gesis Leibniz Institute for the Social Sciences

Digital Preservation Plan

Measures for the preservation of research data at GESIS – Leibniz Institute for the Social Sciences

Document created by	Recker
Date created	21.02.2023
Version	1.0.0
Status	Published
Responsible	Recker, Mauer
Review	Annually. Next review: January 2024
Purpose of document	Overview of digital preservation measures for research data at GESIS - Leibniz Institute for the Social Sciences
Related documents	Digital Preservation Policy

GESIS – Leibniz Institute for the Social Sciences

Data Services for the Social Sciences

Unter Sachsenhausen 6-8

50667 Köln

E-Mail: jonas.recker@gesis.org

Overview

GESIS takes a tiered approach to the preservation of the digital assets in its collections (see below). Data is ingested into the GESIS Data Archive in two distinct workflows. Each workflow is associated with specific curation[1] and preservation levels (see "Digital preservation measures" below).

1) Archiving Basis

Data depositors describe their data with metadata and submit it along with any documentation using an online submission form. A curator performs a basic ingest check and publishes the data if all requirements for publication are met.

2) Archiving Plus and Premium

Data depositors provide their data and documentation to the curation staff via a suitable channel (e.g. encrypted transmission). In case of Archiving Plus the data is checked intensively and corrections might be performed in consultation with the data depositor. In case of Archiving Premium the previous steps are supplemented by quality assurance measures such as the standardization of labels and missing values as well as error corrections. Data and documentation are enhanced, e.g. by adding metadata on study level ("Plus") or variable level ("Premium"). Archival and dissemination copies in different standard formats are created and published.

For further information on these workflows see https://www.gesis.org/en/data-services/share-data.

Digital preservation measures

Tier	(Sub)Collection	Min. retention period	Curation at ingest	Curation and preservation measures for	
				technical usability	understandability
0 – bitstream preservation	Archiving BASIS: files in non-preferred formats	25 years from publication	Minimal: Ingest checks for compliance with deposit criteria	None	Minimal corrections of errors in metadata (reactive)
1 – functional preservation	Archiving BASIS: files in preferred formats	25 years from publication	Minimal: Ingest checks for compliance with deposit criteria	Files in preferred formats migrated to prevent technical obsolescence (reactive)	Minimal corrections of errors in metadata (reactive)
2- full (conceptual & functional) preservation	Archiving PLUS and PREMIUM	indefinite	Correction and enhancement of data Creation and enhancement of study- and variable-level metadata and documentation Creation of archival and dissemination copies in standard formats	Format migration to prevent technical obsolescence (reactive) Update of archival format to reflect new emerging standards and best practices (proactive) Addition of dissemination formats if needs of designated community change (proactive)	Proactive updates to metadata as existing standards evolve or new standards emerge, including re-cataloging Publication of Errata Help desk for user support

---

[1] "Digital curation, broadly interpreted, is about maintaining and adding value to a trusted body of digital information for both current and future use: in other words, it is the active management and appraisal of digital information over its entire life cycle" (Pennock, M. 2007: " Digital Curation: A Life-Cycle Approach to Managing and Preserving Usable Digital Information". http://www.ukoln.ac.uk/ukoln/staff/m.pennock/publications/docs/lib-arch_curation.pdf).

Back-up, integrity, and authenticity

All files ingested into the GESIS Data Archive are stored securely in our systems and backed up to different locations following community best practices. All files are protected against unauthorized access and manipulation by technical and organizational measures. The integrity of all files in archival storage is monitored using checksums and fixity checks.[2] These practices are governed by an IT-security and backup concept. Provenance metadata are added throughout the entire lifetime of the studies to document any curation activity performed ("audit trail").

Retention

The digital objects in Tier 2 will be preserved indefinitely. This may include transferring them to a successor organization.

For objects in Tiers 0 and 1, the minimum retention period is 25 years from the publication of the data. After this time, GESIS will continue to retain studies until a significant change – e.g., in the organizational or legal framework, the designated community, or the technical environment occurs – which affects the conditions of preservation, dissemination, and use of the digital objects in GESIS's care. Such significant changes include, but are not limited to:

- the discontinuation of GESIS preservation services,
- changes in the legal environment, e.g., with regard to data protection or intellectual property rights, significantly affecting the possibilities of preserving and disseminating the collections,
- technical changes significantly affecting the usability of the preserved studies, e.g., file format obsolescence,
- changes in the needs or knowledge base of the designated community which significantly reduce the usability of the holdings for this community.

In the event of such significant change the studies in question will be reappraised (see "Reappraisal"). This may result in the deaccessioning and deletion of studies and return of the files into the care of the depositors or their legal successors (see "Deaccessioning").

Withdrawal and deletion within the retention period

Once a dataset has been transferred to archival storage and a persistent identifier has been registered, it will not be withdrawn within its retention period unless there are important reasons such as, for example, legal or ethical concerns, or when a copy of the data was published elsewhere. In these cases, the published study may be removed from access, and a tombstone page created in accordance with the DataCite recommended practice[3]. Data held exclusively by the GESIS Data Archive will not be deleted within its retention period unless required for contractual, legal, or ethical reasons.

---

[2] See https://www.dpconline.org/handbook/technical-solutions-and-tools/fixity-and-checksums.
[3] https://support.datacite.org/docs/tombstone-pages.

Reappraisal

Holdings are not automatically reappraised during their agreed retention period. A systematic reappraisal of holdings will occur only in the case of a significant change which affects the conditions of preservation, dissemination, and use of the digital objects in GESIS's care (see "Retention" for details).

In addition, individual studies may be reappraised based on, for example, error reports from users. A reappraisal may result in different actions, e.g., conversion of files to a different format, the addition of metadata or documentation, or deaccessioning of the study (i.e. complete removal from the archive).

Deaccessioning

If reappraisal leads to the decision that a study is to be deaccessioned, GESIS will attempt to contact the depositor or their legal successor to inform them about this decision and return the data in question to their care. When studies are deaccessioned, a tombstone page will be created.