

Préservation des feuilles de calcul

Collection de notes d'orientation sur les types de données

Artefactual Systems et Digital Preservation Coalition



**Veille technologique de
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 janvier et mars 2022. 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.

Le manuel a été traduit en français par : Wilfried Prieur (ministère de la Culture) et 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 pour un entrepôt 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 la *UK Nuclear Decommissioning Authority*.

La préservation numérique est un domaine en pleine évolution et 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

Les feuilles de calcul stockent des données sous forme de tableaux divisés en colonnes et en lignes de cellules. Ces cellules de feuille de calcul peuvent posséder des attributs techniques qui influencent l’affichage des données qu’elles contiennent. Il existe environ vingt et un groupes d’attributs techniques identifiés (Wijsman, 2020), notamment :

- une catégorie de format, telle que nombre, devise, comptabilité, date, heure, pourcentage ou fraction ;
- des formules ;
- des commentaires ;
- des tableaux croisés dynamiques ;
- le style, notamment l’alignement du texte, la couleur et le style de la police, la couleur des cellules et les styles de bordure. Le style est souvent utilisé pour mettre en évidence des aspects spécifiques des données, tels que les totaux ou les nombres négatifs, mais aussi pour transmettre du sens (Archaeology Data Service [ADS], 2009).

Les feuilles de calcul existaient avant l’informatique moderne, mais les feuilles de calcul automatisées en tant qu’application logicielle ont été inventées vers 1961 (Nationaal Archief¹, 2003). Au fil des ans, les feuilles de calcul ont évolué pour devenir plus qu’une simple feuille de données délimitées par des tabulations, des virgules ou d’autres séparateurs, et des valeurs basées sur des cellules. Les feuilles de calcul peuvent contenir des informations dynamiques, effectuer des équations mathématiques, créer des visualisations et stocker des données intégrées. Les feuilles de calcul peuvent comporter plusieurs feuilles ou colonnes invisibles. De ce fait, elles peuvent être plus complexes qu’il n’y paraît à première vue, ce qui peut entraîner une perte de données ou une impossibilité d’accéder au contenu si elles ne sont pas traitées avec les niveaux de précaution appropriés.

2 Défis en matière de préservation

La complexité parfois cachée des feuilles de calcul donne lieu à quelques défis notables en matière de préservation. Il s’agit notamment de la manipulation des formules et des macros, de la gestion des droits numériques, de la collecte des anciens² formats de fichiers et des systèmes logiciels basés sur le Cloud qui ne produisent pas de fichiers tels que nous les interprétons habituellement.

2.1 Contenu dynamique

La préservation des formules utilisées pour créer des feuilles de calcul dynamiques présente un défi. Le contenu dynamique peut inclure des styles, des formules, des graphiques, des diagrammes ou des macros. Il a été signalé que 99 % des feuilles de calcul ingérées dans la collection Ejournal des Nationaal Archief étaient considérées comme des feuilles de calcul dynamiques (van Veenendaal, 2019). Les défis spécifiques à la préservation incluent notamment les suivants :

¹ NdT, Archives Nationales des Pays-Bas

² NdT : « ancien » a été choisi pour traduire l’anglais « legacy » qui comprend également la notion d’héritage et d’arriéré.

- Formules : contrairement aux valeurs littérales indiquées dans les cellules de la feuille de calcul, les formules qui effectuent des opérations ou des calculs peuvent ne pas être compatibles entre différents logiciels de feuilles de calcul.
- Macros : il s'agit d'extraits de code qui exécutent des tâches répétitives sur une feuille de calcul. Ils déclenchent automatiquement des séquences de frappes de touches et de mouvements de souris (Technopedia, 2020). Les macros sont généralement écrites en Visual Basic ou en Javascript et peuvent ne pas être compatibles entre différents logiciels de feuilles de calcul.

2.2 Contenu embarqué

Les feuilles de calcul peuvent contenir d'autres types de données, comme des images ou d'autres feuilles de calcul.

Les données embarquées dans les feuilles de calcul rendent la préservation plus difficile.

- Le contenu embarqué peut être lié à un autre emplacement sur l'ordinateur, un réseau d'accès local ou au Cloud. Cette connexion peut être rompue lorsque la feuille de calcul ou le contenu lié est déplacé.
- Le contenu embarqué peut être perdu si la feuille de calcul est migrée vers un format différent.
- Si le logiciel utilisé pour la création de la feuille de calcul n'est pas disponible à l'avenir, le rendu du contenu intégré dans une nouvelle application logicielle peut ne pas représenter fidèlement la structure d'origine. Il peut avoir une apparence différente, se trouver à un autre emplacement dans la feuille de calcul, ou être entièrement perdu.

2.3 Formats basés sur le Cloud

Les feuilles de calcul disponibles sur Internet doivent être exportées vers un format d'état final tel que Microsoft Excel ou Open Document Format pour une préservation à long terme. Les recherches actuelles sur l'exportation des feuilles de calcul ont donné quelques résultats, mais la capture des formules spécifiques à Google et des Sparklines (graphiques linéaires intégrés dans les cellules) pose problème (Young, 2021). Les exportations PDF et HTML sont également moins complètes que Microsoft Excel ou Open Document Format. L'API de Google permet également de recueillir des métadonnées supplémentaires qui ne sont pas disponibles par l'exportation directe à partir de Google Sheets (Young, 2021).

2.4 Gestion des droits numériques

La gestion des droits numériques (GDN)³ désigne 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 cryptage 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

³ NdT, aussi appelé DRM (*Digital Rights Management*).

concerner l'ouverture, la copie, l'enregistrement ou l'impression de fichiers, ce qui peut entraver leur préservation et leur réutilisation.

2.5 Anciens formats

Au fil des ans ont existé de nombreux logiciels permettant de créer des feuilles de calcul. L'accès à ces feuilles de calcul et leur lecture peuvent être compliqués. Même si une application moderne peut les restituer, des propriétés importantes peuvent être perdues. Les formules, les macros, les graphiques, les diagrammes et le style peuvent être perdus ou altérés dans les anciens formats de feuilles de calcul, et le contenu embarqué ou lié peut ne pas être accessible.

3 Formats typiques de feuilles de calcul

Il n'existe pas de format unique parfait pour la préservation et l'utilisation future des feuilles de calcul. Les décisions prises concernant les formats de fichiers doivent dépendre des caractéristiques et des fonctionnalités à préserver, ainsi que 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 | Description sommaire |
|---|---------------|--|
| <i>Valeurs séparées par un délimiteur</i> | .csv, .txt | <p><i>Comma Separated Value</i> (CSV) est un format simple et ouvert couramment utilisé pour stocker et exporter des données de feuilles de calcul. Il utilise des virgules pour différencier les colonnes et les lignes. Le format TSV (<i>Tab Separated Values</i>) utilise des caractères de tabulation à la place des virgules (LC, 2021).</p> <p>Des fichiers séparés par des tabulations ou des virgules peuvent être créés à partir de fichiers de texte brut pour stocker des données structurées. Les tabulations ou les virgules donnent la structure de base des lignes et des colonnes dans le document texte. Il convient toutefois de noter que ces formats de fichier ne peuvent pas préserver le style, les formules, les graphiques, les diagrammes, les relations entre plusieurs feuilles ou d'autres fonctionnalités dynamiques des feuilles de calcul (NCDRCR, 2012). Si aucune de ces caractéristiques n'est présente, ils constituent les formats de préservation et d'accès privilégiés (LC, 2020-2021 ; ADS, 2009).</p> |

| Format de fichier | Extensions | Description sommaire |
|---------------------------------|----------------|---|
| <i>Microsoft Excel</i> | .xlsx | <i>Open Office XML (XLSX)</i> est le format de fichier par défaut du logiciel Microsoft Excel, ayant remplacé le format de fichier binaire Microsoft Excel (XLS) en 2007. Normalisé sous la référence ISO/IEC 29500-1:2016 , il s'agit d'un format conteneur qui regroupe un ensemble de fichiers XML pour fournir une structure et un formatage lors du rendu par le logiciel. Étant donné que le format est basé sur XML, qu'il possède une spécification ouverte (bien qu'incroyablement complexe), une norme internationale et qu'il est largement utilisé, XLSX est un format de préservation privilégié (LC, 2017b ; ADS, 2009). |
| <i>Open Document Format</i> | .ods, .fods | <i>Open Document Format</i> , normalisé sous la référence ISO/IEC 26300-1:2015 , est un format XML ouvert pour les documents bureautiques, notamment les feuilles de calcul. La dernière version, ODS version 1.2, présente moins de problèmes d'interopérabilité que les versions précédentes (LC, 2020). Il est considéré comme un format de préservation privilégié (LC, 2020-2021 ; ADS, 2009). |
| <i>Google Sheets</i> | N/A | La structure et le format d'une feuille de calcul Google Sheets sont opaques pour l'utilisateur qui ne peut voir dans son navigateur Web qu'une version restituée des données (stockées dans le cloud). Les feuilles de calcul Google peuvent être exportées dans différents formats, notamment XSLX et ODS. Ce processus peut modifier le formatage et/ou entraîner la perte de certaines fonctionnalités. Les Archives nationales britanniques travaillent actuellement sur les pratiques d'exportation de feuilles de calcul Google et des métadonnées correspondantes (Young, 2021). |
| <i>Portable Document Format</i> | .pdf .pdf/A | 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 ISO 32000-1:2008 . PDF 2.0 a été publié en 2017 et mis à jour en 2020 en tant que norme ISO 32000-2:2020 . Le PDF/A est le format de préservation conçu par Adobe Systems. Pour des informations détaillées sur le PDF/A, veuillez consulter la note d'orientation sur les documents . Les exportations PDF peuvent conserver « l'aspect et la convivialité » d'une feuille de calcul, mais pas ses fonctionnalités (NCDRCR, 2012). L'exportation de feuilles de calcul au format PDF supprime bon nombre de ses propriétés dynamiques essentielles. Un PDF peut représenter différentes couches et formules associées à une feuille de calcul, mais il |

| Format de fichier | Extensions | Description sommaire |
|-------------------|------------|--|
| | | ne peut pas maintenir les relations avec d'autres feuilles de calcul ou des sources de données externes (Archives nationales néerlandaises, 2003). L'exportation au format PDF n'est pas recommandée comme solution à grande échelle ; cependant, dans certaines circonstances, il peut s'agir d'un format de préservation acceptable, en fonction des propriétés importantes à préserver. |

4 Conseils pour les producteurs

Les producteurs des administrations, des entreprises ou d'autres environnements contrôlés doivent connaître les politiques de leur organisation en matière de traitement des feuilles de calcul et de gestion des documents. Le respect des directives et des exigences en matière de formats de fichiers, de formatage et de style, de gestion des versions et de création de métadonnées contribuera à garantir que les feuilles de calcul pourront être préservées de manière à conserver leur contexte et leur signification au fil du temps. Les milieux universitaires fournissent souvent des directives détaillées pour les données de recherche stockées dans des feuilles de calcul qu'il convient de suivre aussi fidèlement que possible. Les organisations qui n'ont pas de politique sur les formats peuvent trouver des ressources telles que le document Recommended Formats Statement de la Library of Congress comme substituts utiles.

Les bonnes pratiques techniques invitent à envisager la préservation en amont de la création des fichiers. Les suggestions de création et de gestion des feuilles de calcul ci-dessous contribueront à faciliter une préservation et une accessibilité à long terme :

- Notez que ce n'est pas parce qu'une feuille de calcul peut être ouverte que l'on peut accéder à toutes ses propriétés.
- Les plates-formes basées sur le Cloud (par exemple, Google Sheets) créent des feuilles de calcul qui ne sont pas des fichiers (Mitcham, 2017). Étudiez les options de sortie ou d'exportation de ce contenu et vérifiez que les fonctionnalités de la feuille de calcul sont toujours présentes dans les fichiers exportés.
- Les feuilles de calcul doivent être « auto-descriptives » et capables d'exister indépendamment. Créez des titres de lignes et de colonnes significatifs, et décrivez les unités utilisées dans la feuille de calcul. Utilisez des vocabulaires contrôlés et des listes de mots établies, dans la mesure du possible, pour la saisie des données, afin de garantir la cohérence et la clarté des données (ADS, 2009).
- La Library of Congress (2020-2021) recommande de copier les macros ou de les supprimer entièrement pour la préservation à long terme. Dalgleish (2020) fournit des conseils pour stocker les formules et les macros en externe. Ces conseils portent notamment sur la manière d'ajouter des commentaires contextuels au sein du code de la macro. Le code de la macro peut être stocké en externe sous forme de fichiers de texte brut.
- La feuille de calcul doit être « autonome », ce qui signifie que le contenu lié à l'extérieur est capturé et transféré avec la feuille de calcul. Ainsi, la relation et le contexte entre le contenu lié et la feuille de calcul seront préservés (LC, 2020-2021).
- Les auteurs de feuilles de calcul devraient éviter d'utiliser des GDN lorsque cela est possible.

- Les données textuelles des feuilles de calcul ont une plus grande longévité si l'encodage du texte est défini en ASCII ou en UTF-8 (LC, 2020-2021). D'autres encodages de texte peuvent créer des problèmes avec les diacritiques et d'autres caractères spéciaux lors du rendu des feuilles de calcul dans différents logiciels et environnements de systèmes d'exploitation.

5 Conseils pour les archivistes

5.1 Conseils généraux

Les ressources suivantes fournissent des conseils sur la préservation et l'accès aux feuilles de calcul :

- Les documents *Guides to Good Practice on spreadsheets* de l'Archaeology Data Service (2009) comportent des conseils généraux sur la préservation et le format des fichiers.
- Le document *Sustainability of Digital Formats: Planning for Library of Congress Collections* de la Library of Congress (2017a) donne des aperçus détaillés et des informations sur divers formats de fichiers de tableur.
- Le wiki *Let's Solve the Format Problem* (2019) possède une liste de formats de fichiers de feuille de calcul qui décrit les logiciels accessibles pour les anciens formats, tels que Lotus 1-2-3 et Quattro Pro.
- Le rapport des Nationaal Archief (2003) *From digital volatility to digital permanence: Preserving spreadsheets* dispose de conseils sur la préservation de l'authenticité et du contexte des feuilles de calcul. Bien qu'il s'agisse d'une ressource plus ancienne, il s'agit toujours des conseils les plus complets sur la manière de préserver l'authenticité des feuilles de calcul.
- Un certain nombre d'outils logiciels sont disponibles pour travailler avec des données de feuilles de calcul (COPTR, 2021).

5.2 Collecte et évaluation

La préservation des feuilles de calcul dépend largement du mandat de collecte d'un organisme et de l'acquisition d'autres documents liés aux feuilles de calcul.

- Souvent, les feuilles de calcul sont créées pour effectuer des calculs qui sont ensuite intégrés à d'autres documents, ou pour organiser des informations de manière temporaire. Par conséquent, l'évaluation peut montrer que leur valeur en tant que document est limitée et leur conservation inutile.
- Certains domaines (comme les sciences ou la finance) peuvent accorder une plus grande valeur archivistique au contenu d'une feuille de calcul et exiger une préservation et un accès à long terme.
- Envisagez de demander la suppression des GDN dans les feuilles de calcul avant leur acquisition. Les feuilles de calcul dotées de toute caractéristique GDN interdisant l'accès ou l'utilisation ne doivent pas être collectées si possible (LC, 2020-2021).

5.3 Authenticité et contexte

- Les feuilles de calcul doivent être accompagnées de métadonnées descriptives pertinentes afin de fournir un contexte. Assurez-vous que les feuilles de calcul et les documents associés

possèdent au moins les attributs suivants : nom de l'organisation, objet, date et relation avec d'autres fichiers.

- Rédigez une documentation contenant des informations sur les formats de fichiers originaux et actuels, tout logiciel associé utilisé pour créer ou restituer les feuilles de calcul, et toute action de préservation qui a été prise. Les exigences minimales pour assurer l'authenticité des documents sont les suivantes : les feuilles de calcul ont une séquence correcte avec des noms, des lignes et des colonnes correctes ; elles conservent leur structure, les objets intégrés et les formules ; et elles ont des relations correctes entre les cellules (Nationaal Archief, 2003).
- Lorsque cela est possible, conservez les caractéristiques esthétiques, telles que la graisse et les couleurs des polices (en particulier lorsqu'elles sont porteuses de sens) mais sachez que la reproduction exacte n'est pas toujours essentielle.

5.4 Rétrocompatibilité

- Tenez compte du potentiel (ou de l'absence) de rétrocompatibilité. Par exemple, les versions modernes du logiciel Microsoft Excel, comme Office 365, peuvent être utilisées pour lire des fichiers créés dans l'édition antérieure Excel 95 du logiciel et enregistrés au format Excel 95 (Microsoft, 2021).
- La documentation des anciens logiciels peut également fournir des indications sur les logiciels appropriés pour la rétrocompatibilité. Le wiki Let's Solve the Format Problem (2019a) traite de la rétrocompatibilité pour de nombreux anciens formats de feuilles de calcul. Microsoft (2021) dispose de vérificateurs de compatibilité dans les logiciels Microsoft Excel modernes, et fournit une liste des formats compatibles.
- S'appuyer sur la seule rétrocompatibilité peut ne pas suffire à assurer la continuité de l'accès. Les applications commerciales peuvent supprimer la prise en charge des formats obsolètes, ou peuvent elles-mêmes devenir obsolètes.

5.5 Actions de préservation

- Conservez les fichiers originaux pour maintenir des alternatives dans le temps. Le contrôle de la qualité des actions de préservation est généralement coûteux, difficile et sans support technologique pour l'automatisation.
- Une approche par l'émulation (Morrissey, 2020) peut être appliquée pour préserver et donner accès aux feuilles de calcul et au logiciel avec lequel elles ont été créées.
- La migration d'un format à un autre peut constituer une approche viable dans certains cas, mais pourrait entraîner la perte de certaines données et/ou de fonctionnalités, en particulier pour le contenu dynamique.
- La migration vers un format de moindre fidélité (tel que CSV, Texte ou PDF/A) peut simplifier les défis de préservation des fichiers résultants si vous êtes pleinement conscient de la perte de certaines fonctionnalités et/ou données. Cette stratégie peut être utile en parallèle avec les approches plus complexes d'émulation ou de migration complète.
- Faites attention aux macros et à toute autre fonctionnalité de script. Créez des flux de travail d'assurance qualité qui permettent de garantir que les données ne sont pas perdues lors de tout traitement de préservation. Il est recommandé de stocker les macros séparément des feuilles de calcul (Dalgleish, 2020 ; ADS, 2009).

5.6 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 (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 feuilles de calcul ne sont pas disponibles actuellement.

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

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

<https://web.archive.org/web/20210705083538/https://coptr.digipres.org/index.php/Spreadsheet>

Dagleish, 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: Preserving spreadsheets*. 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>

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, ISO26300:2015*. 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/>

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>

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>