



Develop a Team Preservation Process - DIGITAL OBJECTS PRESERVATION CHECKLIST FOR TEAMS

Shelley Stall, Alison Specht, Margaret O'Brien, Jeaneth Machicao, Pedro-Luiz-Pizzigatti Corrêa, Romain David, Rorie Edmunds,, Nobuko Miyairi,, Yasuhiro Murayama,, Solange Santos,, et al.

► To cite this version:

Shelley Stall, Alison Specht, Margaret O'Brien, Jeaneth Machicao, Pedro-Luiz-Pizzigatti Corrêa, et al.. Develop a Team Preservation Process - DIGITAL OBJECTS PRESERVATION CHECKLIST FOR TEAMS. AGU. 2023. hal-04177045

HAL Id: hal-04177045

<https://hal.science/hal-04177045>

Submitted on 3 Aug 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Develop a Team Preservation Process – DIGITAL OBJECTS PRESERVATION CHECKLIST FOR TEAMS

Welcome to “Your Open Science Journey”!

Ensure all digital objects created or used by the team are fully documented, preserved for the long-term, and made openly accessible to the team.

Target Audience

Primary: Team or project lead

Secondary: Team or project researcher

This guidance is generalized and will need to be adjusted based on your institution, lab, research team, and/or funder requirements.

A. DETERMINE THE PURPOSE AND CRITERIA FOR PRESERVATION

Determine the purpose and establish the criteria for making decisions on what is to be preserved. This depends on the motivations of the team and the current stage of the research workflow. Decisions should be reviewed and updated as research progresses. Preservation leading practices include:

1. **Documenting the work for ease-of-use and collaboration.** A full set of documentation and metadata should be created for each digital object as well as for the overall project files to support clarity for the team members. For instance, when a new team member joins, documentation will help them to understand prior work. Also, should there be a catastrophic event (e.g., the primary equipment hosting the project files is damaged), the team will have a complete set of documentation.
2. **Approaches for which files to backup:**
 - a. **Ensuring a complete and full backup of all project files.** This is the easiest option but also requires the most space and energy. It is easiest to ensure a full backup is created by including all related files. It requires iterative updates and can be scheduled to happen automatically in the background.
 - b. **Ensuring reproducibility of the research at that moment in time (partial backup).** This is helpful when a full backup is not possible because of technical restrictions or can be done in conjunction with a full backup for the purpose of supporting distributed teams to make sure they have current versions of all files. Reproducing a current version of the research includes only current versions of files, minimizing the number of files to include. Current files being preserved must be tested to ensure reproducibility. Moreover, information loss could occur if prior manipulation of or changes to the files comprising the current, preserved version are not adequately documented. The entire workflow does not necessarily need to be backed up.
 - c. **Ensuring the integrity of the entire research workflow (partial backup).** This is very important for fully capturing all the research steps taken to support published results. This requires preservation of the initial dataset and the software used for cleaning, aggregating, and analyzing the data, as well as the documentation of the decisions made. Research attempts that were not successful should not be included.

B. PLAN TO PRESERVE AND SHARE FOR THE LONG TERM

1. Determine what to preserve (for the team or beyond).

a. Datasets:

- i. Ensure that the datasets created for the entire project and any derived data products, are preserved. If there is a version that you routinely use for sharing within your group, this is likely to be the version you will archive. *Note: Most publishers require that only data supporting the publication be available in a trustworthy repository. However, the full set of data can be cited in the data availability statement, describing which data were used in the published research. This approach ensures all data are preserved together and improves interoperability and reuse.*
- ii. Clean, reorganize, and document your data to make it understandable.
- iii. Datasets should be machine readable; specifically, they should not require human interpretation or proprietary software. More information: [Data Documentation and Citation Checklist](#).
- iv. **Store datasets created by the team in the [team resource defined](#) and track them along with the other datasets/digital objects you are evaluating.** Ensure good version control.

- b. **Software, code, models, scripts, and algorithms:** Document and reorganize your software to make it understandable. Ask a teammate to review it for understandability and future use. It is important that you include documentation of any relevant configuration information for installation, environment, operation, and use of your software. More information: [Software Documentation and Citation Checklist](#).

- c. **Images and associated digital objects; and Conference, training, workshop reports, and materials:** Consult the team's Resources Summary Checklist for the preservation locations. Review the repository guidance for depositing these objects to ensure they are well-documented and in the best possible format for preservation.

2. Evaluate to what extent the digital objects are Open and FAIR.

It is possible for a digital object to be either Open or FAIR, both, or neither. For instance, sensitive health data should be FAIR, but specific personal or health information must be protected. When deciding to what extent your digital objects should be open and FAIR, consider this guide:

a. Make your digital objects "As Open as Possible, As Closed As Necessary."

- i. **Determine which digital objects are open just to the team, and which should be made openly accessible to anyone outside the team** (e.g., other researchers, communities, citizens, etc.). Research should be as open as possible. It is important to consider that there is a continuum between "being open" and "being closed".

When determining "how open" to make your preserved digital objects, consider the following:

- Ability to reproduce the research.
- Requirements of the lab or facility being used.
- Community(ies) norms and/or best practices for the discipline(s).
- Relevant country, funder, publisher, and institutional requirements.

- Digital object usage licenses (e.g., for data created by others).
- Data request agreements (e.g., for sensitive data).
- Indigenous data governance requirements.
- Other ethical considerations. See [Open Science Practices for Teams](#) for more information.
- *Should any of these be in conflict, consult your institution/organization research office for guidance.*

ii. **For your team**, make working documents and objects open and available to your team members. You might also want to make some of these digital objects open beyond your team. For those objects distributed beyond the team we refer to them as “digital outputs”.

iii. **For digital outputs (re: digital objects distributed beyond your research team)**, you can provide:

- **Open digital objects.** These objects are openly accessible to all.
- **Embargoed digital objects.** These objects are open to the research team and a select group of others (e.g., journal peer reviewers) for a defined period of time, then made fully open. The embargoed time can be a set time limit or a certain date, such as when a paper is published.
- **Managed (or mitigated) access to digital objects.** These objects are preserved in a repository that supports access controls. This approach is commonly used for sensitive data.
- **Closed digital objects.** These objects are available to all or some of the research team by prearranged agreement. Commonly these are commercial objects with closed licenses. For some journals, when digital objects are closed, editors may not accept a paper for publication because the research cannot be fully evaluated.

b. Be "As FAIR as Possible"

The [FAIR Guiding Principles](#) support discovery, understanding and reuse of data and digital research objects. FAIR stands for “Findable, Accessible, Interoperable, and Reusable”. By following your community FAIR practices, your digital objects will be ready for preservation and potential reuse.

- Select the [best possible preservation repository](#).
- Make the metadata that describes your digital objects as meaningful as possible to support discovery, access, and reuse.
- Select a license that is as open as possible allowing for reuse.
- Improve the interoperability of your digital objects. For data, use tools that assess and provide feedback on the “FAIR-ness”. For example, the [F-UJI tool](#) or [ARDC self-assessment](#).

3. Decide where to preserve digital objects

Digital objects should be preserved in an appropriate repository. Trusted preservation repositories provide management and curation support for research teams to make their digital objects as interoperable and reusable as possible. For example, data repositories that specialize in a particular type of data will ensure that the documentation and metadata meet the community standards. These aspects are necessary for [datasets to be FAIR](#). Research

teams should seek out the best possible preservation repository for the type of digital object being preserved.

For help making these decisions, refer to the [Open Science Resources and Guidance for Teams](#).

After your team has chosen a repository,

- a. Ensure all links and persistent identifiers are included in the [Digital Object Registry](#) for the project.
- b. Ensure the repository selected has the necessary protections (access controls) for the project digital objects.
- c. Ensure the repository selected is community-accepted and trusted. See [Repository Guidelines](#) for more information.

4. Schedule when to preserve digital objects

It's important to regularly preserve the research outputs for your team, to guard against equipment failure or other unwelcome events. It's tempting to put off the effort to preserve our work in favor of spending this time on our research. Setting up automatic preservation mechanisms are helpful in reducing the time burden of this effort.

- a. **Refer to the Data and Digital Output Management Plan (DDOMP, or DMP) for the planned digital objects preservation schedule.** Update the DDOMP to reflect the current needs of the team.
- b. **Recommended preservation schedule** could include preservation at the following points:
 - Project milestones such as the yearly report
 - Before and after significant research decisions
 - Public dissemination of research results, such as oral presentations, posters, workshops, training sessions
 - Publication of research
 - Completion of the project

C. TRACK AND MANAGE PRESERVED DIGITAL OBJECTS

A **Digital Object Registry** for the project tracks different types of objects preserved in different repositories, optimizing discovery and reuse. The process is described below and documented in the **DDOMP** for the project. The Digital Object Registry includes the details of what digital objects are preserved and when preservation occurs. Ensure the process and criteria are reviewed and updated periodically to be inclusive of new object types added as the research progresses. Follow this process:

1. **Create and update a Digital Object Registry** in conjunction with **the preservation of** objects described in this guidance. The Digital Object Registry could be in the form of a spreadsheet (e.g., Google Sheet) or another type of list.
2. **Include a description of the object, preferred citation, and the persistent identifier** (e.g., Digital Object Identifier) in each registry entry, as well as any other useful information supporting the project. For objects that do not have a persistent identifier, provide a URL and description.

Recommended information includes:

- a. Type of object (e.g., peer reviewed paper, dataset, software, poster, training material)
- b. Title of object
- c. Participating team members (authors)
- d. Designated team member(s) to contact

- e. Description
 - f. Name of hosting location for the preserved digital object
 - g. DOI or URL to the digital object
 - h. Date digital object is submitted for preservation
 - i. Full citation (if existing)
2. **Don't forget to preserve the Digital Object Registry alongside the other digital files pertaining to your project.** Many funders require grant recipients to list their accomplishments in their yearly reports, including a list of all peer-reviewed publications and digital objects. The Digital Object Registry can be provided to the funder during the reporting process to help showcase research progress and accomplishments or can be used as a tracking tool to assist with completing the report.

Quick Links to related checklists:

- [Your Digital Presence](#)
- [Data Documentation and Citation Checklist](#)
- [Software Documentation and Citation Checklist](#)
- [Open Science Practices for Teams](#)
- [Open Science Resources and Guidance for Teams](#)

To recommend updates, please email datahelp@agu.org. Include the name and DOI for the checklist in your email.

To cite this checklist: Stall, Shelley, Specht, Alison, O'Brien, Margaret, Machicao, Jeaneth, Corrêa, Pedro Luiz Pizzigatti, David, Romain, Edmunds, Rorie, Miyairi, Nobuko, Murayama, Yasuhiro, Santos, Solange, Wyborn, Lesley, Vellenich, Danton Ferreira, Vrouwenvelder, Kristina, & Mabile, Laurence. Digital Objects Preservation Checklist For Teams. Zenodo. <https://doi.org/10.5281/zenodo.7402540>