

# Preserving Email

## Data Types Series

Artefactual Systems and the Digital  
Preservation Coalition



## DPC Technology Watch Guidance Note

July 2021



Digital Preservation Coalition

## The Data Type Guidance Note Series

Each Guidance Note in the Data Types series is designed to provide a primer on the current state of community knowledge about data types commonly encountered by those seeking to preserve digital holdings. Digital preservation is about keeping information findable, usable, and trustworthy over the long-term. The best approach for any repository will vary according to the scope and content of its holdings, available resources, and the expectations of its funders and users. There are however, broadly applicable good practices that have been established as a result of many years of research, practical implementation, and consensus building. These are presented here as a starting point, along with additional resources for further exploration.

This series of Data Type Guidance Notes has been authored by staff at Artefactual Systems in collaboration with the Digital Preservation Coalition. These notes have been developed in conjunction with the UK Nuclear Decommissioning Authority.

Digital preservation is an evolving field and continues to change and develop in response to external drivers and fresh challenges. New formats, standards, and examples of good practice will emerge over time and the information contained within this report will need to be updated. We welcome comments and feedback to: [info@dpconline.org](mailto:info@dpconline.org).

## 1 Overview

Email is a system of asynchronous exchange of messages between computers. An email message is a structured text document which typically includes a header and body, may contain embedded images or other objects, and may include attachments. Email messages are typically structured using the Internet Message Format (IMF) ([RFC 5322](#), 2008) and Multipurpose Internet Mail Extensions (MIME) ([RFC 2405](#), 1996) standards.

A single email message may exist in many instances and email messages can have complex relationships with one another, since they are often part of threads (groups of messages that have been replied to or forwarded to other recipients). Email headers include information that is visible to the user, such as sender, recipient(s), and subject lines, and information typically hidden from the user such as routing and authentication information and date/time stamps. Once an email is in a recipient's mailbox, additional metadata can be generated automatically, such as flags indicating read or unread status, level of importance, or a requirement to follow up. Users can add labels to certain groups of emails and can organize the emails or email threads into folders. These factors, as well as the ability to retain copies of sent messages and add attachments to email, make email accounts de facto recordkeeping systems. For many organizations, email accounts may be one of the most important sources of documentary evidence of activities and transactions, and therefore have enduring archival value ([InterPARES](#), 2009, p.5).

## 2 Preservation Challenges

Email is a complex data type consisting of multiple conversations involving many different people. It can include attachments (which may be of any other data type or file format), it can be large in size and it can be challenging to capture it effectively. These challenges are described in more detail below:

### 2.1 Attachments

The introduction of MIME in the 1990s standardized the way digital files could be attached to email messages. Although in theory files of any size and format can be attached to messages, some email programs apply limits to the size and number of files that can be transmitted, and exclude certain types of files such as executable code ([Tech Terms](#), 2015). Email attachments may need to be preserved along with the email messages to which they were attached. Preserving an email account may therefore necessitate preserving potentially large volumes of other digital files in a variety of formats. Both email messages and their attachments must be preserved in such a way as to maintain the relationships between them.

### 2.2 Extraction and transformation

In the Digital Preservation Coalition's Technology Watch report on email, Chris Prom writes that 'The first, and in some ways, most pressing problem for those wanting to preserve mail is how to acquire it in the first place' ([Prom](#), 2019, p.19). Capture strategies may vary according to how organizations, or even particular individuals within organizations, manage and store email. Most email systems have export options built into them. However, in some cases archivists may not have direct access to an email account, and in such situations it is necessary to use specialized tools (such as [OfflineIMAP](#)) to retrieve the emails. Any effort to preserve email requires some kind of transformation when extracting it from its originating system; there is no 'native' preservation format ([EA-PDF working group](#), 2021, p.7).

### 2.3 Legal issues

The body of an email is unstructured text which may contain personally identifiable information (PII), defined broadly as ‘any information relating to an identified or identifiable natural person’ ([GDPR, Definitions](#), 2016). Email attachments may also contain PII or other sensitive personal data. There may also be copyright complications, as the email author is not the email receiver. Reuse and retention may be affected by varying legal frameworks in different jurisdictions. For example, reuse may be governed by the General Data Protection Regulation ([GDPR](#)), the German Bundesdatenschutzgesetz ([BDSG](#)), the UK Data Protection Act ([DPA](#)), or the U.S. Federal Trade Commission Act ([FTCA](#)). Email accounts can hold significant volumes of PII and sensitive data because of the nature of the organization (such as banks, medical offices, or certain types of government agencies) or position of the person (for example, human resources personnel, executives, or union officials). PII and sensitive personal data can also appear accidentally because the user contravenes organizational policies governing email use, because there are no such policies, or because the user employs an email account for both business and personal purposes. Private individuals donating email accounts to archives may have used their email to correspond with any number of other private individuals who are unaware that their communications are being preserved.

### 2.4 Volume and volatility

Hundreds of billions of emails are sent and received worldwide each day ([Radicati](#), 2021; [Statista](#), 2021). Email use is prolific across many organizations, and multiple individuals within the same organization may be copied on email threads, with the result that numerous duplicates of an email (and its attachments) can reside in multiple email accounts. New threads may also split away from original threads, with recipients added and removed over the life of the thread. Depending on institutional records management policies, users may be free to delete emails and email threads, although some may delete messages from their inboxes without deleting any corresponding emails or threads from ‘sent items’ boxes. ‘Deleted’ messages may not actually be deleted at all, although they appear to be from the user’s perspective, and can be retrieved from a ‘trash’ or ‘recycled’ folder by the user during a certain period. This kind of volatility can make appraisal and acquisition of email exceptionally challenging. As a recent email task force report asks, ‘[W]hat is considered the record copy? The email that was composed and sent? The one that was received and stored in the sent mail folder of the sender’s client application? The copy stored on the recipient’s email server system? A downloaded copy? A copy stored locally in a PST file? A copy on an iPhone?’ ([Council on Library and Information Resources \(CLIR\)](#), 2018, p.30).

## 3 Typical Email formats

There is no single perfect format for the preservation and future use of email. Decisions made on file formats should be dependent on the features and functionality to be preserved and the future use cases to be supported. Note that the table below describes typical transfer formats for email archives that may also be suitable preservation and/or access formats, but should not be seen as an exhaustive list of formats. It is recommended that careful research and analysis is carried out before extracting emails or migrating them to a new format.

File format	Extensions	Brief summary
-------------	------------	---------------

Internet Message Format	.eml	<p>EML is a text-based format designed to conform to the original RFC email specification (<a href="#">RFC822</a>, 1982) and its successors. EML files are typically single email messages with attachments either embedded as encoded base64 content in the file or referenced from the file. The messages (but typically not the attachments) can be opened and read in any text editor. EML can be exported from and rendered by many email programs, including Microsoft Outlook, Zimbra, Eudora, Mozilla Thunderbird, Apple Mail, and Gmail. There is also good tool support for converting email exports in other formats to EML. A useful list of tools used in capturing, converting, and preserving emails as EML and other formats is provided by the Task Force on Technical Approaches to Email Archives (<a href="#">Email Archiving Tools</a>, 2018).</p> <p>EML is based on RFC standards, but there is no specification for the format itself, and the ability to manage threads, attachments, and other features relating to the internal structure of email accounts is limited (<a href="#">EA-PDF Working Group</a>, p.26). Because EML is text-based and ubiquitous it may well be renderable for the foreseeable future. Some major email preservation projects, such as Harvard University’s <a href="#">EASi</a> platform, migrate other formats to EML for preservation.</p> <p>EML messages can be extracted from email accounts and stored as Maildir, a directory format structured to represent the delivery status of the messages. Maildir can be exported from and loaded into a number of email applications; however, it lacks detailed specifications and has some file system compatibility issues which may require remediation before the files can be read by certain email programs (Wikipedia, <a href="#">Maildir</a>, 2020). The open-source <a href="#">OfflineIMAP</a> tool can be used to extract email accounts remotely and store them as MailDir.</p> <p>Library and Archives Canada and the US National Archives and Records Administration both consider EML files to be preferred transfer formats for email (<a href="#">LAC</a>, 2015; <a href="#">NARA</a>, 2020).</p>
MBOX	.mbox	<p>Like EML, MBOX is a text-based format designed to implement RFC email standards. Unlike EML, however, a single MBOX file can contain an aggregation of messages, typically the contents of a folder. MBOX is well supported by numerous email clients; a user can export and store an account as a series of MBOX files, each representing a folder of messages with attachments embedded in them as base64 encoded text. These files can be imported into other email clients or read in <a href="#">EPADD</a>, an open-source tool designed by Stanford University to support ‘appraisal, processing, preservation, discovery, and delivery of historical email archives’ (<a href="#">Stanford Libraries</a>, 2021). University of North Carolina’s <a href="#">RATOM</a> (Review, Appraisal, and Triage of Mail) project has also developed tools for working with MBOX files.</p>

		<p>A single MBOX file can encapsulate an entire email folder, which results in more resilient relationships between email messages and threads and makes it easy to load the folder into an email client or a tool like EPADD. However, due to the large volume of emails that can reside in a single folder, MBOX files can often become excessively large and unwieldy; it also means that, in some cases, the corruption of a single message can prevent the entire file from loading or opening (<a href="#">EA-PDF Working Group</a>, p. 26; <a href="#">DPC and Prom</a>, 2019, p.17). MBOX as a preservation format also suffers from the fact that it is a family of formats with four variants (MBOXO, MBOXRD, MBOXCL and MBOXCL2) that are not entirely compatible with one another (<a href="#">Library of Congress (LC)</a>, 2016). Despite these drawbacks, MBOX files show considerable promise as both a preservation and access format, given the format’s ubiquity, tool support, ability to represent entire email folders, and underlying text-based structure.</p> <p>Library and Archives Canada and the United States Archives and Records Administration consider MBOX to be a preferred email transfer format (<a href="#">LAC</a>, 2015; <a href="#">NARA</a>, 2020).</p>
Microsoft Outlook Personal Folders	.pst .olm	<p>PST is a proprietary format developed by Microsoft to support email download, storage, and offline use. It has two versions, ANSI (the earlier of the two) and Unicode, which are ‘differentiated primarily by software implementation versions, character sets, maximum file size constraints and bit values’ (<a href="#">LC</a>, 2013). A PST file can be an entire email account or a portion of one, and can include not only messages but also contacts, calendar items, tasks, and other products of Outlook and related tools.</p> <p>Microsoft maintains a freely available specification for PST (<a href="#">Microsoft, [MS-PST]</a>, 2020). The ubiquity of Microsoft Outlook and related tools makes PST a relatively common format. Library and Archives Canada lists PST as an acceptable ‘transfer format’, with the preferred format being MBOX and EML, while the National Archives and Records Administration of the United States lists all three as preferred formats (<a href="#">LAC</a>, 2015; <a href="#">NARA</a>, 2020). The <a href="#">RATOM</a> project’s toolset works with PST files in addition to MBOX.</p> <p>However, be aware that in Mac implementations of Microsoft Outlook the export format is called Outlook for Mac (OLM). OLM does not have an openly documented specification like PST and is subject to unidentified changes (<a href="#">Microsoft</a>, 2015). There are a number of free and paid OLM to PST converters such as <a href="#">Softaken</a> and <a href="#">Mailvita</a>, but testing and quality assurance is advised before selecting a tool.</p>
Portable Document Format	.pdf	<p>When single email messages or threads are converted to PDF or PDF/A, they lose much of their metadata and attachments can become inaccessible (<a href="#">CLIR</a>, 2018, pp.82-83). However, a dedicated task force on email preservation recently proposed a specification ‘to set the stage for software developers to create email capture and representation systems leveraging PDF to represent the core metadata, content, attributes, and</p>

		<p>context’ of email accounts’ (<a href="#">EA-PDF Working Group</a>, 2021, p.7). The authors argue that MBOX and EML ‘are actually somewhat amorphous families of formats defined more by client implementations than by authoritative specifications’ and that ‘PDF provides a means to represent these implementations in a normalized packaging model, regardless of the underlying source.’ (<a href="#">EA-PDF Working Group</a>, 2021, p.8).</p> <p>PDF versions of email messages are highly accessible due to widespread software support for the format, and the efforts of the EA-PDF Working Group demonstrate how PDF could be enhanced to serve as both a preservation and access format for entire email accounts rather than individual messages.</p>
Extensible Markup Language	.xml	<p>XML has been explored as a preservation format by a number of digital preservation projects, including <a href="#">PeDALS</a>, <a href="#">CERP</a>, and <a href="#">DArcMail</a>. DArcMail is the most recent of these, and the successor to CERP. DArcMail, which stands for Digital Archives for Email, is run by the Smithsonian Institution Archives, and is the name of both the project and a python-based open-source tool available on the project’s homepage. DArcMail converts MBOX files to XML and includes a graphical user interface that ‘supports appraisal, processing, and providing access to email accounts or large groups of related email messages.’ (<a href="#">Smithsonian Institution Archives</a>, 2017, p. 3). The XML file can represent an entire email account, retaining internal folder structure; header and other metadata; and messages and their attachments. If the Smithsonian Institution Archives receives email in non-MBOX formats, it converts them to MBOX and then uses DArcMail to convert the MBOX files to XML (<a href="#">Smithsonian Institution Archives</a>, 2017, pp.4-5).</p>
Microsoft Outlook Item	.msg	<p>An MSG file is a single email or other Microsoft Outlook Item which can be downloaded by Outlook end users. Unlike EML or MBOX, MSG files cannot be read in text editors; however, Microsoft maintains a published specification for the format (<a href="#">Microsoft, [MS-OXCMSG]</a>, 2020). MSG is not likely to be used for transferring entire email accounts to preserving institutions, but files in that format may appear in transfers of diverse bodies of digital objects from network drives, personal computers, or electronic recordkeeping systems. There has been little discussion in the digital preservation community of MSG as a preservation format, although the open specification and ongoing Microsoft support for MSG suggest that it will be readable for the foreseeable future. When capturing email accounts, archivists should avoid MSG in favour of MBOX, EML, or PST.</p>

## 4 Tips for creators

Creators working in government, business or other controlled environments should be aware of their organizations’ email handling and records management policies. Adherence to guidelines and

requirements for file formats, versioning, retention, and metadata creation will help ensure that emails can be preserved in such a way as to retain their context and meaning over time. Managing email can be a significant challenge for creators outside these environments; however, there are steps that creators can take to ensure that email records of enduring archival value can be handled appropriately.

These tips provide useful guidance for email users in diverse environments:

- If email management and retention policies are lacking, creators should engage in direct communication with archivists or other personnel tasked with preserving email, in order to establish preservation priorities, transfer methodologies, and disposal policies.
- Private individuals considering donating email to a preserving institution should expect to provide information about how they managed their email over time.
- When writing emails, creators should avoid including personally identifying information about third parties unless it is necessary to conduct their business.
- Work and personal email accounts should be separated, and messages sent from a work email address should not be of a personal nature.
- Creators deleting email, either in accordance with organizational policies or as private individuals, should keep in mind that they may have to delete messages from both their inbox and sent items folders -- simply moving them into a “trash” folder is not sufficient.
- If creators are not sure whether or not certain emails should be deleted, they should consult with records managers and/or archivists for guidance.

## 5 Tips for archivists

### 5.1 General guidance

The following resources provide guidance on preserving and providing access to email:

- *The future of email archives* ([CLIR](#), 2018).
- *Preserving email Technology Watch Report* ([DPC and Prom](#), 2019).
- Projects such as EASi ([Harvard](#), 2017), Carcanet Press ([Baker et al](#), 2012), DArcMail ([Smithsonian Institution Archives](#), 2017) ePADD ([Stanford Libraries](#), 2021), and RATOM ([University of North Carolina](#), 2021) offer real-world examples and tools for email preservation.
- A number of software tools are available for working with Email data ([COPTR](#), 2021).

### 5.2 Acquisition and appraisal

- Aim to acquire entire email accounts as opposed to individual messages whenever possible, to maintain folder hierarchies and metadata such as read/unread status, flags, labels, and other metadata generated by email software.
- Obtain direct access to email accounts for acquiring and transferring them to archives, in order to make use of native export capabilities. If such access is not possible, tools like [OfflineIMAP](#) can extract the accounts remotely.
- Be prepared to acquire attachments in diverse file formats and develop preservation plans for them.
- Work closely with Records Managers and IT personnel to establish strategies and technical approaches to acquiring and preserving email.



### 5.3 Characterization

- Identify file formats with a tool such as DROID ([The National Archives](#), n.d.), FIDO ([Open Preservation Foundation](#), 2020), or Siegfried ([Lehane](#), 2020) that uses the PRONOM file format registry ([The National Archives](#), 2020).
- Use PDF validation tools such as JHOVE ([Open Preservation Foundation](#), 2020) and VeraPDF ([Open Preservation Foundation](#), 2020) to check conformance to published PDF specifications and metadata standards. Validation tools for other email formats are not available at this time.

### 5.4 Quality assurance

- If using conversion tools such as OLM to PST converters, make sure that results are quality assured against loss of content, metadata, or significant properties.
- Retain the original email format if possible, being aware that format advice and conversion tools change over time.

## 6 References

Archaeology Data Service [ADS] (2009) *Databases and spreadsheets: A guide to good practice*.

Available at:

[https://web.archive.org/web/20201215095638/https://guides.archaeologydataservice.ac.uk/g2gp/DbSht\\_Toc](https://web.archive.org/web/20201215095638/https://guides.archaeologydataservice.ac.uk/g2gp/DbSht_Toc)

Baker, F. Butler, P. and Green, B (2012) *Carcanet Press Email Preservation Project*. Available at:

<https://web.archive.org/web/20200522083351/https://www.escholar.manchester.ac.uk/uk-ac-man-scw:165096>

COPTR (2021) *Email*. Available at:

<https://web.archive.org/web/20210706065108/https://coptr.digipres.org/index.php/Email>

Dalgleish, D. (2020) Contextures Blog. Available at:

<https://web.archive.org/web/20200623025727/https://contexturesblog.com/archives/2020/01/30/keep-notes-on-excel-formulas-and-macros/>

Dingledy, F. W. and Matamoros, A. B. (2016) *What is Digital Rights Management?* Available at:

<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*. Available at:

<https://web.archive.org/web/20130903083642/http://en.nationaalarchief.nl/sites/default/files/docs/kennisbank/volatility-permanence-spreadsh-en.pdf>

Harvard University (2017) *Submitting Content to EASi*. Available at:

<https://web.archive.org/web/20210708085021/https://wiki.harvard.edu/confluence/display/LibraryStaffDoc/3.+Submitting+Content+to+EASi>

ISO (2020a) *ISO/IEC 26300-1:2015: Information technology — Open Document Format for Office Applications (OpenDocument) v1.2 — Part 1: OpenDocument Schema*. Available at:

<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*. Available at:

<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*. Available at:

<https://web.archive.org/web/20201125031501/https://www.iso.org/standard/51502.html>

Klindt, M. (2017) *PDF/A considered harmful for digital preservation*. Available at:

<https://web.archive.org/web/20200917210028/https://ipres2017.jp/wp-content/uploads/15.pdf>

Lehane, R (2020) *Siegfried*. Available at:

<https://web.archive.org/web/20201028192837/https://github.com/richardlehane/siegfried>

Let's solve the format problem (2020) *Quattro Pro*. Available at:

[https://web.archive.org/web/20200714113418/http://fileformats.archiveteam.org/wiki/Quattro\\_Pro](https://web.archive.org/web/20200714113418/http://fileformats.archiveteam.org/wiki/Quattro_Pro)

Let's solve the format problem (2019a) *Documents*. Available at:

<https://web.archive.org/web/20200701000610/http://fileformats.archiveteam.org/wiki/Document>

Let's solve the format problem (2019b) *Lotus 1-2-3*. Available at:

[https://web.archive.org/web/20200702010518/http://fileformats.archiveteam.org/wiki/Lotus\\_1-2-3](https://web.archive.org/web/20200702010518/http://fileformats.archiveteam.org/wiki/Lotus_1-2-3)

Library of Congress (2021) *TSV, Tab-Separated Values*. Available at:

<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*. Available at:

<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, ISO 26300:2015*. Available at:

<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*. Available at:

<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*. Available at:

<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*. Available at:

<https://web.archive.org/web/2020111210012/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?* Available at:  
[https://web.archive.org/web/20201029193019/http://digital-archiving.blogspot.com/2017/04/how-can-we-preserve-google-documents\\_35.html](https://web.archive.org/web/20201029193019/http://digital-archiving.blogspot.com/2017/04/how-can-we-preserve-google-documents_35.html)

Morrissey, S. (2020) *Preserving Software: Motivations, Challenges and Approaches*. Available at:  
<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*. Available at:  
[https://web.archive.org/web/20201031113655/http://digitalpreservation.ncdcr.gov/file\\_formats\\_in-house\\_preservation.pdf](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)*. Available at:  
<https://web.archive.org/web/20200916134739/https://github.com/openpreserve/fido>

Open Preservation Foundation (2020) *JHOVE*. Available at:  
<https://web.archive.org/web/20201031215050/https://openpreservation.org/products/jhove/>

OpenPreservation Foundation (2020) *veraPDF*. Available at:  
<https://web.archive.org/web/20201031220541/https://openpreservation.org/products/verapdf/>

Smithsonian Institution Archives (2017) *Email Preservation – DarcMail*. Available at:  
<https://web.archive.org/web/20210417235150/https://siarchives.si.edu/what-we-do/digital-curation/email-preservation-darcmail>

Stanford Libraries (2021) *ePADD*. Available at:  
<https://web.archive.org/web/20210413140425/https://library.stanford.edu/projects/epadd>

Technopedia (2020) *Macro*. Available at:  
<https://web.archive.org/web/20210123172223/https://www.techopedia.com/definition/3833/macro>

The National Archives (n.d.) *Digital Object Record Identification (DROID)*. Available at:  
<https://web.archive.org/web/20201015033155/https://github.com/digital-preservation/droid>

The National Archives (2020) *The Technical Registry: PRONOM*. Available at:  
<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*. Available at:  
<https://web.archive.org/web/20200310033845/https://www.nationalarchives.gov.uk/documents/information-management/best-practice-guide-appraising-and-selecting.pdf>

University of North Carolina (2021) *RATOM*. Available at:  
<https://web.archive.org/web/20210301095930/https://ratom.web.unc.edu/>

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*. Available at:  
[https://web.archive.org/web/20191202082644/https://ipres2019.org/static/pdf/iPres2019\\_paper\\_48.pdf](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)*. Available at: <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*. Available at: <https://doi.org/10.5281/zenodo.3971833>

Young, P. (2021) *What's Up, (with Google) Docs? – The Challenge of Native Cloud Formats*. Available at: <https://web.archive.org/web/20210304124326/https://www.dpconline.org/blog/whats-up-with-google-docs>