File format recommendations by content type

Last updated June 2023

Table of Contents

Purpose

These recommendations are intended for depositors who need additional guidance on file formats. The recommendations below may be helpful for depositors submitting a research-creation thesis, or depositors whose research involves creating complex digital objects such as software applications or websites.

Following preservation best practices when preparing your research for deposit ensures that the files will remain accessible and usable in the long-term. The list of content types covered by these guidelines is not exhaustive, and other content types may require care to ensure future preservation and access.

Several factors make certain file formats preferable for others for long-term preservation. When possible, we recommend using formats that are open and non-proprietary well documented and widely used to mitigate the risk of future obsolescence. For general guidance on common file formats, please read the Spectrum Policies to see which file formats are recommended for preservation and access. For any further guidance, please contact lib-spectrum@concordia.ca.

Recommendations by content type

Software

Include the source code, any installation packages, and a README file. Written using a plain text file format, the README should be a note to future users who wish to use the files. It should include information such as scope, purpose, author(s), relevant dates, license for reuse, dependencies, field names/descriptions, and instructions for use. When possible, we recommend making your software

available under an open-source license to increase its longevity and usability. For guidance on open-source licenses, see: https://opensource.org/licenses/.

If the code is hosted on a source code version control system, such as GitHub or Bitbucket, download a ZIP file of the repository and include a link to the hosted version. See the screenshot below for how to do this on GitHub.
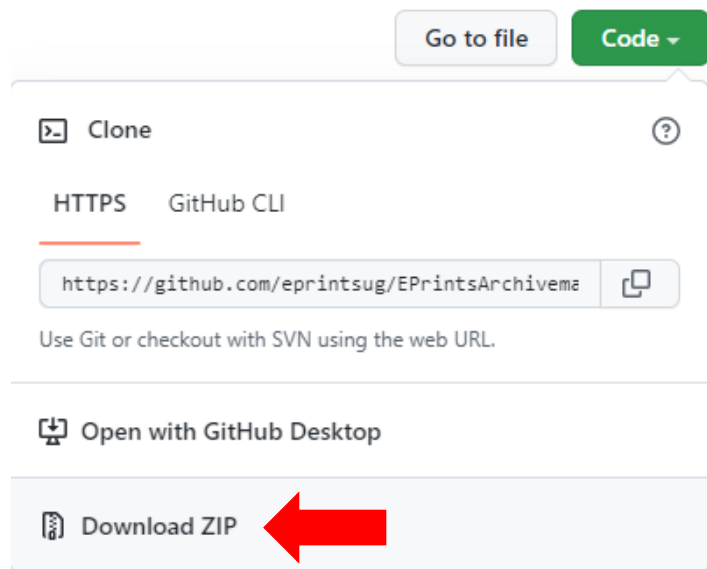
Website/web-based content

If building a website was an integral part of the research project, that website should be preserved as a WARC or WACZ file and included in the deposit. WARC (Web ARChive) is the standardized archival file format for web content. WACZ is a new zipped WARC file format, which allows the archived content to be loaded and replayed more quickly. WARC and WACZ are both recommended formats for preservation and widely supported by web archiving applications.

We recommend using Archiveweb.page, a free and open-source web archiving application, to create a WARC or WACZ file of your website. See the user guide for step-by-step instructions.

Alternatively, it is possible to use a web archiving snapshot service such as archive.today, Perma.cc or Internet Archive's Save Page Now service to archive a single web page and reference the resulting snapshot as an alternative link in the deposit or text. Note that with this option, any dynamic or interactive functionality of the web page will not be preserved.

3D objects (VR, visual effects, digital art, etc.)

Include supporting documentation such as video files and screenshots representing the project.

When possible, export the data in vendor-neutral, standardized formats, to mitigate the risk of format obsolescence. See the Digital Preservation Coalition technology guidance note on Preserving 3D for a list of recommended file formats.

Manage externally referenced files: prefer relative paths to absolute and bind external references prior to deposit.

Review and test work after uploading to different experience platforms such as Sketchfab, 3DHop, and Smithsonian Voyager, to ensure that the files can be rendered in different environments.

Video games/digital games

For applications, include the source code and installation packages (see additional recommendations in the above section on software preservation).

For web-based games, use a web archiving tool to create a WARC file of the website (see additional recommendations in the above section on websites).

If possible, include a full play-through video of the game and any other documentation that would help contextualize the project, such as screenshots or explanatory text files. See the Spectrum Policies for recommended video formats and codecs.

Audio/video

To include an audio or video file as supporting documentation, e.g., a recording of a performance or a video art piece, depositing a copy of the audio or video file itself is preferable for long-term preservation. See the Spectrum Policies for recommended file formats and codecs.

Embedding audio or video files in a textual document such as a PDF can be problematic for preservation. In this case, we recommend also providing an alternative means of accessing the recordings themselves, such as uploading copies of the video or audio as additional files in your deposit, and/or including URLs to a third-party hosting platform like Vimeo or YouTube.

Including a URL instead of a file may also be appropriate in some cases, e.g., should the depositor wish to retain control over access to this content. However, please note that media hosted on external platforms will not be preserved in Spectrum's digital preservation workflow.

Resources consulted

Artefactual Systems and the Digital Preservation Coalition. "Preserving 3D: Data Types Series." *DPC Technology Watch Guidance Notes.* July 2021. http://doi.org/10.7207/twgn21-14.

Artefactual Systems and the Digital Preservation Coalition. "Preserving Documents: Data Types Series." *DPC Technology Watch Guidance Notes.* July 2021. http://doi.org/10.7207/twgn21-07.

Ensom, Tom and Jack McConchie. "Preserving Virtual Reality Artworks." Time-based Media Conservation, Conservation Department, Tate. August 13, 2021. https://www.tate.org.uk/documents/54/tate_pim_preservingvrartworks_01_00.pdf.

Fanning, Betsy A. "Preservation with PDF/A (2nd Edition)." DPC Technology Watch Report 17-01. July 2017. http://dx.doi.org/10.7207/twr17-01.

File formats and standards. In Digital Preservation Handbook (2nd Edition). Digital Preservation Coalition. 2015. https://www.dpconline.org/handbook/technical-solutions-and-tools/file-formats-and-standards.

Greenberg, Jonathan, Karen Hanson, and Deb Verhoff. "Guidelines for Preserving New Forms of Scholarship." September 2021. http://hdl.handle.net/2451/63333.

"OSI Approved Licenses." Open Source Initiative. https://opensource.org/licenses/.

Stapelfeldt, Kirsta, Sukhvir Khera, Natkeeran Ledchumykanthan, Lara Gomez, Erin Liu, and Sonia Dhaliwal. "Strategies for Preserving Digital Scholarship / Humanities Projects." *The Code4Lib Journal*, no. 53 (May 9, 2022). https://journal.code4lib.org/articles/16370.