Open Document Text (ODT) Format Preservation Assessment

Document History

Date	Version	Author(s)	Circulation
20/09/2016	1.0	Michael Day, Paul Wheatley	External

British Library Digital Preservation Team
digitalpreservation@bl.uk

1. Introduction

This document provides a high-level, non-collection specific assessment of the OpenDocument Text (ODT) file format with regard to preservation risks and the practicalities of preserving data in this format.

The OpenDocument Format is based on the Extensible Markup Language (XML), so this assessment should be read in conjunction with the British Library's generic format assessment of XML [1].

This assessment is one of a series of format reviews carried out by the British Library's Digital Preservation Team. Some parts of this review have been based on format assessments undertaken by Paul Wheatley for Harvard University Library. An explanation of the criteria used in this assessment is provided in italics below each heading.

[Text in italic font is taken (or adapted) from the Harvard University Library assessment]

1.1 Scope

This document will primarily focus on the version of OpenDocument Text defined in OpenDocument Format (ODF) version 1.2, which was approved as ISO/IEC 26300-1:2015 by ISO/IEC JTC1/SC34 in June 2015 [2].

Note that this assessment considers format issues only, and does not explore other factors essential to a preservation planning exercise, such as collection specific characteristics, that should always be considered before implementing preservation actions.

1.2 Summary

OpenDocument Text (ODT) is one of the formats defined in the ISO/IEC 26300 Open Document Format for Office Applications and the OpenDocument Format (ODF) specifications developed and maintained by the Organization for the Advancement of Structured Information Standards (OASIS) consortium. ODT is the ODF format used for the mainly text-based documents produced by word processors and - like the other ODF formats - is based on XML.

OpenDocument Format version 1.2 was approved by ISO/IEC JTC1/SC34 as a three-part international standard in June 2015:

- ISO/IEC 26300-1:2015: Information technology -- Open Document Format for Office Applications (OpenDocument) v1.2 -- Part 1: OpenDocument Schema
- ISO/IEC 26300-2:2015: Information technology -- Open Document Format for Office Applications (OpenDocument) v1.2 -- Part 2: Recalculated Formula (OpenFormula) Format
- ISO/IEC 26300-3:2015: Information technology -- Open Document Format for Office Applications (OpenDocument) v1.2 -- Part 3: Packages

According to its website, OASIS is a non-profit consortium driving "the development, convergence and adoption of open standards for the global information society" [3]. Members include both organisations and individuals, and sponsors and contributors include a wide range of technology companies, as well as a number of universities, research institutes, and government agencies (including the UK Cabinet Office). The OASIS Foundational Sponsors are IBM and Microsoft.

ISO/IEC JTC 1 Information Technology is the ISO technical committee dealing with standards relating to Information and Communication Technology; ISO/IEC JTC 1/SC34 is the sub-committee that covers "Document description and processing languages."

Essentially, ODT is one of several competing formats for the documents created by word processors. The dominant formats in this ecology are the formats associated with Microsoft Word, namely the binary .doc format, and the XML-based .docx format - itself now subject to ISO standardisation as part of the Office Open XML (OOXML) series of standards (ISO/IEC 29500:2012) [4], now often simply referred to by Microsoft as OpenXML. Prior to Microsoft Word 2013, the .docx format did not fully implement either of the two standards defined by OOXML (ISO 29500 Transitional; ISO 29500 Strict), so earlier .docx files are perhaps best seen as implementing the ECMA-376 standard that formed the initial basis of ISO 29500 [5 pp. 34-35]. Both ODF and OOXML formats use ZIP-based containers, although ODT files can in some cases also be saved as flat XML files, with the file-extension .fodt.

2. Assessment

2.1 Development Status

A summary of the development history of the format and an indication of its current status

ODF was developed by OASIS and initial versions were based on the OpenOffice XML file format. *Development began in 2002 and ODF has been standardised as follows:*

- *Version 1.0, approved as an OASIS standard in May 2005; published as an ISO standard in 2006 (ISO/IEC 26300:2006).*
- *Version 1.1, approved as an OASIS standard in February 2007; published as an amendment to ISO/IEC 26300 by ISO in 2012 (ISO/IEC 26300:2006/Amd. 1:2012).*
- *Version 1.2, with new functionality, approved as an OASIS standard in September 2011; published in three parts by ISO in July 2015 (ISO/IEC 26300-1:2015; ISO/IEC 26300-2:2015; ISO/IEC 26300-3:2015).*
- *Version 1.3, in development; although the majority of the additional functionality in this new version is already widely deployed under the name 'ODF 1.2 (Extended)'.*

Open Document Text (ODT) is the word processor focused part of the ODF standard. ODT can be found either as single XML documents or as ZIP based containers, where content, styles, metadata, and application settings are separated into 4 XML files.

The applicable MIME media type for ODT is: application/vnd.oasis.opendocument.text

File extensions are: .odt (typically for ZIP container documents; .fodt (typically for single XML file documents). Other format registry references are referenced by Digipres Commons [6].

2.2 Adoption and Usage

An impression of how widely used the file format is, with reference to use in other memory organisations and their practical experiences of working with the format

Although the sector remains dominated by Microsoft Office, the emergence of OpenOffice (and popular forks such as LibreOffice) has provided an open source alternative that natively supports the ODF standard.

In 2014, ODF was selected by the UK Cabinet Office as one of the standard document file formats to be used across all UK government bodies, PDF/A and HTML being the preferred formats for viewing documents, ODF "for sharing or collaborating on government documents" [7]. The rationale given was largely based on interoperability, e.g. that citizens, businesses and other organisations would no longer need specialist hardware or software in order to work with government documents, also that government organisations would be able to share and work with documents using the same formats. The version of ODF mandated is version 1.2 (or later), as expressed in a set of Cabinet Office guidelines [8]:

> *"The default format for saving government documents must be Open Document Format (ODF). Information should be shared in ODF version 1.2 (or later). ODF version 1.1   may be used for transition to the implementation of ODF 1.2. Where users need to calculate formulas in spreadsheets, ODF 1.2 (or later) is preferred for better interoperability. ODF includes filename extensions such as .odt for text, .ods for spreadsheets and .odp for presentations."*

Additional Cabinet Office guidelines are clear that "UK government organisations must use ODF for creating editable documents." In addition to interoperability, additional features deemed to be useful include its mechanism for track changes, the ability to add digital signatures, and its handling of metadata (including linked data) [9]. The Cabinet Office also supports ODF as a preservation format, the guidelines asserting the superiority of XML over binary formats for the future-proofing of content [9].

Most UK government departments are currently dependent on the use of proprietary software. For example, a 2015 adoption plan published by the UK Home Office noted that the department and its associated agencies remained heavily dependent on Microsoft Office 2007 as its "primary office productivity solution" [10]. While there was minimal use of GoogleDrive for collaborative working, the department noted that open source applications like LibreOffice or Apache OpenOffice were not currently available on government computers. The report also noted the department's difficulties with working with ODF within its existing technical infrastructures:

> *Microsoft Office 2007 supports ODF 1.1 but does not implement some features, such as tracked changes, that you can use with other software implementing ODF 1.1. Currently Google Apps*

only supports ODF 1.1. Microsoft Office 2007 supports export to PDF/A but Google Drive only supports export to standard PDF 1.4, which would require a further step in the workflow. By contrast, both LibreOffice and OpenOffice currently support ODF 1.2 and export to PDF/A.

As part of its ODF adoption plan, the Home Office is not aiming to modernise legacy systems in order to ensure that they work with open formats like ODF and PDF/A, so that will become part of a wider technology procurement process for a replacement for Office 2007. There is also no plan to convert legacy content in Microsoft's proprietary formats to ODF.

A similar report on another UK government department, HM Revenue & Customs (HMRC), suggested that they were not planning an immediate move to open source software solutions. In early 2015, the department was still upgrading their systems to Windows 7 and Microsoft Office 2013, the report noting that Office 2013 "can save and open documents in ODF" [11] HMRC were also planning to conduct a review of all of their in house software (known as Business Developed Applications) to assess their compatibility with ODF and to consider the conversion of legacy content whenever these were "refreshed or replaced."

Worldwide, a large number of governmental organisations have moved towards mandating the use of ODF for the sharing of text documents and other office-type files, usually as part of a general move towards the adoption and use of open standards [12]. The State of Massachusetts was one of the first large governmental organisations to propose the mandating of ODF back in 2005, but, following a fierce public debate, it backtracked and permitted the additional use of the recently standardised OOXML from 2007 [13]. Belgium approved the use of ODF as the standard for exchanging office documents within its civil service in 2006, although it also established a transition period to enable migration to happen without compromising service continuity [14]. Since 2012, The Netherlands has identified ODF 1.2 as one of the formats in its 'Comply or Explain' list of open standards [15]; this means that the standard must be implemented, or "an explanation must be provided as to why" [16 p. 5]. In 2009, the Norwegian Ministry of Government Administration and Reform established a set of open standards for the publication of government information which specified the use of ODF 1.1 (or PDF) for the exchange of government information by email [17]. However, more recent Norwegian guidelines (v 4.1, 2014) have mandated the use of specific versions of HTML and PDF for publication, while simply recommending the use of open standards (which would include both ODF and OOXML), when "publishing a text for further processing" [18]. Like Norway, many administrations seem to be happy now to work with multiple open standards, although they typically only cite ODF and OOXML. For example, in 2011 the European Union's Inter-Institutional Committee for Informatics recommended that European institutions use either (or both) ODF and OOXML for revisable documents,[1] the key element being the *choice* of citizens or member states [19]:

> *In cases where documents have to be exchanged using revisable document formats, the principle to be applied by analogy is the same as when dealing with multilingualism. Citizens and the Institutions' partners (e.g. Member States) should be put at the centre, and the Institutions should take all appropriate measures to be able to provide them with documents in the standard format of their choice.*
>
> *The minimum requirement is to support XML-based revisable document formats standardised by the International Organization for Standardization, namely:*
>
> *– Open Document Format for Office Applications, or ODF (ISO/IEC 26300:2006).*
> *– Office Open XML, or OOXML (ISO/IEC 29500:2008).*
>
> *In addition, the Institutions are encouraged to support, on a best effort basis, other widely used document formats.*

Other administrations leave the choice even more open. For example, public bodies in Denmark have since 2011 been committed to *receiving* documents in "all common formats" – including ODF and OOXML – although it was not made mandatory for public bodies to *distribute* editable documents in those formats, as the standards were considered by an expert group to be too immature [20].

One major factor in supporting more recent organisational transitions to ODF has been, as HMRC noted, the ability of Microsoft Office 2013 to both open and save files in ODF (although Microsoft's own documentation notes that, "formatting might be lost when users save and open .odt files" [21]). Microsoft continue to promote the OpenXML format, and the company objected very strongly to the UK's adoption of ODF and HTML as its standard document formats in 2014, arguing that the policy should have

---

[1] PDF and PDF/A remained the Committee's recommendation for non-revisable documents [19].

included both ODF and OOXML [22]. Open standards advocate Andrew Updegrove has commented on the significance of the UK government's promotion of a single standard for sharing and collaboration [13]:

> *The adoption of only ODF - and not also OOXML, as was urged by Microsoft - is most significant for the degree of market adoption and legitimization it will necessarily lead to. Unlike OOXML, which exists in several versions in the marketplace (and was not adopted even by Microsoft in the ISO/IEC approved "strict" form until years after its adoption by those organizations), ODF exists in more settled forms. Use of ODF-compliant software by tens of thousands of U.K. government workers will provide incentives to Microsoft to take greater pains to ensure that documents saved in ODF form will preserve their formatting with greater integrity, since many open source office suites (such as LibreOffice and OpenOffice) are available for free.*

Minutes from meetings relating to the UK government's adoption of open standards are available from the Standards Hub webpages [23].

Support for the use of ODF at governmental level has largely been driven by the open standards agenda. However, the open nature of ODF and its basis in XML means that ODF has also been recommended by a number of organisations as a preservation format. These include:

- *Archaeology Data Service: "As an open XML-based format, ODT is suitable for both deposit and preservation though, in the latter case, the files should be stored in their uncompressed form."* [24].
- The National Archives's Guide to Formats, notes that the ODF family of standards "are designed to be highly re-usable and interoperable," that ODT documents can be accessed by many applications (including Microsoft Word and OpenOffice), and argues that its resilience (defined as recoverability, ubiquity, and stability) is either "very high" or "above average" [5].
- *National Archives and Records Administration: ODF is listed as one of several "Preferred Formats" for text.* [25].
- *Library and Archives Canada: ODF is one of several "Preferred Formats" for text.* [26].
- *The National Archives of Australia: ODF is listed as a "Preferred open file format" and is a target for normalisation from most text-based formats.* [27].
- *Library of Congress: The Recommended Format Statement (RFS) 2015-2016 views all versions of ODF as an "Acceptable" (rather than "Preferred") format for textual works in in digital form, and is second in the first group of "other structured markup formats" in a list described as "in descending order of preference"* [28]. In general, however, the Library of Congress has a stated preference for "formats intended for final publication of textual works, rather than editable formats" [29].
- *Archivematica: Lists ODF as a "Preservation Format" and notes that ODF is a target format for normalisation from some other text based formats* [30].

Recommendations aside, it is difficult to get a feel for the exact numbers of ODF files actually stored in long-term repositories. In his observations on the Library of Congress Format Sustainability assessments of ODF, Carl Fleischhauer noted that examples of ODF and OOXML would be "most-frequently encountered as born-digital segments within collections of personal papers and organizational records, the types of unpublished materials that are acquired by the Library's special collection divisions" [31]. Outside of these types of special collections, it is unclear, therefore, whether ODF would be a format likely to be frequently encountered by national libraries. Suggestive, perhaps, might be the 2014 list of the 50 most prevalent formats stored in the eDepot of the Koninklijke Bibliotheek (National Library of the Netherlands), which did not include any ODF files[2] - although the list included 12m XML files (apparently mostly metadata) and 20k .zip files, as well as a large number of office files in a variety of proprietary or OpenXML-based formats (134.7k .doc, 40.3k .docx, 31.2k .xls, 8.3k .ppt, 3.4k .pptx; also 12.3k RTF) [32] [33] [34]. According to Johan van der Knijff, many of the office files in eDepot are supplementary material to scientific papers [32].

2.3 Software Support

2.3.1 Rendering Software Support

An overall impression of software support for rendering the format with reference to: typical desktop software; and current support on British Library reading room PCs

---

[2] The full list of file extensions in eDepot included the sum total of 2 .odt files [34].

The ODF family of standards were designed with interoperability in mind. ODF formats can be opened in a wide range of different office applications, although the National Archives *Guide to Formats* (2011) notes that "minor changes to formatting may occur in different applications" *[5 p. 33]*. The Community for Open Interoperability Standards (COIS), part of Open Forum Europe, has listed the following software as having full or partial support for ODF 1.2 *[35]*:

- Apache OpenOffice (Open Source) – full support
- Google Docs (proprietary, cloud-based) – partial support
- LibreOffice (Open Source) – full support
- LibreOffice-from-Collabora (Proprietary) – full support
- Microsoft Office 2013 (proprietary) – partial support
- Microsoft Office 365 (proprietary, cloud based) – partial support

Issues

Partial support for ODF does not mean, however, complete interoperability. COIS notes that partial support means that applications may not support all aspects of ODF. For example, they note problems with tracked changes in Microsoft Office [35]:

> *Vital to workgroup interaction, any Tracked Changes will be lost in the Microsoft Office implementations of ODF. Tracked changes become ordinary Body Text with no indication that they represent historical edits, or record of who is responsible.*

Indeed, Microsoft Office's own documentation notes that, "formatting might be lost when users save and open .odt files" [21].

By comparison, Open Source applications like Apache OpenOffice [36] and LibreOffice [37] provide full support for ODF.

2.3.2 Preservation Software Support

An impression of the availability and effectiveness of software for managing and preserving instances of the file format

Format identification

Identification of ODT 1.2 is supported by both Apache Tika [38] and DROID. The PRONOM registry [39] notes that ODT 1.2 can be identified "via a container signature in DROID version 6 or later." The PRONOM database cannot currently represent container signatures. The PRONOM entry for ODT 1.2 also provides a reminder that ODF files are usually packaged in ZIP:

> *ODF is an XML format, which may comprise a single XML document, or a collection of subdocuments within a package. The package format is the commonly used form, and comprises a standard ZIP package containing the XML content and any associated binary data, together with a manifest which lists the package content. An ODF text document contains word-processed text. The content is contained in a 'content' xml subdocument. Additional subdocuments contain style information, document metadata, and application-specific settings.*

Validation and Detecting Preservation Risks

The Apache ODF Toolkit contains an ODF Validator, currently with "incubation" status, which means that as a project-in-progress, it has not as yet been endorsed by the Apache Software Foundation. Apache ODF Validator is described as "a tool that validates OpenDocument files and checks them for certain conformance criteria" [40]. The OpenDoc Society provides an online ODF Validator based on the Apache ODF Validator and Sun's Multi-Schema XML Validator (MSV), although it does not as yet cover all conformance criteria [41].

The UK Cabinet Office has produced some useful guidance on ODF validation and compliance testing. It notes that, "no single validator covers all conformance criteria of the ODF specification," and recommends using several validators in parallel [42]. In addition to the online validator provided by the OpenDoc Society, it notes the existence of downloadable validators like Cyclone3 [43] and Office-o-Tron [44] [45], also the validators provided by RelaxNG [46].

Conformance Checking

The Officeshots tool from the OpenDoc Society and the Netherlands government provides a way to check how applications render ODF files. The UK Cabinet Office elaborates [42]:

If you're considering a particular application, Officeshots can provide an initial but thorough check to see it handles ODF in the way it claims to. It allows you to compare the output and behaviour of a variety of applications, based on documents that are relevant to your organisation rather than on a synthetic benchmark. For instance, do the templates produced by your communications department or design agency look consistent across all the applications they are likely to be used in? How do your outgoing letters look if you view or print them in EuroOffice 2014, different versions of Microsoft Office or Google Docs?

Officeshots is available as an online service [47] and as an open source tool [48]

Metadata Extraction

The open and XML based nature of ODF would, in principle, make metadata extraction straightforward.

Migration

Conversion between ODF and OOXML is complex and a significant range of functionality has been identified as challenging to retain during this process [49]. Support for ODF varies between the main applications with Microsoft Word currently lagging behind [50]. A list of supported ODF functionality in Word is provided by Microsoft [51].

2.4 Documentation and Guidance

An indication of the availability of practical documentation or guidance with specific reference to the facilitation of any recommended actions

OpenDocument Format version 1.2 is published by ISO [2] and is one of the standards available for free download from the ISO Information Technology Task Force [52]. It is also freely available from the OASIS web site [53].

The official community gathering place and information resource for ODF used to be the OpenDocument XML.org web site [54], but this now redirects to the OASIS Open Document Format for Office Applications (OpenDocument) TC [55].

2.5 Complexity

An impression of the complexity of the format with respect to the impact this is likely to have on the British Library managing or working with content in this format. What level of expertise in the format is required to have confidence in management and preservation?

While XML is ultimately human readable, ODF is a rather complex format (although relatively simple documents may not need to use all aspects of it). We have already noted that Tracked Changes may not render correctly in all applications. As with other XML-based formats, other concerns mainly relate to embedded content (e.g. fonts, images, data, etc.).

2.6 Embedded or Attached Content

The potential for embedding or attaching files of similar or different formats, and the likely implications of this

Missing fonts

Embedding fonts in ODTs has not been possible using the major word processing applications, such as LibreOffice and OpenOffice, due to confusion over the ODF standard. It was widely understood that font embedding was not possible, and hence was not implemented in word processor applications, until it was realised that embedding was possible as part of the SVG structure described in ODF 1.0. It appears that more considered support for embedding fonts may be a possibility for ODF 1.3 [56].

LibreOffice has officially supported font embedding since version 4.1 [57], but only when run on Windows and Linux. Furthermore, some basic testing suggests that there are issues with the implementation. OpenOffice remains without the ability to embed fonts in ODTs [58].

The most-recent stable "Still" version of LibreOffice is currently (January 2015) version 4.4.7 (December 2015), with additional versions ("Fresh" v. 5.0.4, preview "Beta" v. 5.1.0) also available.

Referenced or embedded images

Pasting an image from some web browsers into an ODT in OpenOffice results in a reference by URL rather than embedding the image, which would have been a more useful default setting [59]. *Referencing images by URL could be implemented by users regardless of whether this bug existed or not, but its presence is likely to have increased the incidence of this issue. As of November 2014, the bug has still not been fixed* [60], *although a similar bug in LibreOffice was fixed in February 2014.*

2.7 External Dependencies

An indication of the possibility of content external to an instance of the file format that is complimentary or even essential to the intellectual content of the instance

See section 2.6 on Embedded or Attached Content.

2.8 Legal Issues

Legal impediments to the use, management or preservation of instances of the file format

Sun Microsystems has made a covenant available providing a guarantee it will not enforce patents relating to ODF. IBM's ODF contributions are also covered by a similar covenant via their Interoperability Specifications Pledge. Independent review by the Software Freedom Law Center concluded that "ODF, as standardized and licensed by the Organization for the Advancement of Structured Information (OASIS), is free of legal encumbrances that would prevent its use in free and open source software, as distributed under licenses authored by Apache and the FSF". [61]

2.9 Technical Protection Mechanisms

Encryption, Digital Rights Management and any other technical mechanisms that might restrict usage, management or preservation of instances of the file format

None known – although ODF files could presumably be protected by DRM, if that was required

2.10 Other Preservation Risks

Other evidence based preservation risks, noting that many known preservation risks are format specific and do not easily fit under any of the sustainability factors above

None known.

2.11 Preservation Risk Summary

A summary of preservation risks and recommended actions (where possible).

As part of the ODF 1.2 family of standards, ODT is often seen as a preferred choice for a preservation format. This is largely based on the open nature of the standard and the technological basis of the format in XML and ZIP (for packaging). All specifications are freely available and the format can be rendered by a wide-range of office-type programs, although there may be inconsistencies in how individual applications actually support the standard. As an open standard, ODF has been recommended for use by many government organisations worldwide and is the standard specified by the UK Cabinet Office for sharing or collaborating on government documents (while PDF/A and HTML are the recommended standards for viewing documents). The Library of Congress Sustainability of Digital Formats assessment of ODT 1.2 commented that, in general, the Library of Congress "prefers formats intended for final publication of textual works, rather than editable formats" [29]. Consideration, therefore, could be given to the possibility of converting content in ODT into publication formats like PDF or HTML, although this would have to take into account questions of preservation intent.

From a preservation perspective, some of main challenges with ODF include:

Missing font information
- *Where not easily substituted, non-embedded fonts could lead to loss of critical information*
- *Lack of support for font embedding in many tools means that most ODTs will not have embedded fonts*

External References
- *Externally referenced content may be difficult to collect/preserve*

Embedded Resources
- *Resources can be embedded within the main bytestream, each of which may present new preservation issues.*

Rendering support varies
- *Rendering in software different to the creating application may be imperfect*

Encryption
- *Content may be encrypted*

Mitigation of key risks might include:

Mitigating encryption, missing fonts, external references risks
- *Examine content on ingest to check for the presence of these dependencies*

Mitigating varied rendering support
- *Migrate content to alternative easily rendered and stable format (PDF/A) and offer as an alternative for the user; Users within a corporate setting may well render with Microsoft Office, where import and rendering could alter content in significant ways. Providing a PDF as an alternative to the original file (as well as acting to some extent as a benchmark for the original,) along with a short explanation of this issue, would allow the user to make an informed choice.*

3. Recommendations for Action

Recommended actions in usage and handling of the format. Recommend actions in the support or development of software applications that provide, or have the potential to provide, significant risk mitigation for the format. Note that these recommendations do not take into account other requirements such as those driven by specific British Library collections, or non-preservation issues such as resourcing.

When recommended by governments, ODT is often seen as the standard for creating and collaborating on documents, while content would be made available for reading in other formats (in the UK, these would be PDF/A and HTML). It might be useful to review any ODT-based content entering the Library, to see whether the identical content is available in other formats.

Handling Recommendations

- All ODT files received by the Library should be identified (by ODF version) and validated, e.g. using tools like Apache ODF Validator
- The presence of DRM should be identified and, wherever possible, removed
- Likely dependencies on non-embedded fonts, externally-referenced content, and embedded resources should be identified and investigated prior to ingest.

Software Recommendations

- Test and evaluate the use of ODF identification and validation tools (e.g. Apache Tika, DROID, Apache ODF Validator, etc.), prior to ingest of any ODT files.
- Test the migration of ODT files to other "delivery" formats, e.g. PDF.

Monitoring Recommendations

- Review ODT and the other ODF formats on an annual basis, as the format and the applications used to render it are under continual development

4. References

1. **British Library Digital Preservation Team.** XML Format Preservation Assessment. *Digital Preservation Coalition wiki.* [Online] 01 May 2015. [Cited: 04 Nov 2015.] http://wiki.dpconline.org/index.php?title=File_Formats_Assessments.
2. **ISO/IEC 26300-1:2015.** Information technology -- Open Document Format for Office Applications (OpenDocument) v1.2 -- Part 1: OpenDocument Schema. *International Organization for Standardization.* [Online] 2015. [Cited: 04 Nov 2015.] http://www.iso.org/iso/home/store/catalogue_ics/catalogue_detail_ics.htm?csnumber=66363.
3. **OASIS.** About Us. [Online] n.d. [Cited: 04 Nov 2015.] https://www.oasis-open.org/org.
4. **ISO/IEC 29500-1:2012.** Information technology -- Document description and processing languages -- Office Open XML File Formats -- Part 1: Fundamentals and Markup Language Reference. [Online] 2012. [Cited: 04 Nov 2015.] http://www.iso.org/iso/home/store/catalogue_ics/catalogue_detail_ics.htm?csnumber=61750.
5. **The National Archives.** A guide to formats. [Online] 2011. [Cited: 04 Nov 2015.] http://www.nationalarchives.gov.uk/documents/information-management/guide-to-formats.pdf.
6. **Digipres Commons.** .docx format aggregation. [Online] 2014. [Cited: 04 Nov 2015.] http://www.digipres.org/formats/extensions/#*.odt.

7. **UK Cabinet Office.** Worldwide engagement leads to standards that get people working together. [Online] 22 Jul 2014. [Cited: 04 Nov 2015.] https://www.gov.uk/government/news/open-document-formats-selected-to-meet-user-needs.

8. —. Sharing or collaborating with government documents. *GOV.UK.* [Online] 6 Aug 2015. [Cited: 04 Nov 2015.] https://www.gov.uk/government/publications/open-standards-for-government/sharing-or-collaborating-with-government-documents.

9. —. An introduction to Open Document Format (ODF). *GOV.UK.* [Online] 27 Mar 2015. [Cited: 04 Nov 2015.] https://www.gov.uk/government/publications/open-document-format-odf-an-introduction/an-introduction-to-open-document-format-odf.

10. **UK Home Office.** Home Office Open Document Format adoption plan. *GOV.UK.* [Online] 17 Sep 2015. [Cited: 04 Nov 2015.] https://www.gov.uk/government/publications/home-office-open-document-format-adoption-plan/home-office-open-document-format-adoption-plan.

11. **UK HM Revenue & Customs.** Implementing open standards for documents. *GOV.UK.* [Online] 25 Feb 2015. [Cited: 04 Nov 2015.] https://www.gov.uk/government/publications/hmrcs-plan-for-implementing-open-standards-for-documents/implementing-open-standards-for-documents.

12. **Wikipedia.** OpenDocument adoption. *Wikipedia.* [Online] [Cited: 04 Nov 2015.] https://en.wikipedia.org/wiki/OpenDocument_adoption.

13. **Updegrove, Andy.** U.K. Cabinet Office Adopts ODF as Exclusive Standard for Sharable Documents. *The Standards Blog.* [Online] 22 Jul 2014. [Cited: 04 Nov 2015.] http://www.consortiuminfo.org/standardsblog/article.php?story=20140722125241786.

14. Open standards: Belgium's federal Council of Ministers approves ODF (Open Document Format). *Presscenter.org.* [Online] 23 Jun 2006. [Cited: 04 Nov 2015.] http://presscenter.org/fr/node/47487?lang=en.

15. **Netherlands. Forum Standaardisatie.** ODF 1.2. *Forum Standaardisatie | Open standaarden.* [Online] 2012. [Cited: 04 Nov 2015.] https://lijsten.forumstandaardisatie.nl/open-standaard/odf12.

16. **Programma Nederland Open in Verbinding & Forum Standaardisatie.** Open Document Standards for the Goverment Guide. [Online] 13 Apr 2011. [Cited: 04 Nov 2015.] https://www.forumstandaardisatie.nl/fileadmin/os/documenten/Handreiking_ODF_Engelse_versie.pdf.

17. **Norway. Fornyings- og administrasjonsdepartementet.** Referansekatalog for IT-standarder i offentlig sektor, v2.0. [Online] 25 Jun 2009. [Cited: 25 Jan 2016.] https://www.regjeringen.no/globalassets/upload/FAD/Vedlegg/IKT-politikk/Referansekatalogen_versjon2.pdf.

18. **Norway. Direktoratet for forvaltning og IKT.** Referansekatalog for IT-standarder i offentlig sektor, v4.1. *Standardiseringsportalen.* [Online] 03 Mar 2014. [Cited: 25 Jan 2016.] http://standard.difi.no/forvaltningsstandarder/referansekatalogen-html-versjon/.

19. **European Union. Inter-Institutional Committee for Informatics.** Conclusions on document exchange format following the discussion of office automation platforms. [Online] 25 Jul 2011. [Cited: 25 Jan 2016.] http://ec.europa.eu/dgs/informatics/oss_tech/pdf/2011-07-25_ares.pdf. Ref. Ares(2011)80u658 - 25/07/2011.

20. **European Union.** eGovernment in Denmark, v16.0. [Online] Jun 2014. [Cited: 25 Jan 2016.] https://joinup.ec.europa.eu/sites/default/files/7f/1d/a4/eGov%20in%20DK%20-%20June%202014%20-%20v.16.pdf.

21. **Microsoft.** File format reference for Word 2013, PowerPoint 2013, and Excel 2013. *Microsoft TechNet.* [Online] 02 May 2014. [Cited: 04 Nov 2015.] https://technet.microsoft.com/en-us/library/dd797428(v=office.15).aspx.

22. **Glick, Brian.** Microsoft attacks UK government decision to adopt ODF for document formats. *ComputerWeekly.com.* [Online] 23 Jul 2014. [Cited: 04 Nov 2015.] http://www.computerweekly.com/news/2240225262/Microsoft-attacks-UK-government-decision-to-adopt-ODF-for-document-formats.

23. **Standards Hub.** Our meetings. *UK Government Digital Service.* [Online] [Cited: 04 Nov 2015.] http://standards.data.gov.uk/meetings.

24. **Archaeology Data Service; Digital Antiquity.** Section 2. Creating Texts and Documents. *Documents and Digital Texts: A Guide to Good Practice.* [Online] n.d. [Cited: 04 Nov 2015.] http://guides.archaeologydataservice.ac.uk/g2gp/TextDocs_2.

25. **National Archives and Records Adminstration.** NARA 2014-04: Appendix A, Revised Format Guidance for the Transfer of Permanent Electronic Records – Tables of File

Formats. [Online] 03 Mar 2014. [Cited: 04 Nov 2015.] http://www.archives.gov/records-mgmt/policy/transfer-guidance-tables.html#textualdata.

26. **Library and Archives Canada.** Guidelines on File Formats for Transferring Information Resources of Enduring Value. [Online] 01 Oct 2014. [Cited: 04 Nov 2015.] http://www.bac-lac.gc.ca/eng/services/government-information-resources/guidelines/Pages/guidelines-file-formats-transferring-information-resources-enduring-value.aspx.

27. **National Archives of Australia.** Dissecting the Digital Preservation Software Platform, Version 1.0. [Online] 2009. [Cited: 04 Nov 2015.] http://naa.gov.au/Images/Digital-Preservation-Software-Platform-v1_tcm16-47139.pdf.

28. **Library of Congress.** Recommended Formats Statement. [Online] n.d. [Cited: 04 Nov 2015.] http://www.loc.gov/preservation/resources/rfs/textmus.html.

29. —. OpenDocument Text Document Format (ODT), Version 1.2, ISO 26300-1:2015. *Sustainability of Digital Formats Planning for Library of Congress Collections.* [Online] 24 Sep 2015. [Cited: 04 Nov 2015.] http://www.digitalpreservation.gov/formats/fdd/fdd000428.shtml.

30. **Archivematica.** Word processing files. [Online] 26 Nov 2013. [Cited: 04 Nov 2015.] https://www.archivematica.org/wiki/Word_processing_files.

31. **Fleischhauer, Carl.** ODF: The Open Document Format. *The Signal.* [Online] 19 Jan 2016. [Cited: 25 Jan 2016.] https://blogs.loc.gov/digitalpreservation/2016/01/odf-the-open-document-format/.

32. **Knijff, Johan van der.** Top 50 file formats in the KB e-Depot. *KB Research blog.* [Online] 29 Apr 2015. [Cited: 04 Nov 2015.] http://blog.kbresearch.nl/2015/04/29/top-50-file-formats-in-the-kb-e-depot/.

33. —. 50 most prevalent formats in KB e-Depot by file extension, based on March 2014 count. *GitHub Gist.* [Online] 2014. [Cited: 04 Nov 2015.] https://gist.github.com/bitsgalore/21028de28b7f05066585#file-extensionskbdm-md.

34. —. File extensions in KB e-Depot, status March 2014. *GitHub Gist.* [Online] 2014. [Cited: 04 Nov 2015.] https://gist.github.com/bitsgalore/4326300f185eec3d6d48#file-edepotfextentions_v3-md.

35. **COIS Community for Open Interoperability Standards.** Open Document Format principles for government technology. *Open Forum Europe.* [Online] 2015. [Cited: 04 Nov 2015.] http://www.openforumeurope.org/wp-content/uploads/2015/09/OFE-COIS-Open-Document-Format-principles-for-Government-Technology-11th-Plugfest-edition-booklet-EN-20150915-proof.pdf.

36. Apache OpenOffice. [Online] [Cited: 04 Nov 2015.] https://www.openoffice.org/.

37. LibreOffice. *The Document Foundation.* [Online] [Cited: 04 Nov 2015.] https://www.libreoffice.org/.

38. Apache Tika: Supported Document Formats. *Apache Software Foundation.* [Online] [Cited: 04 Nov 2015.] https://tika.apache.org/1.0/formats.html#OpenDocument_Format.

39. **The National Archives.** Details for: OpenDocument Text 1.2. *PRONOM: The Technical Registry.* [Online] 19 Apr 2012. [Cited: 04 Nov 2015.] http://apps.nationalarchives.gov.uk/PRONOM/Format/proFormatSearch.aspx?status=detailReport&id=1034.

40. ODF Validator. *Apache Incubator.* [Online] [Cited: 04 Nov 2015.] http://incubator.apache.org/odftoolkit/conformance/ODFValidator.html.

41. ODF Validator. *OpenDoc Society.* [Online] [Cited: 04 Nov 2015.] https://odf-validator.rhcloud.com/.

42. **UK Cabinet Office.** Open Document Format (ODF): validators and compliance testing. *GOV.UK.* [Online] 27 Mar 2015. [Cited: 04 Nov 2015.] https://www.gov.uk/government/publications/open-document-format-odf-validators-and-compliance-testing/open-document-format-odf-validators-and-compliance-testing.

43. Cyclone3. [Online] [Cited: 04 Nov 2015.] https://www.openhub.net/p/cyclone3.

44. Office-o-tron. [Online] [Cited: 04 Nov 2015.] https://www.openhub.net/p/officeotron.

45. Office-o-tron. [Online] [Cited: 04 Nov 2015.] http://odf-validator2.rhcloud.com/officeotron2/.

46. RELAX NG home page. [Online] 25 Feb 2014. [Cited: 04 Nov 2015.] http://relaxng.org/#validators.

47. Officeshots. [Online] [Cited: 04 Nov 2015.] http://www.officeshots.org/.

48. Officeshots. [Online] [Cited: 04 Nov 2015.] http://code.officeshots.org/trac/officeshots.

49. **ISO/IEC TR 29166:2011.** Information technology -- Document description and processing languages -- Guidelines for translation between ISO/IEC 26300 and ISO/IEC 29500 document formats. *International Organization for Standardization.* [Online] 2011.

[Cited: 04 Nov 2015.]
http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=45245.
50. **Bullock, Ted.** ODF 1.2 Support in Microsoft Office. *Depth of Knowledge blog.* [Online]
19 Apr 2011. [Cited: 04 Nov 2015.] http://tbullock.comlore.com/2011/04/odf-12-support-in-microsoft-office.html.
51. **Microsoft.** Differences between the OpenDocument Text (.odt) format and the Word
(.docx) format. [Online] n.d. [Cited: 04 Nov 2015.] https://support.office.com/en-gb/article/Differences-between-the-OpenDocument-Text-odt-format-and-the-Word-docx-format-d9d51a92-56d1-4794-8b68-5efb57aebfdc?ui=en-US&rs=en-GB&ad=GB.
52. **International Organization for Standardization.** Freely Available Standards.
[Online] n.d. [Cited: 04 Nov 2015.]
http://standards.iso.org/ittf/PubliclyAvailableStandards/.
53. **OASIS.** OASIS Open Document Format for Office Applications (OpenDocument) TC.
[Online] 2015. [Cited: 04 Nov 2015.] https://www.oasis-open.org/committees/tc_home.php?wg_abbrev=office.
54. OpenOffice XML.org. [Online] [Cited: 04 Nov 2015.] http://opendocument.xml.org/.
55. OASIS Open Document Format for Office Applications (OpenDocument) TC. *OASIS.*
[Online] [Cited: 04 Nov 2015.] https://www.oasis-open.org/committees/tc_home.php?wg_abbrev=office.
56. **OASIS.** Public Comment: Font Embedding. *OASIS.* [Online] 2014. [Cited: 04 Nov
2015.] https://issues.oasis-open.org/browse/OFFICE-228.
57. **LibreOffice.** LibreOffice 4.1 ReleaseNotes. *The Document Foundation wiki.* [Online]
2014. [Cited: 04 Nov 2015.] https://wiki.documentfoundation.org/ReleaseNotes/4.1.
58. **Apache OpenOffice.** Issue 20370 - Q-PCD MSInteroperability-33: embedd fonts into
the document. *Apache OpenOffice (AOO) Bugzilla.* [Online] 2013. [Cited: 04 Nov 2015.]
https://bz.apache.org/ooo/show_bug.cgi?id=20370.
59. **Knijff, Johan van der.** Images not embedded because of paste as link bug in
OpenOffice and LibreOffice. [Online] 12 Feb 2014. [Cited: 04 Nov 2015.] http://wiki.opf-labs.org/display/TR/Images+not+embedded+because+of+paste+as+link+bug+in+OpenOffice+and+LibreOffice.
60. **Apache OpenOffice.** Issue 37652 - set default options for pasting images from HTML
content (embed graphics instead of linking them). *Apache OpenOffice (AOO) Bugzilla.*
[Online] 2015. [Cited: 04 Nov 2015.] https://bz.apache.org/ooo/show_bug.cgi?id=37652.
61. **Moglen, Eben.** OpenDocument Opinion Letter. *Software Freedom Law Center.* [Online]
12 Jul 2006. [Cited: 04 Nov 2015.]
http://www.softwarefreedom.org/resources/2006/OpenDocument.html.

< - end - >