et peut inclure non seulement des messages mais aussi des contacts, des éléments de calendrier, des tâches et d'autres produits d'Outlook et d'outils connexes.</p> <p>Microsoft maintient une spécification librement disponible pour PST (Microsoft, [MS-PST], 2020). L'omniprésence de Microsoft Outlook et des outils connexes fait de PST un format relativement courant. L'organisme Bibliothèque et Archives Canada considère PST comme un « format de transfert » acceptable, les formats préférés étant MBOX et EML, tandis que l'agence américaine National Archives and Records Administration considère ces trois formats comme préférés (LAC, 2015 ; NARA, 2020). La boîte à outils du projet RATOM fonctionne avec les fichiers PST en plus de MBOX.</p> <p>Attention : dans les mises en œuvre Mac de Microsoft Outlook, le format d'export est appelé Outlook pour Mac (OLM). OLM n'a pas de spécification ouvertement documentée comme PST et est sujet à des changements non identifiés (Microsoft, 2015). Il existe un certain nombre de convertisseurs OLM vers PST gratuits et payants, tels que Softaken et Mailvita, mais il est conseillé de tester les outils et de s'assurer de leur qualité avant de choisir.</p>

Formats de fichier	Extensions	Brève description
Portable Document Format	.pdf	<p>Lorsque des courriers électroniques uniques ou des fils de discussion sont convertis en format PDF ou PDF/A, ils perdent une grande partie de leurs métadonnées et les pièces jointes peuvent devenir inaccessibles (CLIR, 2018, p. 82-83). Cependant, un groupe de travail dédié à la préservation des messages a récemment proposé une spécification « pour préparer le terrain aux développeurs de logiciels afin de créer des systèmes de capture et de représentation des courriers électroniques tirant parti du format PDF pour représenter les métadonnées de base, le contenu, les attributs et le contexte » des comptes de messagerie (EA-PDF Working Group, 2021⁵, p.7). Les auteurs affirment que MBOX et EML « sont en fait des familles de formats quelque peu informels définis davantage par les mises en œuvre des clients que par des spécifications faisant autorité » et que « le PDF fournit un moyen de représenter ces mises en œuvre dans un format conteneur normalisé, quelle que soit la source sous-jacente ». (Groupe de travail EA-PDF, 2021⁶, p.8).</p> <p>Les versions PDF des courriers électroniques sont très accessibles en raison de la prise en charge généralisée du format par les logiciels, et les efforts du groupe de travail EA-PDF montrent comment le format PDF pourrait être amélioré pour servir à la fois de format de préservation et d'accès pour des comptes de messagerie entiers plutôt que pour des messages individuels.</p>
Extensible Markup Language	.xml	<p>Le format XML a été étudié comme format de préservation par un certain nombre de projets de préservation numérique, notamment PeDALS, CERP et DArcMail. DArcMail est le plus récent de ces projets et le successeur de CERP. DArcMail, qui signifie <i>Digital Archives for Email</i>, est géré par les Smithsonian Institution Archives, et est le nom du projet et d'un outil Open Source basé sur python disponible sur la page d'accueil du projet. DArcMail convertit les fichiers MBOX en format XML et comprend une interface utilisateur graphique qui « prend en charge l'évaluation, le traitement et l'accès aux comptes de messagerie ou à de grands groupes de courriers électroniques connexes » (Smithsonian Institution Archives, 2017, p. 3). Le fichier XML peut représenter un compte de messagerie entier, en conservant la structure interne des dossiers, les en-têtes et autres métadonnées ainsi que les messages et leurs pièces jointes. Si les Smithsonian Institution Archives reçoivent des messages dans des formats autres que MBOX, elles les convertissent en MBOX, puis utilisent DArcMail pour convertir les fichiers MBOX en format XML (Smithsonian Institution Archives, 2017, pp.4-5).</p>

⁵ *Ibid.*

⁶ *Ibid.*

Formats de fichier	Extensions	Brève description
Élément Microsoft Outlook	.msg	Un fichier MSG est un message unique ou un autre élément de Microsoft Outlook qui peut être téléchargé par les utilisateurs finaux d'Outlook. Contrairement à EML ou MBOX, les fichiers MSG ne peuvent pas être lus dans des éditeurs de texte ; cependant, Microsoft maintient une spécification publiée pour le format (Microsoft, [MS-OXCMSG], 2020). Il est peu probable que le format MSG soit utilisé pour transférer des comptes de messagerie entiers vers des institutions de préservation, mais les fichiers dans ce format peuvent apparaître dans les transferts de divers corps d'objets numériques à partir de lecteurs de réseau, d'ordinateurs personnels ou de systèmes d'archivage électronique. La communauté de la préservation numérique a peu discuté de MSG en tant que format de préservation, bien que la spécification ouverte et le soutien continu de Microsoft pour MSG suggèrent qu'il sera lisible à moyen terme. Lorsqu'ils capturent des comptes de messageries électroniques, les archivistes devraient éviter le format MSG en faveur de MBOX, EML ou PST.

4 Conseils pour les créateurs

Les créateurs travaillant dans des administrations, des entreprises ou d'autres environnements contrôlés doivent connaître les politiques de leur organisation en matière de traitement des courriers électroniques et de gestion des documents d'activité (*records management*). Le respect des directives et des exigences en matière de formats de fichiers, de versionnage, de conservation et de création de métadonnées contribuera à garantir la préservation du contexte et de la signification des messages au fil du temps. La gestion des courriers électroniques peut représenter un défi important pour les créateurs en dehors de ces environnements ; cependant, il existe des mesures que les créateurs peuvent prendre pour s'assurer que les messages ayant valeur d'archives historiques peuvent être traités de manière appropriée.

Ces conseils constituent une orientation utile pour les utilisateurs de courriers électroniques dans divers environnements :

- En l'absence de politiques de gestion et de conservation des messages, les créateurs doivent engager une communication directe avec les archivistes ou tout autre personne chargée de préserver les messages, afin d'établir des priorités de préservation, des méthodologies de transfert et des politiques d'élimination.
- Les particuliers qui envisagent de faire don de courriers électroniques à une institution de préservation doivent s'attendre à fournir des informations sur la façon dont ils ont géré leur messagerie au fil du temps.
- Lorsqu'ils rédigent des messages, les créateurs devraient éviter d'inclure des données permettant d'identifier personnellement des tiers, sauf si elles sont nécessaires à la conduite de leurs affaires.

- Les comptes de messageries électroniques professionnels et personnels doivent être séparés, et les messages envoyés à partir d'une adresse électronique professionnelle ne doivent pas être de nature personnelle.
- Les créateurs qui suppriment des messages, que ce soit conformément aux politiques de l'organisation ou à titre privé, doivent garder à l'esprit qu'ils peuvent avoir à supprimer des messages à la fois de leur boîte de réception et des dossiers d'éléments envoyés. Le simple fait de les déplacer dans un dossier « Corbeille » n'est pas suffisant.
- Si les créateurs ne sont pas sûrs de pouvoir éliminer ou non certains messages, ils doivent consulter les *records managers* et/ou les archivistes pour obtenir des conseils.

5 Conseils pour les archivistes

5.1 Conseils généraux

Les ressources suivantes fournissent des conseils sur la préservation et l'accès aux courriers électroniques :

- *The future of email archives* ([CLIR](#), 2018).
- *Preserving email Technology Watch Report* ([DPC et Prom](#), 2019).
- Des projets tels que EASi ([Harvard](#), 2017), Carcanet Press ([Baker et al](#), 2012), DarcMail ([Smithsonian Institution Archives](#), 2017), ePADD ([Stanford Libraries](#), 2021) et RATOM ([University of North Carolina](#), 2021) offrent des exemples concrets et des outils pour la préservation des courriers électroniques.
- Un certain nombre d'outils logiciels sont disponibles pour travailler avec les données issues des courriers électroniques ([COPTR](#), 2021).

5.2 Collecte et évaluation

- Cherchez à acquérir des comptes de messagerie entiers plutôt que des messages individuels chaque fois que cela est possible, afin de conserver les hiérarchies de dossiers et les métadonnées telles que le statut « Lu »/« Non lu », les drapeaux, les étiquettes et autres métadonnées générées par les logiciels de messagerie.
- Obtenez un accès direct aux comptes de messagerie pour les collecter et les transférer aux archives, afin d'utiliser les capacités d'export natives. Si un tel accès n'est pas possible, des outils comme [OfflineIMAP](#) peuvent extraire les comptes à distance.
- Soyez prêt à collecter des pièces jointes dans divers formats de fichiers et à élaborer des plans de préservation pour celles-ci.
- Travaillez en étroite collaboration avec les *records managers* et le personnel des services informatiques afin d'établir des stratégies et des approches techniques pour collecter et préserver les courriers électroniques.

5.3 Caractérisation

- Identifiez les formats de fichiers à l'aide d'un outil tel que DROID ([The National Archives, n.d.](#)), FIDO ([Open Preservation Foundation](#), 2020) ou Siegfried ([Lehane](#), 2020) qui utilise le registre des formats de fichiers PRONOM ([The National Archives](#), 2020).
- Utilisez des outils de validation PDF tels que JHOVE ([Open Preservation Foundation](#), 2020) et VeraPDF ([Open Preservation Foundation](#), 2020) pour vérifier la conformité aux spécifications PDF et aux normes de métadonnées publiées. Les outils de validation pour les autres formats de courrier électronique ne sont pas disponibles pour le moment.

5.4 Contrôles de qualité

- Si vous utilisez des outils de conversion tels que les convertisseurs OLM vers PST, assurez-vous de la mise en place de contrôles de qualité contre la perte de contenu, de métadonnées ou de propriétés importantes.
- Conservez le format d'origine des courriers électroniques si possible, en étant conscient que les conseils en matière de format et les outils de conversion changent au fil du temps.

6 Références

Archaeology Data Service [ADS] (2009) Databases and spreadsheets: A guide to good practice.

Disponible à l'adresse suivante :

https://web.archive.org/web/20201215095638/https://guides.archaeologydataservice.ac.uk/g2gp/DbSht_Toc

Baker, F. Butler, P. and Green, B (2012) Carcanet Press Email Preservation Project. Disponible à l'adresse suivante :

<https://web.archive.org/web/20200522083351/https://www.escholar.manchester.ac.uk/uk-ac-man-scw:165096>

COPTR (2021) Email. Disponible à l'adresse suivante :

<https://web.archive.org/web/20210706065108/https://coptr.digipres.org/index.php/Email>

Dalgleish, D. (2020) Contextures Blog. Disponible à l'adresse suivante :

<https://web.archive.org/web/20200623025727/https://contexturesblog.com/archives/2020/01/30/keep-notes-on-excel-formulas-and-macros/>

Dingley, F. W. and Matamoros, A. B. (2016) What is Digital Rights Management? Disponible à l'adresse suivante :

<https://web.archive.org/web/20200821144227/https://scholarship.law.wm.edu/cgi/viewcontent.cgi?referer=&httpsredir=1&article=1121&context=libpubs>

Dutch National Archives (2003) From Digital volatility to digital permanence:

Preservingspreadsheets. Disponible à l'adresse suivante :

<https://web.archive.org/web/20130903083642/http://en.nationaalarchief.nl/sites/default/files/docs/kennisbank/volatility-permanence-spreadsh-en.pdf>

Harvard University (2017) Submitting Content to EASi. Disponible à l'adresse suivante :

<https://web.archive.org/web/20210708085021/https://wiki.harvard.edu/confluence/display/LibraryStaffDoc/3.+Submitting+Content+to+EASi>

ISO (2020a) ISO/IEC 26300-1:2015: Information technology — Open Document Format for Office Applications (OpenDocument) v1.2 — Part 1: OpenDocument Schema. Disponible à l'adresse

suivante : <https://web.archive.org/web/20201223100915/https://www.iso.org/standard/66363.html>

ISO (2020d) ISO/IEC 32000-2:2020: Document management — Portable document format — Part 2:

PDF 2.0. Disponible à l'adresse suivante :

<https://web.archive.org/web/20201218045916/https://www.iso.org/standard/75839.html>

ISO (2018) ISO/IEC 32000-1:2008: Document management — Portable document format — Part 1:

PDF 1.7. Disponible à l'adresse suivante :

<https://web.archive.org/web/20201125031501/https://www.iso.org/standard/51502.html>

Klindt, M. (2017) PDF/A considered harmful for digital preservation. Disponible à l'adresse suivante : <https://web.archive.org/web/20200917210028/https://ipres2017.jp/wp-content/uploads/15.pdf>

Lehane, R (2020) Siegfried. Disponible à l'adresse suivante : <https://web.archive.org/web/20201028192837/https://github.com/richardlehane/siegfried>

Let's solve the format problem (2020) Quattro Pro. Disponible à l'adresse suivante : https://web.archive.org/web/20200714113418/http://fileformats.archiveteam.org/wiki/Quattro_Pro

Let's solve the format problem (2019a) Documents. Disponible à l'adresse suivante : <https://web.archive.org/web/20200701000610/http://fileformats.archiveteam.org/wiki/Document>

Let's solve the format problem (2019b) Lotus 1-2-3. Disponible à l'adresse suivante : https://web.archive.org/web/20200702010518/http://fileformats.archiveteam.org/wiki/Lotus_1-2-3

Library of Congress (2021) TSV, Tab-Separated Values. Disponible à l'adresse suivante : <https://web.archive.org/web/20210211153542/https://www.loc.gov/preservation/digital/formats/fdd/fdd000533.shtml>

Library of Congress (2020-2021) Recommended formats statement: Datasets. Disponible à l'adresse suivante : <https://web.archive.org/web/20201116041423/http://www.loc.gov/preservation/resources/rfs/data.html>

Library of Congress (2020) OpenDocument Spreadsheet Document Format (ODS), Version 1.2, ISO 26300:2015. Disponible à l'adresse suivante : <https://web.archive.org/web/20201031173456/https://www.loc.gov/preservation/digital/formats/fdd/fdd000439.shtml>

Library of Congress (2017a) Sustainability of digital formats: Planning for Library of Congress collections. Disponible à l'adresse suivante : <https://web.archive.org/web/20201113043001/https://www.loc.gov/preservation/digital/formats/intro/intro.shtml>

Library of Congress (2017b) XLSX Transitional (Office Open XML), ISO 29500:2008-2016, ECMA-376, Editions 1-5. Disponible à l'adresse suivante : <https://web.archive.org/web/20201120235444/https://www.loc.gov/preservation/digital/formats/fdd/fdd000398.shtml>

Microsoft (2021) File formats that are supported in Excel. Disponible à l'adresse suivante : <https://web.archive.org/web/20201111210012/https://support.microsoft.com/en-us/office/file-formats-that-are-supported-in-excel-0943ff2c-6014-4e8d-aaea-b83d51d46247>

Mitcham, J. (2017) How can we preserve Google Documents? Disponible à l'adresse suivante : https://web.archive.org/web/20201029193019/http://digital-archiving.blogspot.com/2017/04/how-can-we-preserve-google-documents_35.html

Morrissey, S. (2020) Preserving Software: Motivations, Challenges and Approaches. Disponible à l'adresse suivante : <http://doi.org/10.7207/twgn20-02>

North Carolina Department of Cultural Resources [NCDCCR] (2012) File Format Guidelines for Management and Long-Term Retention of Electronic records. Disponible à l'adresse suivante : https://web.archive.org/web/20201031113655/http://digitalpreservation.ncdcr.gov/file_formats_in-house_preservation.pdf

Open Preservation Foundation (2020) Format Identification for Digital Objects (FIDO). Disponible à l'adresse suivante : <https://web.archive.org/web/20200916134739/https://github.com/openpreserve/fido>

Open Preservation Foundation (2020) JHOVE. Disponible à l'adresse suivante : <https://web.archive.org/web/20201031215050/https://openpreservation.org/products/jhove/>

Open Preservation Foundation (2020) veraPDF. Disponible à l'adresse suivante : <https://web.archive.org/web/20201031220541/https://openpreservation.org/products/verapdf/>

Smithsonian Institution Archives (2017) Email Preservation – DarcMail. Disponible à l'adresse suivante : <https://web.archive.org/web/20210417235150/https://siarchives.si.edu/what-we-do/digital-curation/email-preservation-darcmail>

Stanford Libraries (2021) ePADD. Disponible à l'adresse suivante : <https://web.archive.org/web/20210413140425/https://library.stanford.edu/projects/epadd>

Technopedia (2020) Macro. Disponible à l'adresse suivante : <https://web.archive.org/web/20210123172223/https://www.techopedia.com/definition/3833/macro>

The National Archives (n.d.) Digital Object Record Identification (DROID). Disponible à l'adresse suivante : <https://web.archive.org/web/20201015033155/https://github.com/digital-preservation/droid>

The National Archives (2020) The Technical Registry: PRONOM. Disponible à l'adresse suivante : <https://web.archive.org/web/20201111032324/http://www.nationalarchives.gov.uk/PRONOM/Default.aspx>

The National Archives (2013) Best practice guide to appraising and selecting records for The National Archives. Disponible à l'adresse suivante : <https://web.archive.org/web/20200310033845/https://www.nationalarchives.gov.uk/documents/information-management/best-practice-guide-appraising-and-selecting.pdf>

University of North Carolina (2021) RATOM. Disponible à l'adresse suivante : <https://web.archive.org/web/20210301095930/https://ratom.web.unc.edu/>

van Veenendaal, R., Kjærskov, F. H., Sein, K., O'Sullivan, J., Nielsen, A. B., Tømmerholt P. M., and Tømmerholt, J. (2019) Significant properties of Spreadsheets. Disponible à l'adresse suivante : https://web.archive.org/web/20191202082644/https://ipres2019.org/static/pdf/iPres2019_paper_48.pdf

W3C (2006) Extensible Markup Language (XML) 1.1 (Second Edition). Disponible à l'adresse suivante : <https://web.archive.org/web/20201023210151/https://www.w3.org/TR/2006/REC-xml11-20060816/>

Wijsman, L. (2020) The Significant Properties of Spreadsheets: Stakeholder Analysis. Disponible à l'adresse suivante : <https://doi.org/10.5281/zenodo.3971833>

Young, P. (2021) What's Up, (with Google) Docs? – The Challenge of Native Cloud Formats.

Disponible à l'adresse suivante :

<https://web.archive.org/web/20210304124326/https://www.dpconline.org/blog/whats-up-with-google-docs>