

# Data documentation for: List of research data repositories that were shut down

Dorothea Strecker (corresponding author)<sup>1</sup>

dorothea.strecker@hu-berlin.de

<https://orcid.org/0000-0002-9754-3807>

Heinz Pampel<sup>1,2</sup>

<https://orcid.org/0000-0003-3334-2771>

Rouven Schabinger<sup>3</sup>

<https://orcid.org/0000-0002-0249-7917>

Nina Leonie Weisweiler<sup>2</sup>

<https://orcid.org/0000-0001-6967-9443>

<sup>1</sup> Humboldt-Universität zu Berlin, Berlin School of Library and Information Science

<sup>2</sup>Helmholtz Association, Helmholtz Open Science Office

<sup>3</sup>Swiss Library Service Platform (SLSP)

Version 2.0 – August 17, 2023

## Abstract

This document describes a dataset that aggregates information on 191 research data repositories that were shut down. The data collection was based on the registry of research data repositories re3data and an extensive content analysis of repository websites and related materials. Documented in the dataset are the period in which a repository was active, the reasons for shutting it down, and the repositories taking over custody of the data after the shutdown.

## Citing the data set

Strecker, Dorothea ; Pampel, Heinz ; Schabinger, Rouven ; Weisweiler, Nina Leonie (2023) List of research data repositories that were shut down. Version 2.0: DOI: 10.5281/zenodo.8233347

## Licenses



This documentation is licensed under Creative Commons Attribution 4.0 International terms.



The data files are licensed under Creative Commons CC0 1.0 Universal terms.

## Motivation

Currently, there is little research investigating the phenomenon of research data repositories being shut down. To address this issue, this dataset collates a list of repositories of all types and disciplines that have been closed and presents information on the shutdown process.

## Methodology

The data collection was based on the registry of research data repositories re3data ("re3data - Registry of research data repositories", n.d.) and an extensive content analysis of repository websites and related materials.

## Inclusion criteria

For this dataset, a repository is considered shut down if data is no longer accessible under the original or a new URL, or if the repository website clearly states that the service has ceased operations (while sometimes maintaining limited access to the data). This definition refers to the life span of a repository and does not relate to the temporal coverage of its collection.

## Data collection

To identify candidates for closed repositories, information on the end date of the 3.069 repositories indexed in re3data was retrieved via the API on January 2, 2023. As shown in Figure 1, this list was then restricted to repositories with an end date; yielding a list of 223 candidate repositories. The authors reviewed the list of candidates. After visiting the website of each repository on the list, duplicate entries (7) were removed, as well as repositories that did not meet the inclusion criteria for identifying closed repositories given above (25).

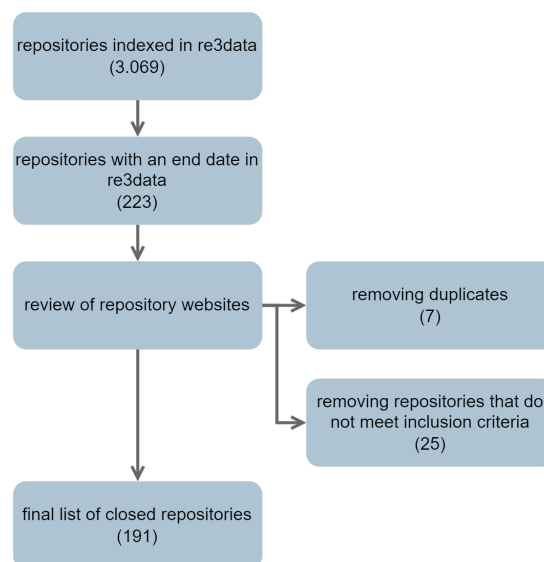


Figure 1: Systematic overview - compiling the final list of repositories that were shut down

For the remaining 191 repositories, information on the start and end date (year) was retrieved from re3data on January 2, 2023. Given that these properties are optional in the re3data Metadata Schema (Strecker et al., 2021), the values were verified intellectually and completed where possible. Between January 2 and 30, 2023, repository websites, both the current version and versions archived by the Internet Archive, as well as additional resources such as data papers describing the repositories were searched for information on the shutdown process. The content analysis of repository websites focused specifically on the risks that resulted in repositories shutting down, and information on the repository taking over custody of the data. This information was collected

for each repository in the dataset and, in addition, the availability of data on the current repository websites was checked. Statements providing information on risks that led to repository shutdown were generalized and summarized into categories. These categories are based on a typology of risks to digital long-term preservation systems by Barateiro et al. (Barateiro et al., 2010) (see Table 1).

vulnerabilities	process	software faults
		software obsolescence
	data	media faults
		media obsolescence
	infrastructure	hardware faults
		hardware obsolescence
		communication faults
network service failures		
threats	disasters	natural disasters
		human operational errors
	attacks	internal attacks
		external attacks
	management	economic failures
		organizational failures
	legislation	legislative changes
		legal requirements

Table 1: Typology of risks to long-term preservation systems, Barateiro et al., 2010, p. 9

## Components of the dataset

### List of research data repositories that were shut down

The resulting data file *01\_repositories.csv* contains information on the 191 closed research data repositories that were identified using the method described above. The following variables are available for all repositories:

- *re3data\_ID*: re3data id of the repository that was shut down
- *start\_date*: start date of the repository; set to *NA* if not available; retrieved from re3data, completed and verified intellectually using repository websites and related materials
- *end\_date*: end date of the repository; retrieved from re3data, verified intellectually using repository websites and related materials
- *risk*: risks resulting in repository shutdown; retrieved from repository websites and related materials, harmonized (see Table 1 for details); set to *NA* if not available
- *succeeded\_by*: re3data id of the repository that has taken over custody of (parts of) the collection after the repository was shut down; set to *NA* if not available; multiple values are separated by semicolons

## About the project

This dataset is part of the project re3data COREF, a joint project by DataCite, the Helmholtz Open Science Office, Humboldt-Universität zu Berlin, and the KIT library.

The project is funded by the German Research Foundation (DFG) under the project number 422587133.

## References

- Barateiro, J., Antunes, G., Freitas, F., & Borbinha, J. (2010). Designing Digital Preservation Solutions: A Risk Management-Based Approach. *International Journal of Digital Curation*, 5(1), 4–17. <https://doi.org/10.2218/ijdc.v5i1.140>
- re3data - Registry of research data repositories. (n.d.). <https://doi.org/10.17616/R3D>
- Strecker, D., Bertelmann, R., Cousijn, H., Elger, K., Ferguson, L. M., Fichtmüller, D., Goebelbecker, H.-J., Kindling, M., Kloska, G., Nguyen, T. B., Pampel, H., Petras, V., Schabinger, R., Schnepf, E., Semrau, A., Trofimenko, M., Ulrich, R., Upmeier, A., Vierkant, P., ... Witt, M. (2021). Metadata Schema for the Description of Research Data Repositories: Version 3.1. <https://doi.org/10.48440/re3.010>