Digital Preservation Policy

Principles of digital preservation at the Data Archive for the Social Sciences

Document created by	N. Schumann
Document translated by	A. Recker, L. Horton
Date created	18.06.2013
Version	This is a translation of the German version 1.4.8 (07.04.2015)
Document saved as	Digital Preservation Policy. Principles of digital preservation at the Data Archive for the Social Sciences
Status	Published
Responsible	N. Schumann, R. Mauer
Review	Annually
Purpose of document	Overview of principles of digital preservation at the GESIS Data Archive for the Social Sciences
Context	See "Related documents"

GESIS – Leibniz Institute for the Social Sciences

Data Archive for the Social Sciences

Unter Sachsenhausen 6-8

50667 Köln

E-Mail: Natascha.Schumann@gesis.org

Introduction

As an infrastructure organization for the social sciences, GESIS – Leibniz Institute for the Social Sciences curates large amounts of empirical social research data. This not only includes data documentation, preparation, indexing and dissemination but also responsibility for digital preservation of data to ensure its long-term usability and understandability.

In this policy the GESIS Data Archive describes its digital preservation principles and strategies. The document expresses GESIS's awareness of responsibilities and measures required to ensure adequate preservation and access. It is addressed at members, committees, management, and employees of GESIS as well as research funders, data producers, data users and other interested parties.

The policy demonstrates the Data Archive is aware of digital preservation challenges. The Archive addresses these challenges through an ongoing process and adapts its workflows where necessary. It has the expertise required for this task and accepts responsibility to ensure usability and interpretability of digital data in its collection. It treats digital assets with necessary care.

The Preservation Policy of the GESIS Data Archive is reviewed every twelve months.

Organizational framework

The GESIS Data Archive was founded at the University of Cologne in 1960 as Zentralarchiv für Empirische Sozialforschung (Central Archive for Empirical Social Research). Since 1987 it has been a member of GESIS – Leibniz Institute for Social Sciences, of which it became a department in 2007. GESIS is member of the Leibniz Association and is publicly funded by the German federal government and the state governments. Since its foundation, the GESIS Data Archive has undertaken continual responsibility for preserving social science research data. Digital preservation activities are integral to GESIS's strategic planning.

Mission

GESIS is an infrastructure organization for the social sciences. As such, its mission is promoting social science research (GESIS statutes, §2, subpar. 1). To support this we offer services and products pertaining to all phases of a research project. Our services are research-based and address current developments to provide best possible support to the scientific community.

An important aspect of this task is permanent provision of research data. Data accessibility is not only prerequisite for transparency and reproducibility, it also offers the possibility of repurposing existing research into new or expanded approaches. It allows exploitation of research data to the fullest possible extent, extending the life of data beyond its original research project. Digital preservation means guaranteeing long-term usability, interpretability, and accessibility of data and is inextricably linked to reliable provision of research data.

Selection and acquisition

The Data Archive primarily collects digital data from empirical social research. GESIS statutes determine the character and content of the service and this accordingly affects data offers. The latter are designed to serve the purpose of investigating "social change in national and international comparative and historical perspective" (Statutes §2, subpar. 2b) and support "international comparative research." The subject of

"social change" is the main focus of GESIS's services and particular importance is assigned to comparability of data across space and time. However, in principle the GESIS Data Archive curates quantitative data sets relevant to social science research questions as long as data is well-prepared and documented.

Access and use

The Data Archive's digital preservation activities make data permanently available for re-use. The Data Archive actively supports and promotes data sharing among researchers.

All publicly accessible datasets receive a persistent identifier, making it possible to clearly reference, locate, and cite datasets over the long term.

The Archive makes studies accessible to its designated user community in empirical social science through different channels and in formats used by this community. The modalities of access and use are laid down in publicly accessible usage regulations.

Challenges

Digital preservation comprises of measures designed to keep data "alive" in the long-term by ensuring interpretability and usability. Achieving this involves more than physical preservation of the bitstream by means of back-ups.

In contrast to traditional (i.e. analog) preservation, digital preservation must address the effects of rapid technological change. Changes to operating environments, hard- and software, file formats, etc. put the use of digital data at risk.

Simultaneously, safeguarding the authenticity and integrity of digital data is of particular importance to ensure its usability is not compromised by damage or unauthorized change.

Digital preservation of research data must also consider legal issues – for example, usage rights, data protection, and intellectual property. Contractual agreements between the Data Archive and rights holder ensure the Archive has the necessary permissions make copies of the data available for re-use or perform format migrations for the purposes of digital preservation.

Another important issue in digital preservation is ensuring data can be understood now and in the future. Among other tasks, this is achieved by describing each study with standardized metadata.

An issue facing digital archives is trustworthiness. To address this, the Data Archive adopts the following principles.

Principles

Digital preservation of research data at the GESIS Data Archive is guided by the following principles:

Strategy

- Preservation activities performed by Data Archive should not be viewed in isolation but in the context of further strategic goals (see also "Related documents"). Accordingly, preservation activities are regularly reviewed and adapted to meet new requirements.

- The Data Archive undertakes active preservation management by means of monitoring procedures and developments in social science research, technical change, and evolving standards.

Standards and co-operation

- The Data Archive embraces the use of recognized standards in fulfilling its tasks (digital preservation, metadata, etc.). Particularly important to planning, implementing, and developing digital preservation is the Open Archival Information System (OAIS) reference model. Recent efforts to establish benchmarks for the trustworthiness of archives are taken into account. Metadata, critical to the long-term preservation, documentation, and use of digital data, are assigned according to recognized standards (e.g. DDI).

- In addition to using standards, the Data Archive is actively involved in their development and promotion. It regards raising awareness of challenges to digital preservation as part of its mission and accordingly offers training and education in this field.

- The Data Archive co-operates with national and international organizations from the social science and humanities as well as the digital preservation domain to exchange insights and experiences, develop and apply standards, and participate in shared monitoring. New research results are taken into account and applied.

Well-documented and traceable processes

- All archival processes are documented and suitably logged. Internal documentation is provided in an appropriate format and regularly updated.

- If data are modified in any way to guarantee their long-term usability and interpretability, e.g. by migrating them to a different format, all changes are documented. Checksums are used to monitor data authenticity and integrity.

Transparency and trustworthiness

- Data depositors and users, research funders, advisory and steering committees as well as the interested public should be familiar with the Data Archive's principles and procedures. The Data Archive provides further information through its website and other channels to meet this interest. In addition, the Archive staff are available for consultation.

- The GESIS Data Archive is a trustworthy long-term archive. It regularly reviews and adapts its workflows in the context of (self-) audits and certification processes drawing on established standards. After carrying out a self assessment it received the Data Seal of Approval.

Technical infrastructure

- The GESIS Data Archive employs a reliable technical infrastructure and suitable measures to ensure the physical safe-keeping of its data collections. Security and risk management are carried out in close co-operation with the IT department. The measures undertaken allow for recovery of the data in case of hardware failures or damages.

- Data is stored redundantly in different locations. Data is frequently backed up to different kinds of storage media and data integrity is constantly monitored. GESIS protects data entrusted to the Data Archive against unauthorized physical or network access. It is aware of the threats posed by extraordinary circumstances or malicious destruction and implements appropriate counter-measures.

Responsibilities

Teams in the Data Archive work together to guarantee long-term usability and comprehension of data. Their work is coordinated and supported by the Data Archive's director and the president of GESIS. The teams are closely connected and engaged in continuing exchanges on the topic of digital preservation. The teams "Akquisition, Sicherung, Datenbereitstellung" ("Acquisition, Archiving, Access") and "Archivinstrumente und Prozesse" ("Archival Tools and Processes") are specifically responsible for digital preservation concerns. They closely cooperate with the GESIS IT department, which is responsible for providing the technical infrastructure.

The director of the Data Archive is responsible for this policy and guarantees its implementation and regular review.

Related documents

- GESIS Mission statement: http://www.gesis.org/en/institute/the-association/mission/

- Statutes: http://www.gesis.org/das-institut/der-verein/satzung/ [in German]

- Usage regulations, Dept. Data Archive for the Social Sciences:

  http://www.gesis.org/en/services/data-analysis/data-archive-service/usage-regulations/

- Archive agreement:
  http://www.gesis.org/fileadmin/upload/institut/wiss_arbeitsbereiche/datenarchiv_analyse/Archivierungsvertrag_GESIS_Datenarchiv_v9__englisch.pdf

- Collection Policy:
  http://www.gesis.org/fileadmin/upload/institut/wiss_arbeitsbereiche/datenarchiv_analyse/2013-07-08_Was_wir_sammeln.pdf [in German]Data Seal of Approval – Assessment:
  https://assessment.datasealofapproval.org/assessment_116/seal/html

- Hinweise zum Datenschutz. Rechtlicher Rahmen und Maßnahmen zur datenschutzgerechten Archivierung sozialwissenschaftlicher Forschungsdaten:
  http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis_reihen/gesis_papers/GESIS-Papers_2015-01.pdf [in German]