

# Préservation des documents

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 2022 et 2024. 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) et André Falut (Archives nationales). 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 la 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](mailto:info@dpconline.org).

# 1 Présentation des types de données

Le terme « document » a plusieurs définitions larges, et même les définitions plus restreintes englobent encore de nombreux formats de fichiers ([Buckland, 1998](#)). La présente note d'orientation envisage les documents comme étant principalement de nature textuelle. Les documents peuvent être nativement numériques ou numérisés à partir de sources analogiques, et peuvent aller des notes de réunion ou des mémos aux documents électroniques, aux documents juridiques et aux publications. Les documents peuvent être liés les uns aux autres et, de fait, tirent souvent une grande partie de leur signification des relations qu'ils entretiennent avec d'autres documents. Les documents peuvent être des versions complètes et finales, des versions provisoires, des versions approuvées<sup>1</sup> ou des copies.

## 2 Défis en matière de préservation

### 2.1 Absence de gestion des documents engageants (*records management*)

Un système de records management médiocre ou absent peut avoir un impact négatif sur la capacité à transférer avec succès les documents vers un service d'archives. Si les documents sont gérés dans un système d'archivage numérique (tel qu'un système EDRMS ou Electronic documents and records management system), les records managers doivent réfléchir à la manière d'extraire les documents et les métadonnées connexes du système pour les transférer. Idéalement, cela devrait être envisagé lors de la mise en œuvre du système.

### 2.2 Embedded content

Les documents nativement numériques peuvent intégrer d'autres objets, souvent des images fixes ou des tableurs, plus rarement des fichiers multimédia et même d'autres documents. Les objets numériques intégrés peuvent ne pas être restitués correctement s'ils sont ouverts dans un logiciel différent de l'application qui a créé le document.

### 2.3 Références entre fichiers

Les documents contiennent souvent des références à d'autres objets numériques tels que d'autres documents, des fichiers multimédia, des tableurs ou des sites Web. Ces objets non intégrés peuvent également nécessiter une préservation afin que le document conserve l'intégralité de son contexte et de sa signification.

### 2.4 Polices manquantes ou incorrectement substituées

Un document peut utiliser des polices qui ne sont pas intégrées dans le fichier lui-même. Lors de la restitution visuelle d'un fichier dont les polices sont manquantes, les applications peuvent substituer des polices similaires, ou parfois totalement différentes. Le résultat peut aller de différences esthétiques mineures à une perte totale de contenu lorsque des équations ou des codes-barres sont présents. Des recherches portant sur des documents Word anciens ont montré que ce problème touchait environ un document sur cinq ([Brown et Woods, 2009](#)).

### 2.5 Formats basés sur le cloud

Les fichiers de documents basés sur le cloud sont largement utilisés et facilitent la collaboration, le contrôle des versions et le partage des documents dans les environnements de travail à distance. Les documents basés sur le cloud peuvent être exportés dans divers formats, notamment DOCX, PDF, HTML, RTF et texte brut. Ce processus peut modifier la mise en forme du texte ou d'autres caractéristiques du document. Par exemple, la pagination peut être modifiée et les dates de dernière

---

<sup>1</sup> « non-draft versions » dans le texte original (NdT).

modification changées ou perdues (Young, 2021 ; Mitcham, 2017). Le fait que de tels documents n'existent que sur le Web et la facilité d'altération par différents utilisateurs en font des objets complexes à préserver.

## 2.6 Gestion des droits numériques

La gestion des droits numériques (DRM ou *Digital Rights Management*) est un ensemble de mesures techniques destinées à limiter l'utilisation des fichiers numériques, généralement pour protéger les droits de propriété intellectuelle tels que le droit d'auteur (Dingledy et Matamoros, 2016). Le chiffrement et la protection par mot de passe peuvent également être utilisés à d'autres fins, comme la restriction de l'accès aux informations personnelles. Des restrictions spécifiques peuvent concerner l'ouverture, la copie, l'enregistrement ou l'impression de fichiers, ce qui peut entraver leur préservation et leur réutilisation.

## 2.7 Anciens formats

De nombreux programmes de traitement de texte ont vu le jour au fil des ans, dont beaucoup produisent des documents dans des formats de fichiers propriétaires. L'accès aux documents anciens et leur lecture peuvent être compliqués, et même lorsqu'un traitement de texte moderne ou un éditeur de texte peuvent les restituer, il peut y avoir une perte de propriétés importantes. Les éléments de mise en forme des documents, comme les polices, la pagination, les en-têtes et les pieds de page, les tables des matières, les listes à puces et le texte souligné ou en italique, peuvent également être perdus ou altérés dans les anciens formats de documents, et les objets intégrés peuvent ne pas s'ouvrir.

## 3 Formats de fichiers

Il n'existe pas un unique format de fichier parfait pour la préservation et l'utilisation future des documents. 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 ne fournit pas une liste exhaustive des formats adaptés à la préservation et à l'accès. Le format le plus approprié pour préserver les caractéristiques et les fonctionnalités importantes d'un fichier peut être le format original dans lequel il a été créé. Il est recommandé d'effectuer des recherches et des analyses approfondies avant de migrer des fichiers vers un nouveau format.

Format de fichier	Extensions	Brève description
Electronic Publication (EPUB)	.epub	EPUB est une norme ouverte utilisée dans l'édition depuis 2007, compatible avec les lecteurs électroniques et les logiciels pour tablettes, ordinateurs et téléphones mobiles. C'est un format conteneur composé de fichiers texte, de fichiers image et de fichiers de métadonnées connexes. Les fichiers EPUB contiennent parfois des mesures de protection technique DRM qui peuvent constituer un obstacle à la préservation à long terme. La Bibliothèque du Congrès considère EPUB comme un format de préservation acceptable « si les fichiers de contenu ne sont pas chiffrés, si le fichier n'est pas soumis à une protection technologique qui compromet la préservation et

Format de fichier	Extensions	Brève description
		l'accès à long terme, et si tout le contenu est stocké dans le conteneur EPUB. » ( <a href="#">LC</a> , 2020).
OpenDocument Text Document Format (ODT)	.odt .ott	ODT est une spécification de format de fichier ouvert qui, comme DOCX, utilise XML comme structure sous-jacente. ODT est pris en charge nativement par l'application de traitement de texte Open Source LibreOffice. Le format fait partie de la famille des fichiers XML Open Document Format (ODF) et il est normalisé selon la norme <a href="#">ISO/IEC 26300-1:2015</a> . Comme DOCX, le fait qu'il soit basé sur XML et pris en charge par des spécifications et des normes ouvertes (bien qu'incroyablement complexes) en fait un format de préservation acceptable ( <a href="#">LC</a> , 2020-2021 ; <a href="#">National Archives and Records Administration [NARA]</a> , 2020).
Texte brut	.txt	Les fichiers de texte brut contiennent des caractères et des espaces blancs (tels que des espaces et des sauts de ligne), mais ne comportent pas de mise en forme ni d'objets intégrés. Ils sont indépendants du programme et n'ont pas besoin d'un logiciel spécialisé pour s'afficher sous forme de texte lisible par l'humain. Le texte brut est omniprésent, puisqu'il s'agit du format sous-jacent des codes logiciels, des fichiers journaux d'ordinateur, des fichiers balisés tels que HTML, XML, JSON, SGML et LaTeX, des fichiers SVG (Scalable Vector Graphics) et des formats de sauvegarde et d'export d'e-mails, MBOX et EML. Tout fichier pouvant être restitué entièrement dans un éditeur de texte est considéré comme un fichier de texte brut et convient généralement à la préservation à long terme.
Portable Document Format/Archive (PDF/A)	.pdf/a	<p>PDF/A est un ensemble de normes PDF conçues par Adobe Systems pour la préservation à long terme. Basé sur les normes PDF 1.4, PDF 1.7 ou PDF 2.0, PDF/A interdit les fonctionnalités qui posent problème pour la préservation à long terme, comme les polices liées et le chiffrement. Il restreint également l'utilisation des signatures numériques, ce qui peut en faire un format de préservation inadapté pour certains types de documents.</p> <p>Les différentes versions et niveaux de conformité (PDF/A-1 à PDF/A-4) permettent aux fichiers conformes de contenir différentes fonctionnalités. La version la plus récente, PDF/A-4, a été publiée en 2017 et révisée de manière substantielle en 2020. Le format PDF/A est largement utilisé tant pour la préservation que pour l'accès.</p>

Format de fichier	Extensions	Brève description
Portable Document Format (PDF)	.pdf	Le format PDF a été développé par Adobe Systems en 1993 en tant que format de présentation propriétaire pour les documents. En 2008, il a été publié en tant que norme ouverte <a href="#">ISO 32000-1:2008</a> . Le format PDF 2.0 a été publié en 2017 et mis à jour en 2020 en tant que norme <a href="#">ISO 32000-2:2020</a> . La <a href="#">PDF Association</a> met à disposition de nombreuses ressources sur les versions et les différents types de PDF. Le format PDF est compatible avec diverses applications de restitution et constitue un format d'accès commun.
Rich Text	.rtf	Le format RTF est un texte brut avec des caractères de mise en forme et des balises supplémentaires, pris en charge nativement par l'application Wordpad de Microsoft et par les premières versions de Microsoft Word. Le format RTF a été conçu à l'origine pour assurer l'interopérabilité des formats entre les différentes versions de Microsoft Word, et entre Microsoft Word et d'autres applications de traitement de texte et de publication assistée par ordinateur. Bien qu'il s'agisse d'un format propriétaire, la spécification RTF est publiée, et la Bibliothèque du Congrès le considère comme un format de préservation acceptable ( <a href="#">LC</a> , 2017).
Google Docs	N/A	La structure et le format d'un document Google sont opaques pour l'utilisateur qui ne peut voir qu'une version restituée des données (stockées dans le cloud), dans son navigateur Web. Les documents Google Docs peuvent être exportés dans différents formats, notamment DOCX, PDF, HTML, RTF et texte brut. Ce processus peut modifier la mise en forme et/ou entraîner la perte de certaines fonctionnalités. Les Archives nationales du Royaume-Uni travaillent actuellement sur les pratiques d'export de documents Google Docs et des métadonnées pertinentes ( <a href="#">Young</a> , 2021).
TeX	.tex	Les fichiers TeX (ou LaTeX) sont des fichiers de texte brut balisés, utilisés pour créer des formats de sortie, tels que le format PDF, pour la publication numérique. L'utilisation de TeX est courante en sciences et en mathématiques en raison de la précision du rendu des formules dans ce format ( <a href="#">The LaTeX Project</a> , n.d.). Les fichiers TeX aboutissent rarement dans un dépôt numérique, car il s'agit d'un format ancien conçu pour générer d'autres formats pour la publication et la distribution. S'ils y parviennent, les versions publiées de ces fichiers doivent également être collectées dans la mesure du possible.

Format de fichier	Extensions	Brève description
Word Document	.docx .doc	<p>Open Office XML (DOCX) est le format de fichier par défaut du logiciel Microsoft Word, ayant remplacé Microsoft Word Document (DOC) en 2007.</p> <p>Normalisé sous la référence <a href="#">ISO/IEC 29500-1:2016</a>, il s'agit d'un format conteneur qui regroupe un ensemble de fichiers XML afin de fournir une structure et une mise en forme lors de la restitution par un logiciel.</p> <p>Étant donné que le format est basé sur XML, qu'il possède une spécification ouverte et une norme internationale, et qu'il est largement utilisé, DOCX est un format de préservation et d'accès acceptable (<a href="#">LC</a>, 2020-2021). Les différentes versions du format DOC sont répertoriées comme des formats de préservation acceptables par la <a href="#">NARA</a> (2020), mais pas par la Bibliothèque du Congrès.</p>

## 4 Conseils pour les créateurs

Les créateurs qui travaillent dans des administrations, des entreprises ou d'autres environnements contrôlés doivent connaître les politiques de leur organisation en matière de traitement des documents et de *records management*. Le respect des directives et des exigences en matière de formats de fichiers, de mise en forme des documents, de gestion des versions et de création de métadonnées contribuera à garantir que les documents peuvent être préservés de manière à conserver leur contexte et leur signification au fil du temps. Les milieux universitaires fournissent souvent des directives détaillées pour différents types de documents (tels que les thèses et les publications), et il convient de les suivre aussi fidèlement que possible. Les organisations qui n'ont pas de politique en matière de formats peuvent trouver des équivalents utiles dans des ressources telles que le [Recommended Formats Statement](#) de la Bibliothèque du Congrès.

Les personnes qui ont l'intention de faire don de documents à un service d'archives doivent utiliser des formats de fichiers répandus et bien pris en charge, dans la mesure du possible, et être prêtes à décrire quand, comment et à quelles fins les documents ont été créés et diffusés.

Voici d'autres conseils à l'intention des créateurs de documents :

- Si vous intégrez d'autres objets numériques dans vos fichiers, conservez des copies externes des objets intégrés et transférez-les aux services d'archives avec les documents chaque fois que cela est possible.
- Si vous migrez vers un format de fichier de document alternatif, envisagez de conserver le fichier source original en même temps que le fichier dans l'état final ; par exemple, conservez les fichiers .tex ou .docx qui ont été imprimés au format PDF.
- Sachez que l'utilisation de polices non standard pour présenter des contenus tels que des équations ou des codes-barres pourrait poser des problèmes pour la préservation.
- Si vous créez une version PDF, utilisez l'application qui a servi à créer le document pour en générer une version PDF ; par exemple, les versions les plus récentes de Microsoft Word

intègrent la possibilité d'exporter au format PDF. Sinon, envisagez d'utiliser des outils tels qu'Adobe Acrobat, PDF Studio ([Qoppa Software](#), 2021) ou pdfaPilot ([Callas Software](#), n.d.).

- En cas d'enregistrement au format PDF/A, n'oubliez pas d'inspecter visuellement le document, et d'utiliser la fonctionnalité de vérification de la conformité PDF/A du logiciel si elle est disponible. Les problèmes sous-jacents du document pourraient entraîner la génération d'un fichier PDF/A non conforme ([Oates et al.](#), 2018).
- Si vous créez un fichier PDF/A pour une préservation à long terme, envisagez d'utiliser le niveau de conformité A pour les documents nativement numériques et le niveau de conformité B pour les documents numérisés ([Oates et al.](#), 2018). Les deux versions ont des exigences en matière de présentation visuelle des fichiers sur un écran d'ordinateur, mais le niveau de conformité A est plus rigoureux, exigeant que le contenu soit structuré et balisé de manière spécifique pour faciliter l'accessibilité.

## 5 Conseils pour les archivistes

### 5.1 Conseils généraux

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

- [Bibliothèque du Congrès](#) (2020-2021) *Recommended formats statement : ii. Textual Works – Digital*.
- [Archives nationales](#) du Royaume-Uni (2012) *Managing digital records without an electronic record management system*.
- Un certain nombre d'outils logiciels sont disponibles pour travailler avec des données de documents ([COPTR](#), 2021).

### 5.2 Collecte et évaluation

- Travaillez en étroite collaboration avec les *records managers* et/ou les créateurs de documents pour vous aider à jeter les bases d'un transfert réussi des documents aux archives.
- Si les documents sont extraits de systèmes d'archivage (*recordkeeping systems*), assurez-vous que les métadonnées exportées en parallèle sont suffisantes pour décrire le contexte de leur création et de leur utilisation pendant les étapes actives de leur cycle de vie.
- En cas de collecte de documents auprès de particuliers, plusieurs entretiens avec le ou les créateurs peuvent être nécessaires pour vous assurer que les documents peuvent être organisés et décrits de manière significative.
- Les archivistes doivent être informés de la présence de tout DRM ou autre mécanisme de protection, et négocier avec le donateur pour assouplir les restrictions, obtenir des clés de chiffrement ou envisager de rejeter les documents si les restrictions sont susceptibles d'entraver leur capacité à les préserver (voir également **Caractérisation**, ci-dessous).

### 5.3 Actions de préservation

- Conservez les fichiers originaux. La migration entraîne souvent une perte de mise en forme et de contenu, mais peut fournir une certaine assurance contre l'obsolescence du format. La préservation des fichiers originaux facilitera une approche d'émulation, si cela s'avère nécessaire à l'avenir.
- Soyez prudent lors de la migration vers le format PDF/A. La génération de documents PDF/A fiables et conformes à partir de fichiers de traitement de texte (y compris les fichiers PDF) peut poser des problèmes ([Klindt](#), 2017 ; [Oates et al.](#), 2018).
- Pour les documents basés sur le cloud, un service d'export dédié tel que Google Takeout Services peut être utilisé, mais une assurance qualité doit être mise en œuvre pour garantir que les propriétés significatives pertinentes ne sont pas affectées négativement ([Young](#),

2021 ; [Mitcham](#), 2017). L'archivage Web est une autre option pour capturer les documents basés sur le cloud dans leur format natif ([Young](#), 2021).

#### 5.4 Documents anciens

- Lors de la collecte de documents de traitement de texte anciens, soyez attentif aux éventuelles versions imprimées de ces documents qui seraient déjà présentes dans les archives. Si des copies imprimées existent, il peut être plus pratique de les préserver que de préserver les versions numériques, à moins que les versions numériques ne comportent des métadonnées et d'autres attributs qui ajoutent un contexte nécessaire aux documents.
- Adoptez une stratégie d'émulation si les documents ne peuvent pas être rendus de manière fiable dans les logiciels modernes. Il est désormais possible d'émuler certains anciens programmes de traitement de texte à l'aide de plateformes d'émulation telles que WordStar Emulator ([Wordstar](#), 2020), PCjs Machines ([PCjs](#), 2021) et vDosWP ([Columbia University](#), 2020). Des projets tels que *Scaling Emulation as a Service Infrastructure* (EaaS) ([Educopia Institute](#), 2020), dirigé par l'université de Yale, et *Emulation as a Service*, dirigé par [l'université de Fribourg](#) (2020), visent à fournir un accès simplifié et évolutif à l'infrastructure d'émulation.

#### 5.5 Caractérisation

La caractérisation peut être utile pour identifier les formats de fichiers, extraire les métadonnées, identifier les contenus corrompus ou chiffrés, ou vérifier la conformité à des profils ou à des normes. La prise en charge et l'efficacité des outils peuvent varier considérablement suivant les différents formats de fichiers.

- 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 utilisent 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. Utilisez le validateur ODF de LibreOffice ([LibreOffice](#), n.d.) pour valider la conformité aux normes du format ODF (OpenDocument Format).
- Des outils de caractérisation peuvent servir à identifier les fichiers dotés de DRM, de chiffrement ou d'autres méthodes de protection.

#### 5.6 Métadonnées

- Documents can contain descriptive and technical metadata generated automatically during creation and modification. Some document creation tools support the ability to add additional descriptive metadata manually.
- Embedded metadata can be viewed and modified using applications such as the open-source ExifTool ([Harvey](#), 2021) and commercial tools such as [GroupDocs Metadata](#) (2020) and [MetaWiper](#) (2021). The ability to view embedded metadata may be diminished over time as the file formats approach obsolescence.

## 6 Références

Brown and Woods (2009) Born Broken: Fonts and Information Loss in Legacy Digital Documents.

Disponible à l'adresse suivante :

<https://web.archive.org/web/20201030203247/https://escholarship.org/uc/item/53z897zb>

Buckland, M (1998) What is a "digital document"? Disponible à l'adresse suivante :  
<https://web.archive.org/web/20201124085206/https://people.ischool.berkeley.edu/~buckland/digdoc.html>

Callas Software (n.d.) pdfaPilot. Disponible à l'adresse suivante :  
<https://web.archive.org/web/20210107171004/https://www.callassoftware.com/en/products/pdfapilot>

Columbia University (2020) vDosWP System for WPDOS under 64-bit Windows. Disponible à l'adresse suivante :  
<https://web.archive.org/web/20201130042537/http://www.columbia.edu/~em36/wpdos/vdoswp.html>

COPTR (2021) Documents. Disponible à l'adresse suivante :  
<https://web.archive.org/web/20210707081618/https://coptr.digipres.org/index.php/Document>

Digital Preservation Coalition (2020) Handbook Glossary: Electronic Records. Disponible à l'adresse suivante :  
<https://web.archive.org/web/20201028205756/https://www.dpconline.org/handbook/glossary#E>

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>

Educopia Institute (2020) Scaling Emulation as a Service Infrastructure (EaaS) (subcontract). Disponible à l'adresse suivante :  
<https://web.archive.org/web/20201129194437/https://educopia.org/emulation-as-a-service-eaasi/>

Google Developers (2019) Structure of a Google Docs document. Disponible à l'adresse suivante :  
<https://web.archive.org/web/20201101023029/https://developers.google.com/docs/api/concepts/structure>

GroupDocs (2020) GroupDocs.Metadata. Disponible à l'adresse suivante :  
<https://web.archive.org/web/20200809000226/https://products.groupdocs.app/metadata/docx>

Harvey, P. (2021) Exiftool. Disponible à l'adresse suivante :  
<https://web.archive.org/web/20210122225751/https://exiftool.org/>

ISO (2020) 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/20201112002833/https://www.iso.org/standard/71691.html>

ISO (2016) ISO/IEC 29500-1:2016: Information technology — Document description and processing languages — Office Open XML File Formats — Part 1: Fundamentals and Markup Language Reference. Disponible à l'adresse suivante :  
<https://web.archive.org/web/20201112002833/https://www.iso.org/standard/71691.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>

ISO (2020) 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>

InterPARES Project (1999-2018) Glossary. Disponible à l'adresse suivante :

[http://interpares.org/display\\_file.cfm?doc=ip1\\_glossary.pdf](http://interpares.org/display_file.cfm?doc=ip1_glossary.pdf) [accessed 25 January 2021]

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>

Library of Congress (2020-2021) Recommended formats statement: ii. Textual Works – Digital.

Disponible à l'adresse suivante :

<https://web.archive.org/web/20201107175543/https://www.loc.gov/preservation/resources/rfs/text.html>

Library of Congress (2020) Sustainability of Digital Formats: Planning for Library of Congress

Zollections: E PUB (Electronic Publication) File Format Family. Disponible à l'adresse suivante :

<https://web.archive.org/web/20201023234157/https://www.loc.gov/preservation/digital/formats/fdd/fdd000310.shtml>

Library of Congress (2017) Sustainability of Digital Formats: Planning for Library of Congress

Collections: Rich Text Format (RTF) Family. Disponible à l'adresse suivante :

<https://web.archive.org/web/20200917211110/https://www.loc.gov/preservation/digital/formats/fdd/fdd000473.shtml>

LibreOffice (n.d.) ODF Validator. Disponible à l'adresse suivante :

<https://web.archive.org/web/20210708081233/https://odftoolkit.org/conformance/ODFValidator.html>

MetaWiper (2021) MetaWiper. Disponible à l'adresse suivante : <https://www.metawiper.com/word-metadata-changer> [accessed 25 January 2021]

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](https://web.archive.org/web/20201029193019/http://digital-archiving.blogspot.com/2017/04/how-can-we-preserve-google-documents_35.html)

Oates, A. I., Downie, S. J., Halvarsson, E., and Popham, M. (2018) Navigating the PDF/A Standard: A Case Study of Theses in Oxford's Institutional Repository. Disponible à l'adresse suivante :

<https://www.ideals.illinois.edu/handle/2142/100236>

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/>

OpenPreservation Foundation (2020) veraPDF. Disponible à l'adresse suivante :

<https://web.archive.org/web/20201031220541/https://openpreservation.org/products/verapdf/>

PCjs (2021) PCjs machines. Disponible à l'adresse suivante :

<https://web.archive.org/web/20210108064056/https://www.pcjs.org/software/pcx86/app/microsoft/word/5.00/>

Qoppa Software (2021) PDF Studio. Disponible à l'adresse suivante :

<https://web.archive.org/web/20210102222204/https://www.qoppa.com/pdfstudio/>.

Society of American Archivists (2005-2020) Dictionary: document. Disponible à l'adresse suivante :

<https://web.archive.org/web/20210107121155/https://dictionary.archivists.org/entry/document.html> (Wayback Machine capture from 7 January 2021)

Tech Terms Computer Dictionary (2020) Rich Text. Disponible à l'adresse suivante :

<https://web.archive.org/web/20201022230151/https://techterms.com/definition/richtext>

The LaTeX Project (n.d.) LaTeX – A document preparation system. Disponible à l'adresse suivante :

<https://web.archive.org/web/20210105071245/https://www.latex-project.org/>

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 (2021) The Technical Registry: PRONOM. Disponible à l'adresse suivante :

<https://web.archive.org/web/20210108231304/https://www.nationalarchives.gov.uk/PRONOM/Default.aspx>

The National Archives (2012) Managing digital records without an electronic record management system. Disponible à l'adresse suivante :

<https://web.archive.org/web/20200809232536/https://www.nationalarchives.gov.uk/documents/information-management/managing-electronic-records-without-an-erms-publication-edition.pdf>

University of Freiburg (2020) bwFLA — Emulation as a Service. Disponible à l'adresse suivante :

<https://web.archive.org/web/20201103031720/http://eaas.uni-freiburg.de/>

Wordstar (2020) WordStar Emulator. Disponible à l'adresse suivante :

<https://web.archive.org/web/20200601225007/http://www.wordstar.org/index.php/wordstar-emulator>

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>