

Digital Preservation Policy

Grundsätze der digitalen Langzeitarchivierung am Datenarchiv für Sozialwissenschaften

Dokument erstellt von	Natascha Schumann
Dokument erstellt am	26.09.2012
Name des Dokuments	Digital Preservation Policy. Grundsätze der digitalen Langzeitarchivierung am Datenarchiv für Sozialwissenschaften
Version	1.4.8 (07.04.2015)
Dokument gespeichert unter	Richtlinie zur digitalen Langzeitarchivierung am Datenarchiv für Sozialwissenschaften (Digital Preservation Policy)
Dokument geändert am / durch	<ul style="list-style-type: none"> • 18.02.2013 N. Schumann, A. Müller, A. Recker, W. Zenk-Möltgen, R. Mauer • 24.02.2013: N. Schumann, Ergänzung • 26.02.2013 Reiner Mauer: Erweiterung um ‚Organisatorischer Rahmen‘; generelle Überarbeitung • 10.06.2013: Version 1.4.6: Genehmigung AL; kleinere Änderungen • 18.12.2013: Version 1.4.7: Kleine Änderungen, N. Schumann • 07.04.2015: N. Schumann: Ergänzungen
Status	Veröffentlicht
Verantwortlich für das Dokument	N. Schumann, R. Mauer
Regelmäßige Überprüfung	Einmal jährlich
Zweck des Dokuments	Darstellung der Prinzipien der Langzeitarchivierung im GESIS Datenarchiv für Sozialwissenschaften
Kontext zu anderen Dokumenten	Siehe Punkt „Verwandte Dokumente“

GESIS - Leibniz-Institut für Sozialwissenschaften
 Datenarchiv für Sozialwissenschaften
 Unter Sachsenhausen 6-8
 50667 Köln
 E-Mail: Natascha.Schumann@gesis.org

Einleitung

Als Infrastruktureinrichtung für die Sozialwissenschaften betreut GESIS – Leibniz-Institut für Sozialwissenschaften Forschungsdaten der empirischen Sozialforschung in erheblichem Ausmaß. Dies umfasst neben Dokumentation, Aufbereitung, Erschließung und Bereitstellung insbesondere auch die digitale Langzeitarchivierung dieser Daten, d.h. den Erhalt von Nutzbarkeit und Interpretierbarkeit der Daten über lange Zeiträume.

Die vorliegende Richtlinie beschreibt grundlegende Prinzipien und Strategien, die das GESIS Datenarchiv zur Langzeitarchivierung von Forschungsdaten einsetzt. Sie bringt zum Ausdruck, dass GESIS sich der damit einhergehenden Verantwortung bewusst ist und entsprechende Maßnahmen ergreift. Das vorliegende Dokument richtet sich an Mitglieder, Gremien, Leitung und Mitarbeiter der GESIS sowie an Forschungsförderer, Datenproduzenten und -nutzer sowie sonstige interessierte Personenkreise.

Die Policy zeigt, dass das Datenarchiv die Herausforderungen der digitalen Langzeitarchivierung kennt, sich mit diesen in einem kontinuierlichen Prozess auseinandersetzt und seine Arbeitsabläufe, wenn erforderlich, entsprechend anpasst. Das Datenarchiv besitzt die dafür notwendige Expertise, übernimmt die Verantwortung für den Erhalt der Nutzbarkeit und Interpretierbarkeit der ihm anvertrauten digitalen Daten und behandelt diese mit der gebührenden Sorgfalt. Die vorliegende Preservation Policy wird alle zwölf Monate auf ihre Aktualität überprüft.

Organisatorischer Rahmen

Das GESIS Datenarchiv wurde 1960 als Zentralarchiv für Empirische Sozialforschung an der Universität zu Köln gegründet, ist seit 1987 Mitglied und seit 2007 eine Abteilung von GESIS – Leibniz-Institut für Sozialwissenschaften e.V. GESIS ist Mitglied der Leibniz-Gemeinschaft und wird durch Mittel, die Bund und Länder zur Verfügung stellen, finanziert. Seit seiner Gründung hat das GESIS Datenarchiv kontinuierlich die Verantwortung für die Langzeitarchivierung sozialwissenschaftlicher Forschungsdaten übernommen. Die Aktivitäten im Bereich Langzeitarchivierung sind integraler Bestandteil der strategischen Gesamtplanung von GESIS.

Mandat, Aufgabe

GESIS verfolgt als Infrastruktureinrichtung für die Sozialwissenschaften das übergreifende Ziel, die sozialwissenschaftliche Forschung zu fördern ([GESIS-Satzung](#), §2, Abs. 1). In diesem Sinne bieten wir Dienstleistungen und Produkte zu allen Phasen eines Forschungsvorhabens in den Sozialwissenschaften an. Unsere Dienstleistungen sind forschungsbasiert und greifen neue Entwicklungen auf, um so der Scientific Community die bestmögliche Unterstützung für ihre Forschung zu bieten.

Ein wichtiger Bestandteil dieser Aufgabe ist die dauerhafte Versorgung der Wissenschaft mit Forschungsdaten. Die Zugänglichkeit von Daten ist nicht nur grundlegende Voraussetzung für Transparenz und Überprüfbarkeit von Forschungsprozessen, sondern bietet auch die Möglichkeit mit erweiterten oder neuen Ansätzen an bestehende Forschung anschließen zu können. Zudem ermöglicht sie die Nutzung bisher nicht ausgeschöpfter Analysepotentiale, d.h. die Auswertung vorhandener Daten unter gänzlich neuen Fragestellungen. Mit einer stabilen und verlässlichen Bereitstellung von Forschungsdaten untrennbar verbunden sind Maßnahmen zur

Langzeitarchivierung, die deren Nutzbarkeit, Interpretierbarkeit und Zugänglichkeit über lange Zeiträume sicherstellen.

Auswahl und Akquisition

Der Sammelschwerpunkt des Datenarchivs liegt auf digitalen Daten der empirischen Sozialforschung. Die Satzung der GESIS bestimmt die inhaltliche Ausrichtung des Dienstleistungs- und damit auch die des Datenangebots. Dieses soll der Erforschung „gesellschaftlicher Entwicklungen in nationaler, international vergleichender und historischer Perspektive“ (Satzung §2, Abs. 2b) dienen und „international vergleichende Forschung“ unterstützen. Dies bedeutet, dass der Vergleichbarkeit von Daten über die Dimensionen Zeit oder Raum eine besondere Bedeutung zukommt und die „gesellschaftliche Entwicklung“ der zentrale Fokus der Dienstleistungen ist. Grundsätzlich werden jedoch alle quantitativen Daten archiviert, die für die Beantwortung sozialwissenschaftlich relevanter Fragestellungen geeignet sind, sofern diese gut aufbereitet und dokumentiert sind.

Zugang und Nutzung

In erster Linie zielen die Maßnahmen zur Langzeitarchivierung am Datenarchiv darauf ab, die archivierten Daten dauerhaft für die Nachnutzung zur Verfügung stellen zu können. Das Datenarchiv unterstützt und fördert aktiv die Bereitschaft von Forschenden, ihre Daten mit anderen zu teilen (data sharing).

Alle öffentlich zugänglichen Studien erhalten obligatorisch einen persistenten Identifikator, damit sie eindeutig identifizierbar, referenzierbar, lokalisierbar und langfristig zitierfähig sind.

Der Datenservice des Datenarchivs stellt seinen Nutzerinnen und Nutzern aus dem Bereich der empirischen Sozialwissenschaften (Designated Community) Studien über unterschiedliche Zugangswege zur Verfügung. Dabei erfolgt die Bereitstellung der Daten in den in der sozialwissenschaftlichen Community verbreiteten Formaten. Die Modalitäten des Zugangs und der Nutzung werden im Rahmen einer öffentlich zugänglichen Benutzungsordnung geregelt.

Herausforderungen

Digitale Langzeitarchivierung beinhaltet eine Vielzahl unterschiedlicher Maßnahmen, die darauf abzielen, digitale Forschungsdaten langfristig zu erhalten und deren Interpretierbarkeit und Nutzbarkeit sicherzustellen. Diese Maßnahmen gehen weit über die rein physische Erhaltung des Datenstroms im Sinne von Back-Ups hinaus.

Anders als im Bereich der analogen Archivierung hat die digitale Archivierung mit den Auswirkungen eines rasanten Technologiewandels umzugehen. So gefährdet die Veränderung von Betriebsumgebungen, Hard- und Software, Formaten etc. die Nutzbarkeit der Daten.

Darüber hinaus ist es gerade bei digitalen Ressourcen besonders wichtig, die Authentizität und Integrität der Daten zu gewährleisten, um sicherzustellen, dass diese nicht unbefugt verändert wurden oder beschädigt sind und sich damit für eine (Nach-) Nutzung nicht mehr eignen.

Rechtliche Fragen zum Beispiel in Bezug auf Nutzungsrechte, Datenschutz, Urheberrechte sind bei der Langzeitarchivierung digitaler Daten zu berücksichtigen. So muss durch entsprechende Verträge festgelegt werden, dass das Archiv zur langfristigen Erhaltung nicht nur Kopien der Daten anfertigen, sondern beispielsweise auch Änderungen in Form von Migrationen durchführen darf.

Ein weiterer wichtiger Punkt bei der Langzeitarchivierung ist die Gewährleistung der korrekten Interpretierbarkeit der Daten (auch für die zukünftige Nutzung). Dies wird u.a. durch die Beschreibung der Studien mit standardisierten Metadaten sichergestellt.

Gerade für digitale Archive stellt die Vertrauenswürdigkeit eine besondere Herausforderung dar. Um diesen gerecht zu werden hat das Datenarchiv für Sozialwissenschaften die folgenden Prinzipien aufgestellt und sich zu eigen gemacht.

Prinzipien

Das GESIS Datenarchiv für Sozialwissenschaften orientiert sich bei der Langzeitarchivierung von digitalen Daten an folgenden Prinzipien:

Strategie und Planung

- Alle Erhaltungsmaßnahmen stehen im Zusammenhang mit weiteren strategischen Zielen des Datenarchivs und sind nicht isoliert zu betrachten (siehe auch „Verwandte Dokumente“). Sie werden in regelmäßigen Abständen überprüft und bei Bedarf an neue Anforderungen angepasst.
- Das Datenarchiv betreibt ein aktives Preservation Management durch Monitoringprozesse v.a. in den Bereichen sozialwissenschaftlicher Forschung, technischer Entwicklungen und Standards.

Standards und Kooperationen

- Das Datenarchiv erkennt die Notwendigkeit an, sich bei der Erfüllung seiner Aufgaben an anerkannten Standards zu orientieren (Langzeitarchivierung, Metadaten, etc.). Dabei kommt dem OAIS-Referenzmodell bei der Planung, Durchführung und Entwicklung der digitalen Langzeitarchivierung eine besondere Bedeutung zu. Die jüngsten Entwicklungen im Hinblick auf die Etablierung von Normen zur Vertrauenswürdigkeit von Archiven finden Berücksichtigung. Metadaten, die für Langzeitarchivierung, Dokumentation und Nutzung gleichermaßen wichtig sind, werden gemäß anerkannten Standards behandelt (z. B. DDI).
- Das Datenarchiv wendet Standards nicht nur an, sondern beteiligt sich aktiv an deren Entwicklung und Verbreitung. Es versteht es als seinen Auftrag, Bewusstsein für die Herausforderungen der digitalen Langzeitarchivierung zu schaffen und bietet zu diesem Zweck Fortbildungs- und Trainingsmaßnahmen an.
- Das Datenarchiv kooperiert mit nationalen und internationalen Einrichtungen, sowohl aus der sozialwissenschaftlichen als auch der Langzeitarchivierungs-Community. Erkenntnisse und Erfahrungen werden ausgetauscht, Standards entwickelt und angewendet und gemeinsames Monitoring betrieben. Neue Forschungsergebnisse werden berücksichtigt und angewendet.

Dokumentation und Nachvollziehbarkeit von Prozessen

- Alle Prozesse der Archivierung sind hinreichend dokumentiert und protokolliert. Die interne Dokumentation wird in geeigneter Form bereitgehalten und laufend aktualisiert.

- Alle Veränderungen, die zum Zwecke der langfristigen Nutzbarkeit und Interpretierbarkeit der Daten vorgenommen werden, z.B. durch eine Migration, werden nachvollziehbar dokumentiert. Das gilt ebenfalls für den aktiven und kontinuierlichen Einsatz von Checksummen als Maßnahme zur Bewahrung der Authentizität und Integrität der Daten.

Transparenz und Vertrauenswürdigkeit

- Sowohl Datengeber als auch Nutzer, ebenso wie Forschungsförderer, beratende u. leitende Gremien und die interessierte Öffentlichkeit sollen mit den Vorgehensweisen und Prinzipien des Datenarchivs vertraut sein. Dazu stellt das Archiv auf seiner Webseite und über weitere Informationskanäle entsprechende Informationen zur Verfügung. Darüber hinaus stehen die Mitarbeiterinnen und Mitarbeiter des Datenarchivs auch zur persönlichen Beratung zur Verfügung.
- Das Datenarchiv ist ein vertrauenswürdige Langzeitarchiv; im Rahmen von (Selbst-) Audits und Zertifizierungen gegen die etablierten Standards werden die Arbeitsabläufe regelmäßig überprüft und ggf. angepasst. Nach einer Selbstevaluierung hat es das Data Seal of Approval erhalten.

Technische Infrastruktur

- Das GESIS Datenarchiv setzt geeignete Maßnahmen und die entsprechende Infrastruktur zur physischen Sicherung der Datenbestände ein. Das Sicherheits- und Risikomanagement erfolgt in enger Abstimmung mit der IT-Abteilung. Die Maßnahmen garantieren im Falle von Ausfällen oder Beschädigungen die Wiederherstellung der Daten.
- Alle Daten werden redundant und an unterschiedlichen Orten vorgehalten. Es erfolgen regelmäßige Sicherungen des aktuellen Datenbestands auf unterschiedlichen Arten von Speichermedien. Die Integrität des Datenbestandes wird kontinuierlich überwacht. GESIS schützt die dem Datenarchiv anvertrauten Daten vor unberechtigtem Zugriff vor Ort als auch vor solchen über das Netzwerk und ist sich der Bedrohung durch höhere Gewalt und gezielte Zerstörung bewusst und hat entsprechende Gegenmaßnahmen ergriffen.

Verantwortlichkeiten

Der langfristige Erhalt von Nutzbarkeit und Interpretierbarkeit der Daten wird durch alle Teams des Datenarchivs sichergestellt. Dabei werden sie von der Leitung des Datenarchivs und der GESIS unterstützt und koordiniert. Sie sind dabei eng miteinander vernetzt und tauschen sich regelmäßig zu Themen der Langzeitarchivierung aus. Besondere Verantwortung für die Belange der Langzeitarchivierung übernehmen dabei die Teams ‚Akquisition, Sicherung, Datenbereitstellung‘ und ‚Archivinstrumente und Prozesse‘. Diese kooperieren eng mit der IT-Abteilung der GESIS, die für die Bereitstellung der technischen Infrastruktur verantwortlich ist.

Der Leiter des Datenarchivs verantwortet die Policy und garantiert deren Umsetzung sowie regelmäßige Überprüfung.

Verwandte Dokumente

- Leitbild: <http://www.gesis.org/das-institut/der-verein/leitbild/>
- Satzung: <http://www.gesis.org/das-institut/der-verein/satzung/>
- Benutzungsordnung:
http://www.gesis.org/fileadmin/upload/dienstleistung/daten/umfragedaten/bgordnung_benutzungsordnung.pdf
- Archivierungsvertrag:
http://www.gesis.org/fileadmin/upload/institut/wiss_arbeitsbereiche/datenarchiv_analyse/Archivierungsvertrag_GESIS_Datenarchiv_annotiert.pdf (dort ist ein Mustervertrag zu finden)
- Collection Policy:
http://www.gesis.org/fileadmin/upload/institut/wiss_arbeitsbereiche/datenarchiv_analyse/2013-07-08_Was_wir_sammeln.pdf
- 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