



Documents and digital texts: A guide to good practice



22nd February 2023





The research leading to these results has received funding from the European Community's Horizon 2020 Programme (H2020-INFRAIA-2018-1) under grant agreement n° 823914.

For any questions regarding this Guide, please email: contact@ariadne-infrastructure.eu

PDF/A-1 (ISO 19005-1), created in MS Word (Version 2016) from the docx format.

Contents

1. Introduction to documents and texts	5
1.1 What are documents and texts?	5
2. Things to consider when creating texts and documents	7
2.1 General considerations	7
Avoid embedded material.....	7
3. Archiving texts and documents.....	8
3.1 Deciding which files to archive.....	8
3.2 Deciding how to archive.....	8
Significant properties	8
File format for archiving.....	8
3.3 Metadata and documentation	9
4 File formats.....	12
Format for marking up text.....	17

Documents and digital texts: A guide to good practice

Acknowledgements: Based on Archaeology Data Service's Guides to Good Practice on Documents and digital texts (<https://archaeologydataservice.ac.uk/help-guidance/guides-to-good-practice>). The information has been reworked, updated and supplemented by the Swedish National Data Service (<https://snd.gu.se/en>) to be more appropriate to data of non-archaeological origin.

1. Introduction to documents and texts

The most common file types created during a research project are various types of documents and text files. Whatever the type of research project, a final report in the form of a text document will be written, if nothing else. Apart from reports, documents that describe various work processes, surveys, metadata documents, etc., are normally created.

This guide is intended to provide an overview of diverse types of binary¹ and plain text files.² In addition to reviewing common file types and file types suitable for archiving, this guide will discuss the different elements/objects that make up a document and how different ways of creating elements affect them. We will also discuss the archival strategy that may be used to ensure that these are maintained after preparation for archiving.

1.1 What are documents and texts?

Put simply, text documents may range in size and complexity from simple reports and short essays to much larger documents, such as dissertations or books. Text files consist of structured text (sentences, paragraphs, pages, chapters) but often include other elements such as images, figures, and tabulated data.

Text documents may be created in different ways, but most are created using various word processing programmes such as Microsoft Word or OpenOffice-based programmes (LibreOffice, Apache OpenOffice, NeoOffice, etc.). In terms of file formats, documents created with word processing programmes used to be saved in proprietary³ binary formats, but there has been a shift towards xml-based formats and human-readable standards such as .docx (Office Open XML⁴ format) and .odt (OpenDocument⁵ format). In addition to these formats, many final versions of documents are saved (by the user) in system-independent formats, usually Adobe's Portable Document Format (PDF),⁶ which allows the format and structure of the document to be maintained regardless of the platform but also prevents the possibility of editing the document.

In addition to the creation of documents using word processing programmes, many documents are created as a result of digitization. Digitization of professional journals to preserve collections created before the digital age or make them accessible is usually the main source of digital texts apart from

¹ **Binary file**, a file that contains data in a format intended to be read by specific computer programmes, with a substantial portion of the information encoded as other than text. Binary files generally cannot be interpreted without knowledge of the file format, except possibly to some extent.

² A **text file** is a file that contains only text. The text may be saved in different character encodings (such as UTF-8 or ISO 8859) and often has the file extension .txt. What distinguishes a text file from other file formats is that it is intended to be readable and understandable as such (provided you use the correct character encoding), without using any specific programme. The file should not contain check characters other than those with an established meaning, such as line breaks.

³ **Proprietary** software is software that has restrictions (usually set by the owner) on using, modifying, or copying it.

⁴ <https://www.ecma-international.org/publications-and-standards/standards/ecma-376/>. Accessed 16 August 2022.

⁵ <http://xml.openoffice.org/xml/>. Accessed 16 August 2022.

⁶ <https://www.adobe.com/acrobat/about-adobe-pdf.html>. Accessed 16 August 2022.

those created with a word processing programme.⁷ This process normally starts with a digitized image of the paper document, which is then processed using a machine-reading/optical character recognition (OCR) programme to turn the image into 'real' (editable/searchable, etc.) text. The final text, which may include images and tables, etc., is normally stored in PDF format although the use of an XML-based format is common, especially when the text is published online.

Following a desire to standardize formats for word processing programmes, the focus was on XML-based open formats (such as .docx and .odt) so that different software systems could use each other's formats. Similar problems existed for the PDF format (incompatibility), so the open standard of PDF/A⁸ was developed to solve these problems and enable archiving (see page 10 on PDF/A-1). Apart from the most common word processing formats and PDF, text can exist in a large number of variants of plain text or marked-up formats such as SGML, HTML, and XML.

⁷ For information on digitizing articles, reports, etc., from analogue format, there is a simple guide from the Archaeology Data Service in York, UK: <https://archaeologydataservice.ac.uk/advice/scanningGuide.xhtml>. Accessed 18 August 2022.

⁸ <https://www.iso.org/standard/50655.html>. Accessed 16 August 2022.

2. Things to consider when creating texts and documents

2.1 General considerations

Various text files and documents are normally part of the set of files created in a project and are often also used in the documentation of the project planning and at various stages of the project. Although little planning is needed to manage this type of file, there are a couple of things to keep in mind:

- It is important for material embedded in text documents, such as images and spreadsheets, to be stored separately with the document. This not only ensures that the files are stored in their original format or in an archive-friendly format, but also allows them to maintain their original resolution and size.
- Avoid including links and dynamic content⁹ as links, etc. may become invalid over time.
- If the intention is to save the final version of the document as PDF, make sure that the original format (Word .docx, OpenOffice .odt, etc.) is also saved separately.
- If documents are saved as PDF, make sure they are saved as PDF/A.¹⁰ The file should also not be user-protected, should not include javascript/video/sound, etc., and fonts and images should be correctly embedded. Also make sure that the file is properly marked-up.¹¹
- Ensure that non-proprietary fonts and UTF-8 encoding are used. This is because the appearance of the document may change if it is opened by someone who does not have access to the specific font. UTF-8 encoding should be used because it contains most characters (letters and numbers, etc.).

In general, it is important to ensure that all information in the document is complete and understandable. Sources should be cited, etc.

Avoid embedded material

One of the most common types of embedded material in text documents is images. For some types of documents (for example Microsoft Word and PDF), more complex material (for example spreadsheets and videos) can be embedded and in formats that should be kept separately with the text document. The recommendation is therefore that, in addition to embedding the material in the text documents, these files should be stored and preserved separately (in accordance with the respective file formats) to ensure that the files retain their original quality (for example image resolution).

⁹ Creating **dynamic content** may, for example, involve linking the contents of an Excel file to a table in a Word document. Changing the contents of the Excel file changes the contents of the table. Unlike an embedded object, where you copy the values into the table, the information for a dynamic content object is stored in a separate file. The problem with this is that moving the files breaks the link between the Excel file and the table in the Word document.

¹⁰ To save as PDF/A-1 from Microsoft Word, select Save as type, select PDF from the list, click the Options button and for PDF options select ISO 19005-1 compliant (PDF/A).

¹¹ **Markup** is the process of adding instructions or additional information to text or other files to facilitate automatic handling and information retrieval. Instructions usually concern the graphic appearance (style) and layout of the text. Additional information is usually keywords (metadata) that are entered to facilitate automatic searches.

3. Archiving texts and documents

3.1 Deciding which files to archive

As mentioned above, file format changes for documents do not occur to any great extent during the creation process. An exception is PDF files, which are often created at the end of this process to make the material available/disseminate it. However, it is recommended that the original file be saved in parallel with the PDF file. Embedded material such as images should also be stored separately with the text document to ensure that they are best preserved according to the file type.

Make sure that what is to be preserved is indeed the final version of the document. To make it easier, it is a good idea to keep track of all the different working versions and to delete notes/comments, etc. for the final version.

3.2 Deciding how to archive

When deciding which file format to use for archiving documents, it is a good idea to choose a format that preserves important features of the document, is commonly used and, if necessary, can be migrated¹² by different applications.

Significant properties

Significant properties, i.e., the basic elements of texts and documents to be preserved and maintained, are described below:

- The words and word order in the document.
- The hierarchical structure of the document (for example different heading levels).
- Formatting within the document (for example bold, italics).
- The page numbering of the document. This is important whether the document is a report or a thesis, published or not. If a user wants to cite and refer to the document, the page reference must be correct. It is important to be extra observant if the document is migrated several times.
- Embedded material, such as images and data tables. Ensure they are kept separate.

There are also properties that are not always seen as important to preserve. These include font type and size (provided it does not affect formatting and pagination) and the Track Changes feature.

However, significant properties of a document may change depending on the document to be preserved. Regardless, when reviewing a document for archiving, it is necessary to assess which of the above elements must be preserved.

File format for archiving

In terms of file formats for long-term storage and archiving, there is now a general recommendation to use standardized XML formats such as Microsoft's OOXML (.docx) and OpenOffice ODF (.odt). A JISC technical report, *XML-based Office Document Standards* (Ditch 2007),¹³ reviews and compares the different specifications. The main advantage of both formats is that they are internationally recognized open standards and text-based (unlike binary files, which are only machine-readable) and

¹² **Migration** may involve transfer between different media and transfer between different file formats.

¹³

<http://www.webarchive.org.uk/wayback/archive/20140615220449/http://www.jisc.ac.uk/media/documents/techwatch/tsw0702pdf.pdf>. Accessed 17 August 2022.

thus also intended for human reading. Both formats are mutually accepted as well as accepted by several third-party solutions, such as Google Docs. The formats are similar in that they use a zipped archive format in which the different parts of the documents are stored separately to form one file.

ODF makes better use of open and existing standards such as SVG (Scalable Vector Graphics). The available documentation of ODF is significantly shorter, and more complete, than OOXML, which may mean that third-party support for the format will spread more rapidly. Microsoft's OOXML, on the other hand, has better support for earlier versions of MS Word, as backward compatibility was one of the aims of the development of the standard. When converting from MS Word to ODF, the format is not completely correct. Among other things, conversions of graphic elements do not work fully on account of incompatibility between formats.

As a supplement to these XML-based formats, PDF/A may be a potential format for archiving, but for documents that otherwise only exist in PDF format. Although PDF/A is a binary format, it is an open standard and the software to read the files is free and easy to find, including through increased third-party support. Since extraction or migration from PDF documents to other formats is problematic, PDF/A offers a good, accurate way to preserve existing PDF material in a known open standard format, even if it is binary. In addition, there are several third-party programmes for importing text from PDF even if the text is saved as an image. The technology, OCR – Optical Character Recognition, converts digital images of text into letters and numbers. The quality of the import may vary, especially if the text was originally placed over background images, so some editing may be required.

There are also a few other more general problem areas that apply to text documents and archiving them. The first, which also applies to many other file formats, is the ongoing changes to the file formats used in word processing software. The second problem concerns file formats for which the software no longer exists. Another problem arises when the development and improvement of file formats used by existing word processing programmes results in incompatibility between older file versions and the current version of the software.

3.3 Metadata and documentation

Although many documents and texts are self-explanatory, some metadata should be documented for each individual document or group of documents for the sake of searchability and history. There are many different metadata standards that have been developed in different subject areas, for example MARC¹⁴ in the library sector. The type of metadata needed for searches and provenance checks, as opposed to technical metadata about the file documented in formats such as textMD,¹⁵ is described below and occupies only a minimum of metadata records which also exist in most metadata standards. These include the 15 elements contained in the Dublin Core Metadata Element Set.¹⁶ For documents, it is common for some metadata to overlap with metadata for the project (for example for reports describing the project).

The table below shows the information that should be included for each document/group of documents:

¹⁴ <http://www.loc.gov/marc/>. Accessed 17 August 2022.

¹⁵ <http://www.loc.gov/standards/textMD/>. Accessed 17 August 2022.

¹⁶ <https://www.dublincore.org/specifications/dublin-core/dcmi-terms/>. Accessed 17 August 2022.

Element	Description
Title	Title of the document.
Abstract	Brief description/summary of the document.
Publication date	Year of publication.
Published	Full publication reference. Series/journal, issue, edition, start and end page or number of pages, etc., should be noted.
Publisher or equivalent	Details of the publisher or equivalent (name, location, etc.).
ISBN	ISBN ¹⁷ (where available).
DOI	Digital Object Identifier (DOI) ¹⁸ (where available).
URL	URL (where available).
Related material	Information on related material such as files, databases, and other material.
Language	Specify the language in which the document is written.
Author	Names of authors.
Contributor	Names of contributors.
Email	Email address of the author/contact.

Apart from the elements listed above, there are additional elements (see below) that should be recorded at document level. Some of these elements may also apply to a group of documents created at project level. Metadata may therefore already exist at that level, but it is still recommended that these elements are recorded for each document:

¹⁷ <http://www.isbn-international.org/>. Accessed 17 August 2022.

¹⁸ <http://www.doi.org/>. Accessed 17 August 2022.

Element	Description
Project name	Name of the associated project, and any sub-projects.
Subject	Specify the subject of the project. Map to DC subject or other controlled vocabulary.
Investigation type	Specify how the investigation was conducted, for example by questionnaire, observation, experiment, field survey, etc.
Geographical coverage	Parish, municipality, county, region, country, etc.
Time period of the investigation	Keywords for time periods, start/end date of investigation, dating of material (for example C14).

4 File formats

The table below shows some common file formats and associated applications as well as some information about them and their potential for use in archiving documents. Some of these formats may appear obsolete, but they are described since it may be necessary to handle older file formats.

Adobe PDF	
File format/ extension	PDF/.pdf
Format	PDF (Portable Document Format), created by Adobe, is essentially an open standard for transfer between different systems. The format has evolved several times since it was first created and there are several versions from 1.0 to 1.7.
Description	PDF is a format intended to be used across different platforms. Although it is a proprietary, binary format, it works very well for disseminating material as it is designed to maintain the format of the original document. In addition to displaying plain text, the format can contain a wide variety of embedded file formats or linked media including raster and vector graphics, JavaScript, and 3D data. PDF files can be protected so that it is not possible to edit or print the document.
Recommendations	<p>Although PDF is an open standard, the format is proprietary and binary, making it unsuitable for deposition or archiving.¹⁹ In most cases, PDF files are created from other file formats (for example docx). It is better to deposit and archive the original format, but where this is not possible, PDF/A is preferred. You must then check that all embedded material is included correctly.</p> <p>If the format is to be used, ensure that features such as text search, embedded fonts, lossless compression, high-resolution images, standardized colour information²⁰ and content tagging are included in the document.²¹</p>
File format/ extension	PDF/A / .pdf/a-1, .pdf/a-2, .pdf/a-3
Format	The format is based on the PDF format and was created by Adobe. PDF/A was developed as an open standard to be used for archiving.

¹⁹ For further reading on the problems of using PDF for archiving, read the report *Preserving the Data Explosion: Using PDF* (Fanning 2008) <https://www.dpconline.org/docs/technology-watch-reports/86-preserving-the-data-explosion-using-pdf/>, written by DigitalPreservationCoalition (DPC). See also <https://www.loc.gov/preservation/digital/formats/fdd/fdd000030.shtml>. Accessed 17 August 2022.

²⁰ Device-independent specifications of color space, i.e., standard terms for describing colours such as brightness, hue, saturation, and intensity.

²¹ <https://www.loc.gov/preservation/resources/rfs/text.html>. Accessed 17 August 2022.

Description	<p>PDF/A-1 is based on PDF version 1.4 and is designed to serve as a reliable open standard format for archiving. The format is an ISO standard (ISO 19005-1:2005/COR 2:2011²²).</p> <p>PDF/A-2 is based on PDF version 1.7 and is also an ISO standard (ISO 32000-1:2008.²³ Note that ISO 19005-2:2011²⁴ describes the use of ISO 3200-1).</p> <p>PDF/A-3 is also based on version 1.7 but has a different ISO standard (ISO 32000-2:2020²⁵).</p> <p>The difference between PDF and PDF/A is that the latter does not include certain features from PDF to ensure that it can be used indefinitely. PDF/A-2, on the other hand, allows the use of JPEG2000 and makes it possible to enclose/attach PDF/A files. PDF/A-3 allows you to attach any type of file.</p> <p>There are then three different conformance levels for the PDF/A documents: b, u, and a. The basic level 'b' conforms to the basic requirements of the PDF specification, which include the ability to visually reproduce a document over time. The 'u' level (which only applies to PDF/A-2 and -3) means that the 'b' level must be met as well as the Unicode requirements of the PDF/A specifications, while the 'a' level must meet all the requirements of the PDF/A specifications, which means, among other things, that it must include markup and searchable document structure.</p>
Recommendations	<p>PDF/A-1 has been accepted as a viable format for archiving (for example by the Library of Congress²⁶) and has been reviewed and assessed by DigitalPreservationCoalition (DPC).²⁷</p> <p>It is recommended that files with another provenance (for example .doc or .odt) be kept and saved in parallel with the PDF/A file. In order to be considered a secure format for archiving, certain elements (fonts and colours) must be specified or embedded in the file, while other elements (javascript, audio, video, encrypted material) must not be present.</p>

²² <https://www.iso.org/standard/60603.html>. Accessed 17 August 2022.

²³ <https://www.iso.org/standard/51502.html>. Accessed 17 August 2022.

²⁴ <https://www.iso.org/standard/50655.html>. Accessed 17 August 2022.

²⁵ <https://www.iso.org/standard/75839.html>. Accessed 17 August 2022.

²⁶ <http://www.digitalpreservation.gov/formats/fdd/fdd000125.shtml>. Accessed 17 August 2022.

²⁷ *Preserving the Data Explosion: Using PDF* (Fanning 2008) <https://www.dpconline.org/docs/technology-watch-reports/86-preserving-the-data-explosion-using-pdf/>, written by DigitalPreservationCoalition (DPC). Accessed 17 August 2022.

Microsoft Word	
File format/ extension	DOC/.doc
Format	A proprietary binary format for Microsoft Word.
Description	A popular file format and the default format for all versions of MS Word from 1.0-6.0, 95 and 97-2003. The files can also be read by OpenOffice and converted to .pdf. Although backward compatibility has existed between different versions of Word, with service pack 3 support for version 2.0 and earlier has been removed. After 2008, the specifications for a number of Microsoft binary file formats were made available on the Microsoft website. ²⁸
Recommendations	Although the file format is not suitable as an archive format or as a format for the dissemination of material, it is so common that there is no obstacle to its use.
File format/ extension	DOCX/.docx
Format	Part of the Office Open XML (OOXML) format created by Microsoft. An ECMA (ECMA-376 ²⁹) and ISO (ISO/IEC 29500-1:2016 ³⁰) standard.
Description	The file format was introduced with Office 2007. Microsoft chose to develop its own standard (OOXML) instead of using the ODF standard (ISO/IEC 26300:2006, see ODT below) to create better conditions for backward compatibility with earlier versions of the MS Word file format. The format consists of readable XML files which are packed with other files (images, etc.) in a zipped archive. ³¹
Recommendations	Suitable for deposition, dissemination, and archiving, but embedded material should be stored separately. Since the file format is built as a zipped archive, it should be stored in uncompressed format. ³²

²⁸ https://docs.microsoft.com/en-us/openspecs/office_file_formats/ms-offfflp/6ae2fd93-51fc-4e75-a54a-1b175c627b51. Accessed 18 August 2022.

²⁹ <http://www.ecma-international.org/publications/standards/Ecma-376.htm>. Accessed 18 August 2022.

³⁰ <https://www.iso.org/standard/71691.html>. Accessed 18 August 2022.

³¹ To unpack the file: change the file suffix from .doc to .zip. Unzip the file. The result is an xml file and a number of folders that contain more folders and xml files. The text itself is in the 'word' folder, in the document 'document.xml'. <http://www.digitalpreservation.gov/formats/fdd/fdd000397.shtml>. Accessed 17 August 2022

³² MS Office 2007/2010 does not support the ODF 1.2 format, which means that you may get an error message if you try to open an .odf document in MS Word, for example. Convert ODF to 1.0/1.1 and it should work. Office 2013 supports the ODT 1.2 format.

Unformatted text	
File format/ extension	TXT/.txt and unformatted text
Format	A simple unformatted text file. Different from formatted text where style sheets are included, and binary files where part of the information is encoded. Unformatted text files are the basis for marked-up texts (see below).
Description	Simple unformatted text files are the simplest format for text information and are compatible across a wide range of platforms and software. Since the format barely supports any formatting, it should only be used for the simplest form of document. Some form of character encoding (ASCII or UNICODE) should be specified for all variants.
Recommendations	Works well for archiving and disseminating text files but only if they are simple files and if character encoding is specified.

OpenDocument Text	
File format/ extension	ODT/.odt
Format	Open Document Text ³³ is one of the formats developed as part of the OpenDocument Format, ³⁴ an ISO standard (ISO/IEC 26300:2006 ³⁵) for XML-based document formats. The format is supported and used by several office applications.
Description	Like .docx, .odt consists mainly of XML files, intended for human reading, which are packed with other files (images, etc.) in a ZIP file.
Recommendations	As .odt is an open XML format, it works well for both deposition and archiving. For archiving, the files should be uncompressed. If the files contain images or other embedded material, this material should be stored separately in a suitable format for archiving.

³³ <http://www.digitalpreservation.gov/formats/fdd/fdd000427.shtml> and

<http://www.digitalpreservation.gov/formats/fdd/fdd000428.shtml>. Accessed 18 August 2022.

³⁴ <http://www.digitalpreservation.gov/formats/fdd/fdd000247.shtml>. Accessed 17 August 2022.

³⁵ http://www.iso.org/iso/catalogue_detail.htm?csnumber=43485. Accessed 17 August 2022. There are three parallel ISO standards that are also valid: ISO/IEC 26300-1:2015/-2:2015/-3:2015, which define different parts of the standard.

OpenOffice/StarOffice	
File format/ extension	Sxw/.sxw
Format	The XML format used by applications such as OpenOffice/StarOffice from version 1.0 to 2.0. Later replaced by OpenDocument Format (.odt).
Description	Although the format has been replaced by .odt, it is structurally similar (zipped xml files) and can be read by OpenOffice.org 2.0.
Recommendations	Works for archiving but it is better to update to .odt where possible.

Rich Text Format	
File format/ extension	RTF/.rtf
Format	RTF (Rich Text Format) is a marked-up text format developed by Microsoft.
Description	Although the format is readable plain text, and therefore suitable for both archiving and use, there are some compatibility issues regarding formatting (for example text boxes and tables) when the files are opened in some word processing programmes. An .rtf file is normally larger than the corresponding .doc, .pdf or .odt file.
Recommendations	Although the format works for deposition and storage, there are better and newer formats such as .docx and .odt which are smaller and have better compatibility. PDF/A is also preferable.

WordPerfect	
File format/ extension	WPD/.wpd
Format	A binary, proprietary file format developed for WordPerfect.
Description	Use of the once-popular WordPerfect has declined since its introduction in the early 1980s, on account of Microsoft Word. Although .wpd (.wp and .wp5 for earlier versions) is the default format, later versions of the programme support

	the import and export of various file formats. The file format can be read by Microsoft Office Word and OpenOffice.
Recommendations	Not recommended as a format for disseminating material or storage/archiving. Although the latest version of WordPerfect can import and export ODF and OOXML files, it is better to use the later XML-based open file format options.

Format for marking up text

Although markup languages are not usually used to create reports and similar documents (HTML is usually used for web pages and XML for data exchange), some information about some of the formats is provided below.

HTML/XHTML	
File format/extension	HTML/.htm/.html, XHTML/.xhtml/.xht
Format	Hypertext Markup Language (HTML) is a markup language written as unformatted text and developed as a subset of SGML.
Description	HTML is a markup language mainly used for websites. Apart from the unformatted text content of the HTML file (including either a style sheet or a link to a style sheet), web pages often consist of various types of linked media (images, video, audio, documents, etc.).
Recommendations	The format works for archiving and dissemination, but character encoding must be specified. For versions earlier than HTML5, the document must also follow and specify a valid DTD (document type definition). If CSS (style sheets) are used, they should either be specified in the document or attached separately. Images and other media should be treated as individual objects according to the file type.

SGML	
File format/extension	SGML/.sgml
Format	Standardized Generalized Markup Language (SGML). A certified ISO standard (ISO 8879:1986 SGML ³⁶) for markup languages.

³⁶ http://www.iso.org/iso/catalogue_detail.htm?csnumber=16387. Accessed 22 August 2022.

Description	SGML is a metalanguage used to define other markup languages such as HTML and XML.
Recommendations	The format works for archiving and dissemination, but documents must follow a defined scheme according to the ISO 8879 standard.

XML	
File format/ extension	XML/.xml
Format	Extensible Markup Language (XML) is an open standard based on unformatted text and developed by the World Wide Web Consortium (W3C).
Description	XML was developed as a subset of SGML ³⁷ and is mainly used for the web and for exchanging data between different systems (for example databases).
Recommendations	Works for archiving and dissemination, but documents must follow and specify a valid XSD/DTD and character encoding.

³⁷ <http://www.digitalpreservation.gov/formats/fdd/fdd000075.shtml>. Accessed 22 August 2022.