Long-term Preservation Appraisal Checklist

V2 (2024)

This checklist is designed for repository curators and data stewards to record information from curation reviews that can inform long-term retention decisions for datasets in the repository. This checklist is intended to be used at initial deposit and again throughout the life of the dataset in the repository. It can also be used to facilitate conversations with researchers about the preservability of their datasets.

Section 1 asks questions about what the researcher intends, and what the repository and its designated community values.

<u>We recognize that it is hard to evaluate data</u>. "As a result, most research repositories only evaluate the process of data management (practices surrounding description, preservation, dissemination, etc.), and not the data themselves, nor their volume, their quality, their relevance or value."[1]

The questions presented here are asking if the data merit the effort to retain them. Decisions might be informed by what the researcher considers important to preserve, and may include the informational content as well as appearance, interactivity, or some other aspect.

<u>We recommend that you work with what you know now.</u> Is the data prepared in a manner that is useful to the community your repository serves, and does it provide unique information not available elsewhere? Was the data curated to a high standard or is this an uncurated dataset? Consider an evaluation in relation to the broader goals of Open Science, such as the [UN Sustainable Development Goals](#), which could be helpful in determining what is considered timely and useful to contemporary researchers.

Preservation intent

Q1: Is long-term preservation an important factor in the researcher's decision to deposit with your repository?

❏ Yes

If Yes -> Does the deposit contain the necessary documentation (consent forms, project documentation, etc.) to support a recommendation for long-term preservation?

❏ Yes

If No -> Please advise the researcher that long-term preservation may not be possible.

---

[1] Schöpfel J., Prost H., Rebouillat V. (2016). "Research data in current research information systems". In CRIS 2016 : 13th International Conference on Current Research Information Systems, 8-11 June, St Andrews http://dspacecris.eurocris.org/handle/11366/501

Relevance to mission

Q2: Do the materials (continue to) meet your institution or repository's acquisition mandate, collection policy, domain specialty, or other priorities?

❏ Yes

If No -> Is there another repository that would be a better fit? Does your depositor agreement allow for the transfer of materials to other platforms?

Value

Q3: Does the data concern under-documented or marginalized peoples?

❏ Yes

If Yes -> If applicable, have the relevant community stakeholders been consulted about the appropriate custodianship of the data (particularly in the case of First Nations, Inuit, or Métis data)?

❏ Yes

If No -> Are there resources that can help engage the appropriate communities?

Q4: Do you perceive the data to provide valuable evidence of research activity, and demonstrate potential ongoing social, scientific or historical value?

❏ Yes

If No -> Is it possible that certain parts of the dataset are important to retain? Is there enough evidence from the research community or a rationale provided by the depositor to support that?

Uniqueness

Q5: Is the deposit unique (i.e. it is not comprised solely of third-party data and it has not been shared anywhere else)?

❏ Yes

If No -> Is there a need to retain a duplicate copy to facilitate access for a particular community?

Q6: Is the data unique (i.e. it is not the result of a model or generated by code)?

❏ Yes

If No -> Is this a computationally intense output (i.e. one which would require significant resources to re-create and maintain accessibility?)

❏ Yes

Section 2 asks questions about the ease of preservation for the repository. Long-term preservation implies a level of care different from simply holding onto the data. How much effort will it take to keep this dataset useful in the future? Is your organization prepared for the ongoing human, technological, and environmental costs required to preserve the data?

<u>Accessibility is one of the most important factors in appraisal for research data.</u> It matters in the initial curation of the data, and it counts for long-term preservation as well. Without the ability to enable reasonable community access to the data, the effort and costs associated with long-term preservation is not justifiable.

Cost/Economic case

Q7: Is the dataset of a reasonable size (not very large or complex nor organized into many folders, sub-folders, and files)?

❏ Yes

If No -> Is there a subset of the data that could be retained without compromising the utility of the dataset? What are the computational costs associated with re-creating the dataset (if possible)?

❏ Yes

Q8: Are the data easy to access via most personal computers and/or are platform-independent?

❏ Yes

If No -> If this dataset will require support for researchers to effectively reuse (i.e. require emulation of original software environments or access to compute resources), is that possible to support that at your repository?

❏ Yes

Rights and restrictions

Q9: Has the researcher assigned an open license to the files?

❏ Yes

If No -> If the data can be made openly accessible after copyrights or other restrictions expire, is that timeframe acceptable for your repository?

❏ Yes

Preservability of content and context

Q10: Are the data files well documented (i.e. there is sufficient information to ensure that the files will be correctly interpreted over time, including a README, clear description of methodology and variables, and/or links to scholarly publication)?

> ❏ Yes

> *If No ->* Poorly documented data is rarely worth keeping. Is it understandable to all users (non-disciplinary specialists, citizen scientists, etc.) even with the low level of documentation?

>> ❏ Yes

Q11: Have the data been provided in your repository's preferred or accepted file formats, supporting both preservation and access?

> ❏ Yes

> *If No ->* Are the file formats appropriate for the dataset and the intended user community? Is it possible to convert to a non-proprietary format version of the data?

>> ❏ Yes

Overall Preservation Recommendation

Count the number of 'yes' answers to help inform your recommendation.

❏ YES	❏ NO

Appraisal date: _____

Re-appraisal date: _____

Retention notes/concerns: