

## Feuilles de calcul : analyse de la situation et perspectives pour le Catalogue des formats de données d'archivage (Cfa)

Lambert Kansy, Archives de l'État de Bâle-Ville,

en collaboration avec le groupe d'experts Preservation Planning du CECO

Septembre 2021

|     |                                      |   |
|-----|--------------------------------------|---|
| 1   | Introduction.....                    | 1 |
| 1.1 | Remarque préliminaire.....           | 1 |
| 1.2 | Contexte et problématique.....       | 1 |
| 2   | Délimitation.....                    | 2 |
| 3   | Bonnes pratiques.....                | 2 |
| 4   | OOXML et ODF.....                    | 4 |
| 4.1 | OOXML.....                           | 4 |
| 4.2 | ODF.....                             | 4 |
| 4.3 | Interopérabilité d'OOXML et ODF..... | 5 |
| 5   | Conclusion.....                      | 5 |

### 1 Introduction

#### 1.1 Remarque préliminaire

La présente étude est née dans le cadre de l'élaboration de la version 6.2 du Catalogue des formats de données d'archivage (Cfa)<sup>1</sup> du CECO. Un des points principaux de cette version a été le remaniement des recommandations concernant les formats de feuilles de calcul. L'étude rapporte l'état des connaissances de septembre 2021 ; la version 6.2 du Cfa a été publiée en décembre 2021.

#### 1.2 Contexte et problématique

Un tableur (logiciel de feuilles de calcul ou de tables) représente des données numériques et alphanumériques et permet notamment de traiter ces dernières au moyen de fonctions. Depuis la version 5.1 de 2017, le Cfa recommande de procéder de la manière suivante en ce qui concerne l'archivage de fichiers de tableurs :

« À titre de solution provisoire, il faut archiver dans le format d'origine, en général [XLS](#), [ODF](#) ou [OOXML](#). Les alternatives comme archiver dans des formats d'archivage [Base de données](#) ou convertir en [PDF/A](#) ne peuvent être considérées comme bonnes pratiques et doivent être évaluées de manière critique. »<sup>2</sup>

La question se pose de savoir si de nouvelles connaissances depuis 2017 nécessitent de modifier la recommandation. Les aspects dont il faut tenir compte restent inchangés :

- a. Fonctionnalité (aspect calcul)
- b. Nature du tableau
- c. Aspect visuel

<sup>1</sup> <https://kost-ceco.ch/cms/formats-de-fichiers.html>

<sup>2</sup> [https://kost-ceco.ch/wiki/whelp/KaD\\_cmsv5.1/tableurs-sommaire.html](https://kost-ceco.ch/wiki/whelp/KaD_cmsv5.1/tableurs-sommaire.html)

## 2 Délimitation

Les explications n'incluent aucun format en nuage comme Google Docs. L'étude ne contient pas d'explication approfondie sur l'interopérabilité entre ODF et OOXML ni sur les risques lors de la conversion des formats précédents respectifs (XLS ou SXC) vers les formats actuels. L'étude n'apporte pas non plus d'éclaircissement pour déterminer lequel des formats actuels (XLSX et ODS) se prête le mieux suivant les circonstances comme format cible pour la conversion depuis d'autres formats de fichiers obsolètes.

## 3 Bonnes pratiques

Sur la base de ces questions, les recommandations actuelles concernant le traitement des données de tableurs des institutions d'archivage suivantes ont été examinées<sup>3</sup>:

- Nationaal Archief (NL): Handreiking voorkeursformaten Nationaal Archief, **2016**  
<https://www.nationaalarchief.nl/archiveren/kennisbank/handreiking-voorkeursformaten-nationaal-archief>
- SLUB Dresden, Langzeitarchivfähige Dateiformate, Version 1.3.4, **2021-02-22**  
[https://slubarchiv.slub-dresden.de/fileadmin/groups/slubsite/slubarchiv/SLUBArchiv\\_langzeitarchivfaehige\\_Dateiformate\\_v1.3.4.pdf](https://slubarchiv.slub-dresden.de/fileadmin/groups/slubsite/slubarchiv/SLUBArchiv_langzeitarchivfaehige_Dateiformate_v1.3.4.pdf)
- Bibliothèque de l'EPFZ, Archivtaugliche Dateiformate, **s. d.**  
<https://documentation.library.ethz.ch/display/DD/Archivtaugliche+Dateiformate>
- Bibliothèque et Archives Canada, Lignes directrices sur les formats de fichier à utiliser pour transférer des ressources documentaires à valeur continue, **2014-10-01**  
<https://www.bac-lac.gc.ca/fra/services/gestion-ressources-documentaires-gouvernement/lignes-directrices/Documents/formats-fichier-rdvc.pdf>
- Archaeology Data Service, GUIDELINES FOR DEPOSITORS : Downloads, Version 4.1, **April 2021**  
<https://archaeologydataservice.ac.uk/advice/Downloads.xhtml>
- Archaeology Data Service / Digital Antiquity, Guides to Good Practice: Databases and Spreadsheets: A Guide to Good Practice, **2013**  
[https://guides.archaeologydataservice.ac.uk/g2gp/DbSht\\_1](https://guides.archaeologydataservice.ac.uk/g2gp/DbSht_1)
- Florida Digital Archive, Recommended Data Formats for Preservation Purposes in the Florida Digital Archive, **2013**  
<https://libraries.flvc.org/documents/181844/502298/Recommended+Data+Formats/0b25496f-33ac-4f56-9550-12c34f3d5d7c>
- UK Data Archive, Recommended formats, **s. d.**  
<https://www.ukdataservice.ac.uk/manage-data/format/recommended-formats>
- NARA, Transfer Guidance Format, Structured Data Formats, **s. d.**  
<https://www.archives.gov/records-mgmt/policy/transfer-guidance-tables.html - structuredata>
- The National Archives, File formats for transfer, **s. d.**  
<https://www.nationalarchives.gov.uk/information-management/manage-information/digital-records-transfer/file-formats-transfer/>
- Library of Congress, Recommended Formats Statement, VI. Datasets, **s. d.**  
<https://www.loc.gov/preservation/resources/rfs/data.html>
- Archives fédérales suisses, Formats de fichiers adaptés à l'archivage, version 2020/04 **avril 2020**

---

<sup>3</sup> Sans autre indication, les sites et liens mentionnés dans la présente étude ont été consultés la dernière fois le 30.06.2021.

[https://www.bar.admin.ch/dam/bar/fr/dokumente/konzepte\\_und\\_weisungen/archivtaugliche\\_dateiformate.1.pdf.download.pdf/archivtaugliche\\_dateiformate.pdf](https://www.bar.admin.ch/dam/bar/fr/dokumente/konzepte_und_weisungen/archivtaugliche_dateiformate.1.pdf.download.pdf/archivtaugliche_dateiformate.pdf)

Les Archives nationales néerlandaises privilégient **ODS**, CSV et PDF/A pour l'archivage de données de tableurs, alors que XLS et **XLSX** sont qualifiés de formats acceptables.

La Bibliothèque d'Etat de Saxe — Bibliothèque d'Etat et universitaire de Dresde (SLUB Dresden) désigne les formats **OOXML** comme formats adaptés à l'archivage s'ils ont été créés avec MS Office 2013 et suivants. Les formats **ODF** sont qualifiés sans réserve de formats adaptés à l'archivage.

La Bibliothèque de l'EPFZ désigne uniquement CSV et TAB-delimited comme formats adaptés pour une utilisation au-delà de 10 ans, alors qu'OOXML (**XLSX**) est accepté pour une utilisation dans un laps de temps allant jusqu'à 10 ans. **XLS** n'est en revanche plus considéré comme adapté à l'archivage. **ODS** n'est pas mentionné.

Bibliothèque et Archives Canada privilégie ASCII-Text et CSV, mais accepte aussi DBF, EBCDIC, **OOXML**, **ODS** et XLS. Les directives formulent des exigences claires en matière de métadonnées descriptives.

Le Archaeology Data Service privilégie le format CSV, alors que **XLSX** et **ODS** sont choisis lorsqu'il faut conserver des fonctionnalités. Les formats anciens de MS Office et ODF doivent être convertis en XLSX et ODS. Le service formule des exigences pour décrire les données de tableurs. L'examen approfondi de la conservation de données issues de feuilles de calcul désigne **XLSX** et **ODS** comme formats pour la conservation à long terme si certaines propriétés des formats sont prises en compte et que des mesures adéquates sont prises.

Les Archives numériques de Floride classent CSV, TXT et SQL DDL dans le groupe des formats avec « *High Confidence Level* », alors que les formats DBF, **ODS** respectivement SXC et **XLSX** sont classés dans le groupe avec « *Medium Confidence Level* ». XLS fait partie du groupe « *Low Confidence Level* ».

Le centre britannique d'archives de données UK Data Archive désigne CSV et TAB comme formats acceptables pour le partage, la réutilisation et la conservation alors que TXT, XLS, **XLSX** et **ODS**, mais également MBD sont qualifiés de formats acceptables pour la conservation des données.

Les recommandations des services d'archives précités ont à peine évolué depuis 2017. L'étude a réexaminé les recommandations des Archives nationales américaines (NARA), britanniques (National Archives), de la Bibliothèque du Congrès américaine et des Archives fédérales suisses.

Les Archives nationales américaines (NARA) recommandent de privilégier CSV, ODS, ASCII Textm JSON et XML pour les données structurées. Microsoft Excel, Office Open, XML et Microsoft Excel 97 Binary File Format (XLS) sont considérés comme des « formats acceptables ».

Les Archives nationales britanniques (National Archives) énumèrent au côté de CSV, OpenDocument Spreadsheet, ainsi que le format qui l'a précédé OpenOffice Calc, Lotus 1-2-3 Worksheet et Microsoft Excel ainsi que Microsoft Works dans presque toutes les versions.

La Bibliothèque du Congrès privilégie les formats basés sur des caractères et indépendants des plateformes et précise :

« Preferred formats include well-developed, widely adopted, de facto marketplace standards, e.g.

- a. Formats using well known schemas with public validation tool available
- b. Line-oriented, e.g. TSV, CSV, fixed-width
- c. Platform-independent open formats, e.g. .db, .db3 »

Font en outre partie des formats de fichiers privilégiés également « Any proprietary format that is a de facto standard for a profession or supported by multiple tools (e.g. Excel .xls or .xlsx, Shapefile) ». Les jeux de caractères utilisés sont dans l'ordre décroissant UTF-8 et UTF-16 (with BOM), US-ASCII ou ISO 8859-1 et « Other named encoding ».

Les formats de la catégorie « acceptable » sont ceux présentant les propriétés suivantes :

« Non-proprietary, publicly documented formats endorsed as standards by a professional community or government agency, e.g. CDF, HDF » et « Text-based data formats with available schema ».

Les Archives fédérales suisses recommandent CSV comme format de fichiers adapté à l'archivage de tables.

## 4 OOXML et ODF

### 4.1 OOXML

En 2006, Office Open XML (OOXML) a dans un premier temps fait l'objet d'une certification en tant que norme ECMA-376. En 2008, le format a été certifié en tant que norme ISO 29500 avec quatre parties, mais en 2 versions, « *Transitional* » et « *Strict* ». ECMA-376 n'est pas complètement compatible avec ISO 29500. La version actuelle est ISO/IEC 29500:2016.

La version « *Strict* » d'ISO 29500 n'est complètement prise en charge qu'à partir de MS Office 2013. MS Office 2007 prend en charge uniquement ECMA-376 et MS Office 2010 peut certes lire des documents « *Strict* » et « *Transitional* », mais ne peut sauvegarder des documents que dans la version « *Transitional* ».

À propos d'OOXML/XLSX, voir :

- Library of Congress, XLSX Transitional (Office Open XML), ISO 29500:2008-2016, ECMA-376, Editions 1-5  
<https://www.loc.gov/preservation/digital/formats/fdd/fdd000398.shtml>
- Wikipédia, Office Open XML (fr) [https://fr.wikipedia.org/wiki/Office\\_Open\\_XML](https://fr.wikipedia.org/wiki/Office_Open_XML)
- Wikipedia, Office Open XML (en) [https://en.wikipedia.org/wiki/Office\\_Open\\_XML](https://en.wikipedia.org/wiki/Office_Open_XML)

### 4.2 ODF

En 2005, l'Organization for the Advancement of Structured Information Standards (OASIS) a publié le format OpenDocument et l'a adopté en tant que standard. En 2006, le format OpenDocument pour applications Office (OpenDocument) v1.0 a fait l'objet d'une certification en tant que norme ISO 26300. En 2015, la version 1.2 a été certifiée en tant que norme ISO/IEC 26300:2015 en trois parties ; il s'agit de la version actuelle de la norme ISO. OASIS a publié en 2019 ODF 1.3.

À propos d'ODF/ODS, voir :

- Library of Congress, OpenDocument Spreadsheet Document Format (ODS), Version 1.2, ISO 26300:2015  
<https://www.loc.gov/preservation/digital/formats/fdd/fdd000439.shtml>

- Wikipédia OpenDocument (fr)  
<https://fr.wikipedia.org/wiki/OpenDocument>
- Wikipedia OpenDocument (en)  
<https://en.wikipedia.org/wiki/OpenDocument>

### 4.3 Interopérabilité d'OOXML et ODF

Il n'est pas possible de convertir sans autre d'ODF vers OOXML et inversement des fichiers de feuilles de calculs sans perte de mise en page et de fonctionnalités. Nous renvoyons aux recherches de la Bibliothèque du Congrès et du Fraunhofer Institut pour une argumentation détaillée :

- Library of Congress, XLSX Transitional (Office Open XML), ISO 29500:2008-2016, ECMA-376, Editions 1-5, Notes/General : Conversion between XLSX and ODS (<https://www.loc.gov/preservation/digital/formats/fdd/fdd000398.shtml> [page consultée le 01.07.2021])
- Fraunhofer-Institut für Offene Kommunikationssysteme, FOKUS (Hg.), Interoperabilität von Dokumentenformaten: Open Document Format und Office Open XML; white paper, Stuttgart 2009

## 5 Conclusion

Nous pouvons affirmer en conclusion que l'utilisation d'ODF et OOXML comme formats de fichier archivistes est toujours préconisée pour le domaine des feuilles de calcul. Aucun nouveau format pouvant régater avec ces deux formats concernant la diffusion et la normalisation ne se profile à l'horizon. Cette recommandation est toutefois à prendre moyennant deux réserves.

D'une part, toutes les versions de ces deux formats ne peuvent être utilisées de la même façon. Il est recommandé d'utiliser la version « *Strict* » d'ISO 29500 avec OOXML. Par exemple, lorsque l'on utilise les paquets MS Office, il faut employer MS Office 2013 ou les versions plus récentes afin de pouvoir aussi générer et sauvegarder cette version. En cas d'utilisation d'autres outils Office, il faut le spécifier clairement. Avec ODF, il faut privilégier la version 1.2. Celle-ci a été certifiée en tant que norme ISO alors que la version 1.3 n'est reconnue qu'en tant que standard par l'OASIS.

D'autre part, la recommandation d'actualiser les fichiers produits avec d'anciennes versions dans les versions recommandées s'applique toujours ; en particulier pour le format de fichiers binaires XLS et pour les formats qui ont précédé ODF. Peu d'informations sont toutefois disponibles sur les modifications ou pertes d'informations qui en découlent ou sur les conditions générales dans lesquelles cela peut se produire.

Il paraît toujours opportun de se pencher de manière approfondie sur ces deux formats. Nous constatons que la littérature spécialisée comporte très peu d'études actuelles spécifiquement consacrées aux tableurs. La seule référence dans le Cfa jusqu'à la version 6.1 incluse date de 2003. Une nouvelle étude qui résume la situation actuelle et donne de nouvelles références bibliographiques est parue en 2021 : Artefactual Systems, Digital Preservation Coalition (eds.), Preserving Spreadsheets. DPC Technology Watch Guidance Notes, Data Types Series. 2021. <http://doi.org/10.7207/twgn21-09> [page consultée le 08.09.2021]. À cela s'ajoutent les informations de la Bibliothèque du Congrès, Sustainability of Digital Formats : Planning for Library of Congress Collections, [https://www.loc.gov/preservation/digital/formats/fdd/dataset\\_fdd.shtml](https://www.loc.gov/preservation/digital/formats/fdd/dataset_fdd.shtml) [page consultée le 09.09.2021]