

GESIS - Datenarchiv für Sozialwissenschaften Köln/D



Begehung am 27.11.2012, aktualisiert am 2.12.2014

Aufgaben	
Hauptaufgaben:	Archivierung, Aufbereitung, Dokumentation und Bereitstellung von Forschungsdaten, vorwiegend quantitative Daten der empirischen Sozialforschung; Registrierungsagentur für PIDs; Nachweiskatalog für Daten der empirischen Sozialforschung im Aufbau (seit Ende 2011)
weitere Aufgaben:	Beratung zur Datenerhebung, Schulung/ Training Datenmanagement, Sekundäranalysen archivierter Daten
nicht Teil der Aufgaben:	keine amtlichen Daten, keine qualitativen Daten, Bereitstellung von Forschungswerkzeugen
relevante Fachrichtung:	Sozialwissenschaften mit Schwerpunkt auf Soziologie und Politikwissenschaften

Allgemeine Angaben	
Rechtsform:	„Datenarchiv für Sozialwissenschaften (DAS)“ ist eine von fünf Abteilungen innerhalb des GESIS-Institutes, das Mitglied der Leibniz-Gemeinschaft ist und als Verein organisiert ist (80 Mitglieder)
Gründung/Geschichte:	1960 als erstes sozialwissenschaftliches Datenarchiv in Europa gegründet (zugehörig zur Uni Köln) 1986 Zentralarchiv wird Mitglied beim neu gegründeten GESIS-Institut 2007 Zusammenführung der drei Teil-Institute von GESIS zu einem Institut, seitdem ist das Datenarchiv eine wissenschaftliche Abteilung 2008 Leibniz-Institut für Sozialwissenschaften
Anzahl Mitarbeiter:	GESIS insgesamt: 268 (davon ca.170 wiss. MA) Datenarchiv insgesamt: 50 FTE (davon 30 Vollzeit und 30 wiss. MA) Anzahl FTE (Planstellen) ca. 2/3 Anzahl Hilfskraftstellen (Planstellen) unbekannt Anzahl FTE (befristet, Drittmittel) ca. 1/3 Anzahl Hilfskraftstellen (befristet, Drittmittel) unbekannt Anzahl WissMA 30 Anzahl IT max. 3 für Datenarchiv alleine + zentrale GESIS-IT Abteilung + projektbezogene Entwickler + Mitarbeiter in GESIS-Abteilung „Wissenstechnologien für Sozialwissenschaften“
Zertifizierung:	Data Seal of Approval 2014-2015
Referenzmodell:	OAIS

Datenbestand	
Verpflichtung Daten zu sichern?	Es besteht keine generelle Verpflichtung, aber BMBF und DFG empfehlen, Daten bei GESIS zu archivieren
Findet gezielte Datenakquise statt?	Ja (Daten, die zum wissenschaftlichen Profil des Archivs passen, werden gezielt gesucht)
Anzahl an Projektarchiven:	6.100 Studien (darunter einige Langzeitstudien)
Anzahl an „grauer Literatur“:	nicht relevant

Anzahl an Doktorarbeiten:	nicht relevant
Anzahl an retrodigitalisierten Zeitschriften:	nicht relevant
Anzahl an Einträgen Nachweiskatalog:	ca. 5.200 (Stand Dezember 2014) Der zentraler Katalog für die Daten im Archiv ist erreichbar unter: http://www.gesis.org/dbk Im Aufbau befindet sich ein zentrales Nachweissystem für die Sozialwissenschaften, das nicht nur Daten von GESIS nachweist.
Gesamtgröße Archiv:	1.000 – 2.000 GB (1 – 2TB) mit ca. 600.000 digitalen Objekten (Dateien/Ordernern)
Datensatzgröße:	zwischen 200-400 MB/Studie
Datenzuwachs jährlich:	150 Studien = 30 – 60 GB
Altdaten vorhanden:	Ja

Datengenerierung

technische Unterstützung bei Datenerhebung (z. B. durch Tools, Software etc.):	Nein
fachliche Unterstützung bei Datenerhebung:	Ja

Datenupload

Online-Upload möglich?	Ja, das Datenarchiv betreibt als Ergänzung zum Standardarchiv eine Data Sharing Platform (datorium), die eine eigenständige Dokumentation und Publikation von Forschungsdaten ermöglicht: https://datorium.gesis.org/
Auswahlkriterien für Aufnahme von Daten:	In das Datenarchiv aufgenommen werden Studien, die: - Aussagen über die deutsche Bevölkerung oder über Teile von ihr erlauben - von deutschen Forschern durchgeführt wurden, unabhängig davon, ob sie sich auf Deutschland beziehen - ganz allgemein für sozialwissenschaftliche Benutzer von Interesse sein könnten
Werden angebotene Daten abgelehnt? (Gründe)	geringer wissenschaftlicher Wert (wenig nachgefragte Studien), sensible oder vertrauliche Daten, datenformatbezogene Gründe (Lochkarten)
Vertragsverhältnis Datengeber - Datenarchiv:	Benutzungsordnung, Archivierungsvertrag http://www.gesis.org/unser-angebot/archivieren-und-registrieren/datenarchivierung/
Software/Collection Registry:	da ra = DOI-Registrierungsservice für Sozial- und Wirtschaftsdaten in Deutschland

Kuratierung von Daten

Arbeitsschritte zur Datenarchivierung:	<ul style="list-style-type: none"> - Ablage in temporärem, internen „Postkasten“ - Vergabe Archivnummer - Virencheck - Prüfung der Lesbarkeit der Dateien - Checksummen erzeugen - inhaltliche Aufbereitung (Dokumentation, Anonymisierung, Vereinheitlichung etc.) - technische Aufbereitung - Dokumentation Kommunikation - Ablage in Ordnerstruktur auf Archiv-Server 				
Werkzeuge für Datenarchivierung:	<table style="width: 100%; border: none;"> <tr> <td style="width: 50%;">Eigenentwicklung</td> <td>individuelle Scripts (z. B. zur Anonymisierung von Variablen)</td> </tr> <tr> <td>verfügbare Software (kommerziell, OpenSource)</td> <td>keine</td> </tr> </table>	Eigenentwicklung	individuelle Scripts (z. B. zur Anonymisierung von Variablen)	verfügbare Software (kommerziell, OpenSource)	keine
Eigenentwicklung	individuelle Scripts (z. B. zur Anonymisierung von Variablen)				
verfügbare Software (kommerziell, OpenSource)	keine				

Versionskontrolle:	Ja, vor allem wichtig für Langzeitstudien, die regelmäßig aktualisiert/ergänzt werden: dreistellige Versionsnummer, in separaten Unterordnern auf Archiv-Server bzw. einzelnen Datensätzen in Web-Nachweiskatalog erfasst, jede neue Version erhält neue PID
Verhältnis manuelle/automatisierte Datenkuratierung:	überwiegende manuelle Kuratierung, z. T. unterstützt durch selbstprogrammierte Scripts
Aufwand Datenkuratierung insgesamt (Personen, Zeit):	unbekannt
wissenschaftliche Bewertung von Daten:	Ja
Verwendung Repository-System:	Das Datenarchiv betreibt als Ergänzung zum Standardarchiv eine Data Sharing Platform (datorium) auf DSpace-Basis, die eine eigenständige Dokumentation und Publikation von Forschungsdaten ermöglicht: https://datorium.gesis.org/
Software zur Verwaltung/Ablage der Archivdateien:	strukturierte, verzeichnisbasierte Dateiablage Eigenentwicklung verschiedene Scripts für Synchronisation, Checksummen etc. verfügbare Software keine (kommerziell, OpenSource)
Software zur Verwaltung Metadaten/Rechte/Versionen:	Ja, alles über eine Web-Anwendung Eigenentwicklung Web-Anwendung auf Basis von MySQL verfügbare Software keine (kommerziell, OpenSource)

Archivierung von Daten

technischer Partner für LZA:	kein externer Partner
Anzahl redundanter Datenkopien:	3 (an GESIS Standorten in Köln, Mannheim, Berlin)
Administration Archivsystem inhouse/extern:	inhouse
garantierter Zeitraum:	unbegrenzt
Erfahrungen mit technischer Migration von Archivdateien:	Ja

Bereitstellung von Daten

Arbeitsschritte zur Datenbereitstellung:	- Dokumentation und Metadaten in Studie in Web-Nachweiskatalog - automatisierter Abgleich Archivdateien (→ Dateiliste) und Angaben Web-Nachweiskatalog - Rechte für Dateien festlegen - Freigabe von Metadaten und Dateien
Werkzeuge für Datenbereitstellung:	Eigenentwicklung Scripts, um Dateien aus Archiv-Verzeichnis in Web-Nachweiskatalog einzubinden / anzuzeigen verfügbare Software keine (kommerziell, OpenSource)
Registrierung bzw. Login für Nutzer notwendig?	für Suche in Datenbestandskatalog nicht für Download von Daten Anmeldung erforderlich
Wird die IP und der Zugriff geloggt?	Ja
Gibt es verschiedene Authentifizierungsstufen?	Ja

Zugriff:	- Online über Portal (download) oder FTP - postalisch per Datenträger - bei sensiblen Daten nur Einsicht in sog. Safe-Räumen von GESIS (on-site Nutzung)
Wie wird ein Missbrauch von Daten verhindert/kontrolliert (z. B. wirtschaftliche Nutzung)?	verschiedene technische Maßnahmen, um Missbrauch vorzubeugen Nutzer müssen Nutzungsordnung zustimmen; bei bestimmten Daten müssen zusätzliche Nutzungsverträge geschlossen werden, besonders schützenswerte Daten können nur vor Ort unter besonderen technischen und organisatorischen Bedingungen im Secure Data Center genutzt werden.
Rollenkonzept:	Datengeber Datennutzer Datenarchivar Administrator
Qualitätssicherung:	Daten werden nach methodischer Qualität beurteilt
Schutzmechanismen für bestimmte Informationen:	keine automatisierten Mechanismen, aber es wird ein hoher manueller Aufwand betrieben, um Datenschutzbestimmungen einzuhalten personenbezogene Metadaten/Dateien Ja, bei Datenaufbereitung erfolgt Anonymisierung individuelle Scripts für Veränderung der Daten raumbezogene Metadaten/Dateien ja, sofern auf Personen beziehbar individuelle Scripts für Veränderung der Daten

Metadaten & Interoperabilität

Mindestanforderungen an Daten:	- inhaltliche Relevanz - ausreichende Dokumentation und Materialien (Methodenberichte, Codepläne etc.) - keine technischen Formatvorgaben - Berechtigung der Datengeber zur Weitergabe von Daten an ein Archiv
Verwendung von Standards:	Ja
Metadatenmodell:	DDI (Data Documentation Initiative) SDMX (Statistical Data and Metadata Exchange) DataCite 2.1 eigene Anpassungen an Metadaten? Nein, falls erforderlich aktive Beteiligung an der Weiterentwicklung von Standards
vorhandene Schnittstellen:	unbekannt
Sichtbarkeit der Metadaten in anderen Portalen/Aggregatoren:	diverse sozialwissenschaftliche (Meta-)Datenportale: ZACAT, HISTAT, CESSDA-Portal, sowiport, DBK (Retrieval, Download, Online-Analyse)
Sichtbarkeit der archivierten Dateien in anderen Portalen/Aggregatoren:	unbekannt
System für Persistente Identifikatoren:	DOI (pro Studie), jede neue Version einer bereits archivierten Studie erhält eine neue DOI; DOIs werden nicht nur für Studien im eigenen Datenarchiv vergeben, sondern auch für externe Daten (mittels da ra - Daten-Registrierungs-Agentur, Mitglied in DataCite) Anzahl PIDs insgesamt 5.915 DOIs für GESIS-Datenarchiv (Stand 09.2013)

Nicht-Technische Dienstleistungen

Bereitstellung Guidelines, Ratgeber etc.:	GESIS insgesamt unterstützt alle Phasen eines Forschungsdatenzklus durch verschiedene Dienstleistungen und führt auch aktiv Forschungsprojekte durch; einzelne Aufgaben umfassen - Recherche von Forschungsdaten und Informationen - Studiendesign und -planung - Erhebung und Aufbereitung von Forschungsdaten - Analyse der Forschungsdaten - Registrierung und Archivierung von Forschungsdaten - Projektplanung und -management
---	---

Antragsberatung:	Ja
Schulungen/Workshops:	Ja
Weiterentwicklung Standards:	Ja, aktive Beteiligung
Support Datenmanagement:	ja

Finanzen: Einnahmen

Jahresbudget:	20 Mio. davon speziell für das Datenzentrum unbekannt
Basisfinanzierung:	GESIS wird als Institut der Leibniz-Gemeinschaft durch eine Bund (80%)-Länder (20%)-Finanzierung getragen, damit auch das Datenarchiv zu 100 % Zeitraum der vertraglich Finanzierung 7 Jahre Nachhaltige Finanzierung? Ja, über die Leibniz-Gemeinschaft und die in der Satzung festgeschriebenen Verpflichtungen (Garantie für LZA)
Drittmittel:	Ja, ca. 3 Mio. = ca. 15 % am Gesamtbudget
restliche Finanzierung:	gestaffelte Gebühren für Benutzung (vernachlässigbarer Anteil am Gesamtbudget)
Gebühren für das Archivieren von Daten:	Standardarchivierung ist frei, sofern eine besondere Aufbereitung (added value) erwünscht ist, müssen Mittel durch Datengeber oder Drittmittel eingeholt werden; Kosten werden dann individuell festgelegt
Gebühren für die Nutzung von Daten:	prinzipiell werden keine Gebühren erhoben; für die manuelle Bereitstellung von Daten und vereinzelt Dienstleistungen werden feste Preise gemäß einer Gebührenordnung verlangt, die je nach Status des Nutzers, Anzahl der erwünschten Daten, Art der Datenzustellung Preise festlegt (min. 15 € für 1 Studie bis max. 250 € für 40 Studien); Beträge dienen vor allem als Schutzgebühren / handling charges und sind in keiner Weise kostendeckend; Datengeber sind von diesen Gebühren befreit

Finanzen: Ausgaben

Kostenstruktur nach Kostenträger:	Personalkosten	2/3
	Planstellen	unbekannt
	Drittmittelstellen	unbekannt
	Hardware/Software	1/3 Sachausgaben
	Reisekosten, Gebäudeinstandhaltung	1/3 Sachausgaben
Kostenstruktur nach Kostenstellen:	Daten akquirieren und aufbereiten	unbekannt
	Daten archivieren	unbekannt
	Daten zugänglich machen	unbekannt
	Softwareentwicklung	unbekannt
	Kosten für PIDs	keine Kosten für individuelle PIDs, es wird eine Pauschale für die Mitgliedschaft in DataCite bezahlt
Höhe Durchschnitts-Kosten pro 1 GB Daten:	unbekannt	

Negativa

nicht funktionierende/ optimierbare Workflows oder Prozesse:	DAS bietet zwar umfassenden Archivierungsservice an, aber sehr starker Fokus auf quantitative Umfragedaten, keine qualitativen Daten
--	--

nicht gelöste Probleme:

Ausblick für die Zukunft:

stärkere Modularisierung des Archivierungsservices, um bedarfsgerechte Dienste bieten zu können:

- Langzeitarchivierung (projektspezifisch, z. B. als Hintergrunddienst für Datenzentren)
- Datenservice: Datenzugang u. Nutzerservice (projektspezifisch)
- Aufbereitung und Dokumentation (projektspezifisch, z. B. große Erhebungsprogramme)
- Datenregistrierung
- zentrale Datennachweisdienste (also auch für Daten außerhalb der GESIS)
- „Selbstarchivierung“ für Daten (niedrigschwelliger Weg ins Archiv)