Implementation of the CoreTrustSeal

The CoreTrustSeal board hereby confirms that the Trusted Digital repository GESIS Data Archive for the Social Sciences complies with the guidelines version 2017-2019 set by the CoreTrustSeal Board.
The afore-mentioned repository has therefore acquired the CoreTrustSeal on September 15, 2017.

The Trusted Digital repository is allowed to place an image of the CoreTrustSeal logo corresponding to the guidelines version date on their website. This image must link to this file which is hosted on the CoreTrustSeal website.

Yours sincerely,

The CoreTrustSeal Board

Assessment Information

Guidelines Version: 2017-2019 | November 10, 2016
Guidelines Information Booklet: CTS Requirements 2017-2019
All Guidelines Documentation: Documentation

Repository: GESIS Data Archive for the Social Sciences
Seal Acquiry Date: Sep. 15, 2017

For the latest version of the awarded   https://www.coretrustseal.org/why-certification/certified-repositories/
CoreTrusSeal for this repository :

Previously Acquired Seals: Seal date: May 8, 2014
                                           Guidelines version: 2014-2017 | July 19, 2013

This repository is owned by: **GESIS Data Archive for the Social Sciences**
                                           Unter Sachsenhausen 6-8
                                           50667 Cologne
                                           Germany

                                           T +49 (0)221-47694-423
                                           E Natascha.Schumann@gesis.org
                                           W http://www.gesis.org/

Assessment

0. Context

Applicant Entry

Self-assessment statement:

Repository Type: **Domain or subject based repository**

Level of Curation Performed:

B. Basic curation – e.g., brief checking, addition of basic metadata or documentation
C. Enhanced curation – e.g., conversion to new formats, enhancement of documentation
D. Data-level curation – as in C above, but with additional editing of deposited data for accuracy

According to its statutes, the core task of GESIS – Leibniz Institute for the Social Sciences (1) is to promote social science research. With a workforce numbering about 300 people, GESIS provides essential research-based services of national and international importance.
The Data Archive for the Social Sciences was first established in 1960 at Cologne University as the Central Archive for Empirical Social Research (ZA).
GESIS is divided in five scientific departments covering with their research based services the whole range of empirical social research (2):

• Survey Design and Methodology

• Monitoring Society and Social Change

• Data Archive for the Social Sciences

- Computational Social Science

- Knowledge technologies for the Social Sciences

The application covers the standard archival and data services provided by the Data Archive (3). This includes a centralized infrastructure for registering, curating and archiving quantitative social science research data, and facilitates data searches through a central catalogue.

The data archive currently distinguishes between two main levels of curation, standard archiving and added value archiving. Standard archiving contains checks whether the delivered material is complete, correct, and in a suitable technical condition (e.g. readable, virus free, etc.). Further checks concerning the plausibility, consistency, weighting and data protection are carried out. Besides adding a couple of administrative variables like study number, version number and date of version, data are only changed in case of errors or inconsistencies. These alterations are communicated with the respective data depositors. An example for our standard archiving service can be found here (4).

Added value archiving is carried out for selected surveys, like the International Social Survey Programme ISSP, the European Values Study EVS etc. These are typically large-scale, national or cross-national survey programs as well as longitudinal surveys. Data is changed or enhanced to a substantial extent. For example, variables are harmonized and standardized in order to facilitate comparisons across time or regional units. Beyond that, extensive metadata on variable level is produced and methodological reports and variable reports are created to facilitate usage of the data. An example for our added value archiving service can be found here (5).

Additionally GESIS offers also a service for sharing social science research data, even for smaller research projects. The service is called datorium (6) and is not part of this application, because it garantuees only bitstream preservation for a minimum of 10 years.

The Data Archive holds data of more than 5900 studies (2016). About 63.000 data sets were distributed in 2016 to users from more than 100 countries.

The Data Archive for the Social Sciences consists of seven teams with different tasks which are working closely together:

- **Producer Relations and Outreach:** Pre-Ingest activities, communication and training

- **National Surveys:** Preparation of documents and publication and user support of German long-term reference studies (added value)

- **International Surveys:** Preparation, international integration, documentation and user support regarding various international comparative data collection programs (added value)

- **Archive Instruments and Metadata Standards:** Preparation, development and use of social science standards for metadata and data

- **da|ra**: National center for registration of research data from social sciences and economics, assignment and management of DOI® in cooperation with DataCite and jointly with the Leibniz Information Center for Economics (ZBW)

- **Archive Operations**: Ingest, preservation, and provision of data to secondary users via the main dissemination and access channels

- **Data Linking and Data Security:** GESIS' Secure Data Center and development and offering of data linking services

Data predominantly originate from national and international comparative surveys from the social sciences. The studies are acquired, processed and documented on a needs-oriented basis, according to the preferences of the data depositor and availble resources, archived and made accessible to the scientifically interested public: These services are backed up by consultancy and training.

The Designated Community is constituted of researchers in empirical social research with a focus on the areas of sociology and political science as well as social science in its entirety. Other target groups include those working in related political, social and commercial social science environments. The services are offered to researchers (both in universities and non-university research institutions) and students.

Scope and nature of data collected by the Archive is described in our collection policy (7). Data must meet defined (minimum) criteria to be accepted by the Archive as it is suitable for social science research and allows reuse. Data has to be well-documented and survey instruments are available. The collection policy is currently under development and a new version will be published by the end of 2017.

Links (last accessed 03.08.2017):

(1) GESIS Website: http://www.gesis.org/en/en/home/
(2) GESIS Organigramm:
https://www.gesis.org/fileadmin/upload/institut/organigramm/GESIS%20Organigramm_en.pdf
(3) Data Archive Website: https://www.gesis.org/en/services/archiving-and-registering/data-archiving/
(4) Example study: People with Migration Background in Germany 2014:
https://dbk.gesis.org/dbksearch/sdesc2.asp?no=6604&search=bundespresseamt&search2=&field=all&field2=&db=e&tab=0&notabs=&n
(5) Example EVS:
https://www.gesis.org/en/services/data-analysis/survey-data/rdc-international-survey-programs/european-values-study/
(6) datorium: https://datorium.gesis.org
(7) Collection Policy (in German):
http://www.gesis.org/fileadmin/upload/institut/wiss_arbeitsbereiche/datenarchiv_analyse/2013-07-08_Was_wir_sammeln.pdf

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

1. Mission/Scope

Minimum Required Statement of Compliance:

0. N/A: Not Applicable.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

According to GESIS's statute, among the association's primary objectives is the "archiving, documentation, and long-term preservation of social sciences data, including the indexing of data as well as the high-quality enhancement of particularly relevant data to prepare them for re-use" (Statute § 2) (1). Thus, GESIS voices its commitment to preserve and provide access to social sciences research data in its by-laws and mission statement (2).

We formulated a preservation policy (3), which describes the main principles of digital preservation activities within our archive. The activities of the archive are documented on our website. All these activities that include documentation and processing as well as archival storage and access ensure that the data can be re-used and can be seen under this perspective as part of long-term preservation.

The Data Archive's preservation principles and practices have also been communicated in contributions to relevant publications.

GESIS is actively communicating its services and resources in numerous ways (exhibition stands at conferences, social media, publications, web pages, brochures etc.).
CESSDA Training (4), hosted by GESIS explicitly promotes research data management and data curation in the social sciences. It offers workshop and training events in data management for researchers and in digital preservation for archive and repository staff.
We have several (mostly internal) strategic and planning documents which outline how the mission statement is implemented. The annual report (5) covers research and research based services from all departments of GESIS. Program planning and program budget are the central strategic instruments. Program planning defines the strategic goals for five years and the program budget defines the focus for the next year. These documents are only for internal use. All member institutions of the Leibniz Assoziation use these instruments. Long-term planning with regard to GESIS' data infrastructure is described in our archiving strategy that is currently in a development phase.These documents can be handed over on request on a confidential basis.

Links (last accessed 03.08.2017):
(1) GESIS Statute (in German): http://www.gesis.org/das-institut/der-verein/satzung/

(2) GESIS Mission Statement: http://www.gesis.org/en/institute/the-association/mission/
(3) Preservation Policy:
http://www.gesis.org/fileadmin/upload/institut/wiss_arbeitsbereiche/datenarchiv_analyse/DAS_Preservation_Policy_eng_1.4.8.pdf
(4) CESSDA Training: https://www.cessda.eu/Research-Infrastructure/Training
(5) Annual reports (in German):
https://www.gesis.org/fileadmin/upload/institut/Jahresbericht%202016%20Web.pdf

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

2. Licenses

Minimum Required Statement of Compliance:

0. N/A: Not Applicable.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

GESIS's offers are primarily directed at researchers (both in universities and non-university research institutions) and students – in particular in empirical social research with a focus on the areas of sociology and political science as well as social science in its entirety. The Data Archive strongly promotes data sharing and re-use and hence seeks to make data available as openly and easily accessible as possible. However, both legal regulations and respect for the needs and requirements of data depositors make it necessary to manage access according to principles laid out in our usage regulations.

To enable the Data Archive to preserve and offer data for re-use, data depositors must sign an archive agreement (1) when submitting data for archiving. According to this agreement, the Data Archive may archive all data and documentation and process them further for the purpose of long-term preservation and re-use. The archive receives all necessary rights of (non-exclusive) use as laid down in German copyright law (especially §§16 and 19 UrhG). Thus the Data Archive receives permission from the data producers to carry out long-term preservation actions, e.g. migration to a different file format, as well as making several copies of the data and their documentation for backup and distribution.
In the archive agreement data depositors also determine under which standard licensing conditions data will be made available to data users. Typically data producers choose to make data and documents available for scientific analysis carried out in academic research and teaching.

All data users must agree to the usage regulations (2). The Data Archive provides guidance on how to use and cite data obtained through GESIS (3). These guidelines are available on the website.
The data deposited with the Data Archive are distributed through various channels. To facilitate access to the data, all data sets without special access restrictions are available for download via different platforms. Users are required to register once (by doing so accepting the terms of use/usage regulations) (4) and log in before downloading a data set. These data sets are available free-of-charge and at any time convenient to users.

Different access categories determined by the data producer and fixed in the archiving contract define the way data can be accessed and used. The conditions are outlined in the usage regulations as well as in the charge regulations (5).
The usage regulations also contain general access conditions. Users have to agree that they will inform the Data Archive when their project is completed and quote all used documents according to scientific conventions.

The use of metadata from GESIS' data catalogue DBK is possible under a Creative Commons license (6). All metadata are available free of restriction under the Creative Commons CC0 1.0 Universal Public Domain Dedication. However, GESIS requests that users actively acknowledge and give attribution to all metadata sources, such as the data providers and any data aggregators, including GESIS.

Individual-level data in social science research concerns behavior, opinions, attitudes, as well as social and economic living conditions of individuals. These contents are all subject to codes of conduct for the protection of the participants in research projects. Thus, personality rights, e.g. the right to informational self-determination or the right to privacy, have to be considered.

Individual-level data is especially protected by the Federal Data Protection Act (7): "The purpose of this Act is to protect the individual against his/her right to privacy being impaired through the handling of his/her personal data" (BDSG §1, 1). For archiving purposes, the legal framework under which the data was collected must allow for data archiving and the individual's right to privacy has to be protected. Usually data has to be anonymised (unless no informed consent exists which allows for transfer of personal information to the archive). If checks during the ingest phase show, that data is not anonymised, data depositors are asked to catch up on. Otherwise data is not accepted.

GESIS offers information of data protection and anonymisation (10,11,12).

The work of GESIS is based on a code of conduct and is explicitly bound to the rules of good scientific practice (8). An ombudsperson (9) is in place who can be involved in the case of discrepancies concerning codes of conduct. The Ombudsperson shall neither be a GESIS employee nor a member of a supervisory body and shall advise on and investigate discrepancies, suspicion(s) and disputes relating to good scientific practice. If the Ombudsperson receives an allegation of scientific misconduct, he or she shall conduct a preliminary inquiry independently and without delay and shall inform the person who made the allegation, the President, and the Chair of the Board of Trustees of the results of this preliminary inquiry. In the case of concrete suspicions, the facts on which the suspicions are based shall be determined immediately. The investigation shall be conducted or initiated by the Ombudsperson.

There are a range of different consequences for non-compliance to these codes:

• Consequences under labor and employment law

- Academic consequence

- Consequences under civil law

- Consequences under criminal law

- Scientific papers that contain factual errors due to proven scientific misconduct shall be withdrawn if they have not yet been published or shall be rectified if they have already been published (retraction).

Access to confidential data is only given in our Secure Data Center. It provides controlled and secure access to data deserving special protection.

Links (last accessed 03.08.2017):
(1) Archive agreement:
https://www.gesis.org/fileadmin/upload/institut/wiss_arbeitsbereiche/datenarchiv_analyse/Archivierungsvertrag_GESIS_Datenarchiv_v9
(2) Terms of use/ usage regulations (incl. access conditions):
https://www.gesis.org/fileadmin/upload/dienstleistung/daten/umfragedaten/_bgordnung_bestellen/Usage_regulations.pdf

(3) Bibliographic citation of research data and study related documents:
https://www.gesis.org/en/services/data-analysis/data-archive-service/citation-of-research-data/

(4) Registration: https://dbk.gesis.org/dbksearch/register.asp
(5) Charge regulations:
https://www.gesis.org/fileadmin/upload/dienstleistung/daten/umfragedaten/_bgordnung_bestellen/Charges.pdf

(6) Usage guidelines for GESIS DBK Metadata: https://dbk.gesis.org/dbksearch/guidelines.asp?db=e
(7) Federal Data Protection Act: https://www.gesetze-im-internet.de/englisch_bdsg/index.html
(8) Guidelines of good scientific
practice: http://www.gesis.org/fileadmin/upload/institut/leitbild/Gute_Praxis_GESIS_engl.pdf

(9) Ombudsperson: https://www.gesis.org/en/institute/the-association/

(10) Watteler, Oliver (2010). (in German): Datenschutz und die Archivierung von Daten in der qualitativen empirischen Sozialforschung. In: Medjedovic, Irena; Witzel, Andreas; Mauer, Reiner (Mitarb.); Watteler, Oliver (Hrsg.): Wiederverwendung qualitativer Daten: Archivierung und Sekundärnutzung qualitativer Interviewtranskripte, Wiesbaden: VS Verl. für Sozialwiss., S. 55-94: http://www.springer.com/springer+vs/soziologie/book/978-3-531-15571-5

(11) Recommendations for anonymising quantitative research data (in German): http://www.gesis.org/fileadmin/upload/institut/wiss_arbeitsbereiche/datenarchiv_analyse/Anonymisierung_quantitiativer_Daten-0150512

(12) Kinder-Kurlanda, K. & O. Watteler, 2015: Hinweise zum Datenschutz. Rechtlicher Rahmen und Maßnahmen zur datenschutzgerechten Archivierung sozialwissenschaftlicher Forschungsdaten. GESIS Papers 2015|01: http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis_reihen/gesis_papers/GESIS-Papers_2015-01.pdf

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

3. Continuity of access

Minimum Required Statement of Compliance:

0. N/A: Not Applicable.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

GESIS – Leibniz Institute for the Social Sciences is legally registered as a non-profit association and is sponsored jointly by the federal government and the federal states according to Article 91b of the German Federal Constitution (1). Funding in accordance with Article 91b is assigned in seven-year cycles. As a member of the Leibniz Association (2), the umbrella organization to currently 91 research institutions which "conduct research and provide infrastructure for science and research and perform research-based services – liaison, consultation, transfer – for the public, policy-makers, academia and business", GESIS is part of a strong network of publicly funded research institutions. In addition, GESIS has long-lasting and strong ties with universities. GESIS has three partner universities (GESIS' president and heads of departments are professors at these universities) and 60 German universities are members of GESIS e.V.

All holdings curated as part of our standard and added value archiving services are preserved for the long-term, i.e. in perpetuity.

Currently no formal succession plan is in place. Even though, GESIS as the largest infrastructure for the social sciences in Germany operates under a relatively stable financial framework, this issue needs to be addressed in the future.
Before funding through the Leibniz Association can be ceased a multi-staged procedure, starting with an evaluation of a Leibniz Institute, has to be passed. In case that after this procedure and subsequent negotiations an end of funding is recommended the financial support will end within the budget year. After that an interim funding for a maximum of three years (with 100% for the first two years and if no other decision is made for the third year, too) is provided. Thus, in case of a withdrawal of funding the Data Archive would be able to use that period to organise the transfer of its holdings to another appropriate institution. Due to the Archive's systematic approach to archive and preserve research data, a transfer of data holdings back to data owners or to another institution taking over responsibility is – at least in principle - possible any time.

Links (last accessed 03.08.2017):
(1) Annual Report 2016 (in German):
https://www.gesis.org/fileadmin/upload/institut/Jahresbericht%202016%20Web.pdf

(2) Leibniz Association: https://www.leibniz-gemeinschaft.de/en/home/

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

4. Confidentiality/Ethics

Minimum Required Statement of Compliance:

> 0. N/A: Not Applicable.

Applicant Entry

Statement of Compliance:

> 4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

> Individual-level data is especially protected by the Federal Data Protection Act (1): "The purpose of this Act is to protect the individual against his/her right to privacy being impaired through the handling of his/her personal data" (BDSG §1, 1). For archiving purposes, the legal framework under which the data was collected must allow for data archiving and the individual's right to privacy has to be protected. Usually data has to be anonymised (unless no informed consent exists which allows for transfer of personal information to the archive). If checks during the ingest phase show, that data is not anonymised, data depositors are asked to catch up on. Otherwise data is not accepted.

> Studies deposited in the archive which contain confidential data are either anonymised or access is only given through our Secure Data Center (2), which provides controlled and secure access to data deserving special protection. Data depositors and producers are advised on dealing with disclosive data and anonymisation. Trained staff is available who can give advice on handling and processing of sensitive data as well as on accessing it through the Secure Data Center (SDC).

> Checks for disclosure risks take place during the ingest phase on a regular base, e.g if there are still real names in the data or if data granularity, e.g. with respect to geographic coverage or occupational classification would allow for re-anonymising individuals. If problems are detected, data producers will be informed and advised, if and how data can be made accessible in a secure way. Within the Data Archive two specially trained staff members are responsible for the management of data with disclosure risks. Furthermore GESIS has an external data protection officer who is regularly involved in all relevant issues. Additionally, GESIS consults lawyers for advice and expert opinion in cases where further legal advice is required. The Data Archive has published recommendations for data anonymisation and refers to other publications 3, 4, 5, 6).

> Access to confidential data is only given through our Secure Data Center (SDC). It provides controlled and secure access to data deserving special protection. Data protection legislation requires that the possibility of re-identifying individuals in data provided by GESIS must be avoided. The SDC offers restricted access to data which has not been fully anonymised. It uses special contracts (7, 8) and user guidelines (9) which are published on the SDC website. The SDC applies special procedures to manage data with disclosure risk, such as carefully vetting researchers, requiring the signing of an agreement and applying various measures of organisational and technical control to protect data (10). The service comprises possibilities for off-site access (after signing a usage agreement (8) and submitting a data protection plan) and on-site access at GESIS' safe room. At the safe room researchers

work in an enclosed, offline workspace from which they cannot download any data. Analysis results are only transmitted after an output control has been performed. In the safe room researchers are not allowed to use mobile phones or laptops.

The user contract for the use of confidential data contains different measures in case of non-compliance: In case of misuse the user has to delete all data and supplementary material. In addition, a report will be send to other data service centers as well as to the German Data Forum (RatSWD) (11). It is also possible that the data user's account is blocked temporarily or permanently. Beyond that, users agree to the obligation to pay a fine of €10,000 (Euro), in the event of willful, deliberate or grossly negligent breach of contractual obligations.

Currently it is not yet clear if and how the new European Data Protection Regulation (12) will affect our work. Article 89 "Safeguards and derogations relating to processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes" paragraph 2 allows national exceptions.

Links (last accessed 03.08.2017):

(1) Federal Data Protection Act: https://www.gesetze-im-internet.de/englisch_bdsg/index.html

(2) GESIS – Secure Data Center: http://www.gesis.org/en/services/data-analysis/data-archive-service/secure-data-center-sdc/

(3) Recommendations for data anonymisation (in German): http://gesis.org/fileadmin/upload/institut/wiss_arbeitsbereiche/datenarchiv_analyse/Anonymisierung_quantitiativer_Daten-0150

(4) Watteler, Oliver (2010). (in German): Datenschutz und die Archivierung von Daten in der qualitativen empirischen Sozialforschung. In: Medjedovic, Irena; Witzel, Andreas; Mauer, Reiner (Mitarb.); Watteler, Oliver (Hrsg.): Wiederverwendung qualitativer Daten: Archivierung und Sekundärnutzung qualitativer Interviewtranskripte, Wiesbaden: VS Verl. für Sozialwiss., S. 55-94: http://www.springer.com/springer+vs/soziologie/book/978-3-531-15571-5

(5) Recommendations for anonymising quantitative research data (in German): http://www.gesis.org/fileadmin/upload/institut/wiss_arbeitsbereiche/datenarchiv_analyse/Anonymisierung_quantitiativer_Daten

(6) Kinder-Kurlanda, K. & O. Watteler, 2015: Hinweise zum Datenschutz. Rechtlicher Rahmen und Maßnahmen zur datenschutzgerechten Archivierung sozialwissenschaftlicher Forschungsdaten. GESIS Papers 2015|01, Verfügbar unter: http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis_reihen/gesis_papers/GESIS-Papers_2015-01.pdf

(7) Contract for on-site use: https://www.gesis.org/fileadmin/upload/dienstleistung/daten/secure_data_center/GESIS_Data_Use_Agreement_SDC_On-Site.pdf

(8) Contract for off-site use:
https://www.gesis.org/fileadmin/upload/dienstleistung/daten/secure_data_center/GESIS_Data_Use_Agreement_Off-Site.pdf

(9) Technical and organizational
measures: http://www.gesis.org/fileadmin/upload/dienstleistung/daten/secure_data_center/Guidelines_for_the_Secure_Data_Center_Safe
(10) Security requirements for off-site access:
https://www.gesis.org/fileadmin/upload/dienstleistung/daten/secure_data_center/GESIS_Leaflet_Secure_Data_Handling.pdf
(11) German Data Forum: https://www.ratswd.de/en
(12) REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April
2016: http://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&from=EN

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

5. Organizational infrastructure

Minimum Required Statement of Compliance:

0. N/A: Not Applicable.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

GESIS – Leibniz Institute for the Social Sciences is legally registered as a non-profit association and is sponsored jointly by the federal government and the federal states according to Article 91b of the German Federal Constitution. Funding in accordance with Article 91b is assigned in seven-year cycles. As a member of the Leibniz Association (1), the umbrella organization to currently 91 research institutions which "conduct research and provide infrastructure for science and research and perform research-based services – liaison, consultation, transfer – for the public, policy-makers, academia and business", GESIS is part of a strong network of publicly funded research institutions. In addition, GESIS has long-lasting and strong ties with universities. GESIS has three partner universities (GESIS' president and heads of departments are professors at these universities) and 60 German universities are members of GESIS e.V. (2)

Founded in 1960 as one of the first archives for social science data worldwide, the GESIS Data Archive for the Social Sciences (DAS) looks back on a history of over 50 years of curating, preserving, and disseminating data.

Today the DAS is a department of GESIS–Leibniz Institute for the Social Sciences, the biggest social science infrastructure institution in Germany. In addition to carrying out social science research projects of its own, GESIS offers support services throughout the complete data lifecycle (3): from initial research and study planning to data collection and analysis to data registration and archiving.

GESIS is also a service provider for CESSDA (4), the Consortium of European Social Science Data Archives, an umbrella organization dedicated to fostering cooperation and the creation of synergies between the contributing archives. Employees of the Data Archive are affiliated to several national and international organisations and initiatives, e.g. German Data Forum (RatSWD), DataCite, International Federation of Data Organizations for Social Sciences (IFDO) (5), CESSDA, European Social Survey (ESS) (6), and European Value Study (EVS) (7) etc.

Most of the currently about 70 staff members in the archive hold degrees in social sciences and thus have an

excellent understanding of the data they are curating. In addition, some members of the Data Archive hold degrees in library and information sciences. Staff members regularly take part in internal and external trainings on data management, metadata, long-term preservation and other relevant fields. Continuing education and training in different fields is part of GESIS' human resource development (8) and offered to all staff members.

Links (last accessed 03.08.2017):
(1) Leibniz Association: https://www.leibniz-gemeinschaft.de/en/home/
(2) General Assembly: https://www.gesis.org/en/institute/the-association/general-meeting-of-members/
(3) GESIS Services: https://www.gesis.org/en/services/
(4) CESSDA: http://www.cessda.net
(5) IFDO: http://ifdo.org/wordpress/
(6) ESS: http://www.europeansocialsurvey.org/
(7) EVS: http://www.europeanvaluesstudy.eu/
(8) GESIS Human resources development: http://www.gesis.org/en/institute/career/human-resources-development/

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

6. Expert guidance

Minimum Required Statement of Compliance:

0. N/A: Not Applicable.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

Services offered by the Archive and the corresponding underlying infrastructure are constantly evaluated either as part of internal procedures or through external evaluation.
GESIS offers are always at the forefront through our own research in the areas of survey methodology, attitudes and behavior and structural research as well as in applied computer science. Working together with universities and other partners enables us to maintain close contact with social science researchers. Internal procedures contain e.g. strategic planning, portfolio analysis, reports to and evaluations by the scientific advisory board (1) and the user advisory board (2). Both boards are regularly invited to evaluate existing services and discuss future developments.

Next to these internal measures GESIS – like any other Leibniz institute – is evaluated by the Leibniz Senate every seven years at the latest (3). In particular, the members of the review board evaluate the quality of work carried out in science and research, consultancy and services as well as in other specific fields of activity. They also examine to what extent the Leibniz institution has produced a convincing strategy for combining and developing the individual strands. In addition, attention is directed to cooperation with other institutions, such as neighbouring universities, and international visibility; the transfer of results to other sections of society; the promotion of junior researchers and efforts to achieve gender equality. The institution's performance in terms of quality assurance is also assessed.

Furthermore GESIS has an external adviser who supports strategic planning and an external data protection officer who is regularly involved in all relevant issues. Additionally, GESIS consults lawyers for advice and expert opinion in cases where further legal advice is required.

In an internal strategic planning document the data archive has determined its strategic goals and work planning for a period of five years. One of the tasks identified in this document is to establish a yearly "trend report research data" to gain a systematic overview about relevant developments in the eco-system of the Archive. One important aspect with regard to this is new types of data that might be of interest for our designated communities and thus for the archive.

GESIS is involved in numerous national and international initiatives and projects dealing with topics such as persistent identifiers (e.g. DataCite (4)), metadata standards (e.g. DDI (5)), and building of data infrastructures (e.g. CESSDA (6)) etc. to continuously advance its services and to contribute to future developments in relevant areas. All these activities contain promotional aspects as well.

Communication with users takes place in several ways. GESIS carries out user surveys, e.g. in 2016 researchers from our designated community were asked about the different services offered by GESIS. The results will be published as a GESIS Paper in summer 2017 (7). Furthermore GESIS started an Online Access Panel for Interactive Information Retrieval Research (IIIRpanel) (8) which allows our users to participate in studies in which we will ask for their opinion on new ideas or concepts. The goal is to align our digital portals with user demands.

Links (last accessed 03.08.2017):
(1) Scientific Advisory Board: https://www.gesis.org/en/institute/the-association/scientific-advisory-board/
(2) User Advisory Board: https://www.gesis.org/en/institute/the-association/user-advisory-board/
(3) Evaluation of Leibniz institutes: http://www.leibniz-gemeinschaft.de/en/about-us/evaluation/
(4) DataCite: https://www.datacite.org/
(5) DDI: https://www.ddialliance.org/
(6) CESSDA: https://cessda.net/
(7) GESIS Papers: http://www.gesis.org/en/services/publications/gesis-papers/
(8) IIRpanel: https://multiweb.gesis.org/iirpanel/Home.php

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

7. Data integrity and authenticity

Minimum Required Statement of Compliance:

0. N/A: Not Applicable.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

After a submission has been received, the data and all accompanying materials are assessed with regard to content, structure and format. After this initial check for completeness on the basis of an internal checklist and technical control, all files are transferred to the archival storage in their original versions and formats – this SIP (Submission Information Package) will not be altered anymore and will be retained in its original form. Information about depositor(s), archive agreement, deposition (data, responsible staff, and composition of the SIP etc.) is compiled and documented. Corresponding checksums are produced and stored. Subsequently, the data are converted to or saved in archival formats and undergo further checks and – where necessary – corrections. Any corrections carried out at this stage are documented in a way that allows for reverting to the original version at any time. All significant corrections/changes to the data will be discussed with data depositors beforehand. All changes are documented in syntax/ setup or further documentation files. These always contain additional information about when and why changes were made and by whom. Handling of different versions is supported by a versioning policy (1).

Identity of depositors is not checked in a formalized way but as a rule the staff has personal communication with every depositor (e-mail, telephone or face to face). Data depositors need to sign an archive agreement (2). Thus, contact information of the depositors is available. In many cases longstanding relationships exist.

The integrity of data is monitored with the help of checksums among other measures. All checksumming is done using sha 256. Automated scripts generate and compare continuously checksums of all objects in the central archival storage. Output is logged and checked by staff on a weekly basis.

The central metadata management system (DBKEdit) is accessible to authorized staff only and all activities and changes are logged.

Each study receives a version number. Versioning of data is governed by a versioning guideline that is strictly adhered to in order to meet the requirements of DOI-assignment and the standards of trusted digital preservation.

The version number is a three-digit number (major.minor.revision). The major number is incremented when there are changes in the composition of the data set (e.g. additional variables or cases), the minor or second number is incremented when significant errors have been fixed (e.g. coding errors, misleading value labels), and the third or revision number is incremented when minor bugs are fixed (e.g. spelling errors in variable or value labels). The version number is included as a variable in the data set and added to file names of corresponding objects in accordance with a set of naming conventions. The version history, indicating the major/minor changes made to the data, is documented in the metadata management system and made available to end users through our data catalogue (DBK). All data set versions – from the original to the latest version – are kept in the archive. Syntax/ setup files documenting the changes between the different versions are kept in addition.

Links (last accessed 03.08.2017):
(1) Versioning guidelines: https://www.gesis.org/en/services/archiving-and-registering/data-archiving/ingest/
(2) Archive agreement:
https://www.gesis.org/fileadmin/upload/institut/wiss_arbeitsbereiche/datenarchiv_analyse/Archivierungsvertrag_GESIS_Datenarchiv_v9

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

8. Appraisal

Minimum Required Statement of Compliance:

0. N/A: Not Applicable.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

The GESIS Data Archive is specialized on quantitative survey data, but holds other types (e.g. aggregate/ time series data or textual materials) as well. Our collection policy (1) defines the criteria a study should meet to be admitted into the Data Archive. Since other types of data (e.g. transactional data, social media data, and experimental data (2,3)) are becoming more and more important the archive evaluates opportunities and challenges with regard to archiving and re-use and is also engaged in corresponding projects.

For each submission, archive staff checks whether the delivered material is complete, correct, and in a suitable technical condition (e.g. readable, virus free, etc.). Further checks concerning the plausibility, consistency, data weighting and data protection are carried out. This ingest control is carried out based on an internal checklist which at same time is used for documentation purposes.

Most of the metadata are created by archive staff on the basis of the documentation delivered by the depositors. Data depositors are asked to fill in an off-line study description template (DBKForm) (4). It requests information such as "Title", "Alternative Title", "Date of data collection" and some others. If the metadata are not sufficient, an attempt is made to extract it from publications. Otherwise data depositors are asked to deliver additional information.
The metadata scheme used by the Data Archive is compliant with the DDI Standard (5), as well as with the da|ra and DataCite metadata schemes. Resource discovery metadata is available for download in DDI2 and DDI3 format under a CC0 license. Datasets of particular importance and important study collections falling into the Data Archive's core areas of collection are processed (cumulated, harmonized, standardized), documented, and enhanced in much greater depth – not only on study level, but on the level of individual questions and variables. This metadata is used for different purposes (e.g. production of codebooks/ variable reports; long-term preservation) and is made available online via portals like ZACAT (6). The deployed tools (e.g. Dataset Documentation Manager, CodebookExplorer) produce DDI compliant metadata as well.

Further structural and administrative metadata is created for internal use. Among others, this provides relevant technical and provenance information. Detailed information on the metadata scheme was published in Zenk-Möltgen 2012 (7).
Data are usually delivered in formats that are well-established within the social science community, in particular statistical formats like SPSS or Stata. After a submission has been received, the data and all accompanying

material are assessed with regard to content, structure and format. Incoming objects in non-compliant formats will be converted to defined archival formats suitable for long-term preservation. If that is not possible studies are not accepted for archiving or, exceptionally, bit-level preservation is offered. However, to provide users with more comprehensive information on this matter we have published a list of preferred formats (8).

Links (last accessed 03.08.2017):
(1) Collection Policy (in German):
http://www.gesis.org/fileadmin/upload/institut/wiss_arbeitsbereiche/datenarchiv_analyse/2013-07-08_Was_wir_sammeln.pdf
(2) Project x-Hub: https://www.gesis.org/forschung/drittmittelprojekte/projektuebersicht-drittmittel/x-hub-ii/
(3) Project GeorefUm: https://www.gesis.org/forschung/drittmittelprojekte/projektuebersicht-drittmittel/georefum/

(4) DBKForm: https://dbk.gesis.org/dbkform/indexen.htm
(5) DDI Codebook: http://www.ddialliance.org/Specification/DDI-Codebook/2.5/
(6) ZACAT: http://zacat.gesis.org/webview/
(7) Zenk-Möltgen, Wolfgang; Habbel, Norma (2012): Der GESIS Datenbestandskatalog und sein Metadatenschema, Version 1.8. GESIS Technical Reports, 2012/01. (in German):
https://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis_reihen/gesis_methodenberichte/2012/TechnicalReport_2012-01.pdf
(8) Recommended formats:
https://www.gesis.org/en/services/archiving-and-registering/data-archiving/preparing-data-for-submission/

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

9. Documented storage procedures

Minimum Required Statement of Compliance:

0. N/A: Not Applicable.

Applicant Entry

Statement of Compliance:

3. In progress: We are in the implementation phase.

Self-assessment statement:

Empirical research projects producing research data generally go through a process of multiple phases – the research data cycle. Each individual phase of this cycle requires specific know-how to obtain significant results. GESIS has the know-how and optimizes utilization of this by offering a unique package of services to accompany the entire research data cycle. Archiving is well integrated into this data life cycle. Workflows within the archive are organised according to an archival life cycle, ranging from acquisition/ pre-ingest to dissemination of data (1). The central functions of the OAIS reference model can be mapped to the existing structure of the archive. For most parts of the corresponding workflows, procedures, standards and rules are in place (2,3,4). Even though internal documentation exists it currently is not complete and up to date for every activity. Significant steps to improve this situation were already taken, but there is still some work to be done. Existing internal documentation is available to all staff members through a wiki.

The main principles of long-term preservation as carried out by the GESIS Data Archive are documented in our preservation policy (5).
The security and risk management is carried out in close co-operation with GESIS's IT department, which administers the servers and takes care of backups, media monitoring and refreshing.

To protect the data, the following backup and access control procedures are in place to guarantee the (physical) safety of the digital archive holdings:
1) Physical protection measures:
a. The computing centers and server rooms are secured against unauthorized access by means of an electronic access control system.
b. Smoke and water detectors are in place, temperatures in the computing center are monitored.
2) Redundant data storage in different locations (Cologne and Mannheim):
a. Frequent (up to daily) incremental and complete back-ups to onsite disk and tape libraries (tapes stored in suitable vault). In addition, frequent backups to offsite tape libraries.
3) Diversity of storage media (hard disk, tape) and frequent media refreshment.
a. The backup and storage procedures (redundant and distributed storage) in place allow for fast and complete recovery / restoral of the archive holdings in case of a disaster.

In addition to the backup procedures described above, the archive has the following technological and

organizational measures in place to assure that the data bitstream is securely stored and cannot be altered without authorization: Write access to the archive server is highly restricted and governed by a set of strict rules and regulations. Only two members of the Data Archive staff are authorized to add, delete or change files on the archive server. All changes to the archival storage are logged and compiled in a weekly report. Special events (e.g. deletion of objects, movement of objects to unusual locations) must be commented by the originator and verified by a supervisor. The transfer of files into the archive takes place by means of a special transfer folder in the network, from which data and documentation to be archived are picked up, checked once again for conformity with the Data Archive's preservation standards (file formats, naming conventions, etc.), and transferred onto the archive server by one of the two authorized staff members.

All source and archive files' checksums are stored to ensure data consistency. All checksumming is done using sha 256.

All transformations made to data are documented. All significant corrections/changes of the data will be discussed with data depositors beforehand. General information about handling of the data is given on our website and as well during pre-ingest communication with depositors. Standard procedures for changes applied to the data during ingest or at later stages are available (e.g. naming conventions, handling of missing data, versioning rules).

Supported by a constant monitoring of technology (storage technology and media, software and file formats) as well as a normalization of file formats during ingest, the Data Archive pursues a migration strategy to ensure long-term access to its holdings. Data and documentation are archived in well-defined, standardized file formats to ensure that efficient migration strategies can be developed when this becomes necessary. Syntax/ setup files documenting the changes between different versions are kept in addition.

While the refreshing of storage media takes place continuously, format migrations are undertaken only if the readability / interpretability of archive holdings is endangered by technological obsolescence, if they cannot be processed and used anymore in a state-of-the-art manner, and/or if a format migration brings considerable advantages with regard to user-friendliness and the work of the archive.

During format migrations utmost care is taken not to alter the significant properties of the archival objects. Being domain experts, all central aspects of social science data (e.g. characteristics of data matrices and variable definitions) and the special features of the different file formats are well-known to archive staff. Migration procedures are documented thoroughly in order not to compromise the archival objects' authenticity during the migration process.

Links (last accessed 03.08.2017):
(1) Data Archive Webpage: https://www.gesis.org/en/services/archiving-and-registering/data-archiving/

(2) Data Service Infrastructure for the Social Sciences and Humanities. "Report about Preservation Service Offers", Deliverable: D4.2, p 47-70:
http://dasish.eu/publications/projectreports/D4.2_-_Report_about_Preservation_Service_Offers.pdf
(3) Mauer, Reiner (2012): Das GESIS Datenarchiv für Sozialwissenschaften. In: Altenhöner, Reinhard; Oellers, Claudia (Hrsg.): Langzeitarchivierung von Forschungsdaten. Standards und disziplinspezifische Lösungen, Berlin: Scivero, S. 197-215, http://ratswd.de/dl/downloads/langzeitarchivierung_von_forschungsdaten.pdf
(4) Recker, Jonas A., and Stefan Müller. 2015. "Preserving the Essence: Identifying the Significant Properties of Social Science Research Data." New Review of Information Networking 20 (1-2): 231-237. doi: http://dx.doi.org/10.1080/13614576.2015.1110404

(5) Preservation Policy:
http://www.gesis.org/fileadmin/upload/institut/wiss_arbeitsbereiche/datenarchiv_analyse/DAS_Preservation_Policy_eng_1.4.8.pdf

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

10. Preservation plan

Minimum Required Statement of Compliance:

0. N/A: Not Applicable.

Applicant Entry

Statement of Compliance:

3. In progress: We are in the implementation phase.

Self-assessment statement:

A complete preservation plan is not yet formulated. But almost all activities undertaken by staff of the archive serve to preserve archive holdings. The main principles of long-term preservation as carried out by the GESIS Data Archive are documented in our preservation policy (1).

Information on depositing and archiving of social science research data is provided on the GESIS web pages (2) and described in publications (3). This includes, among others, a list of minimum requirements for a submission information package and a list of preferred formats. Remaining questions and details are clarified during the pre-ingest phase.

After submission, data and documentation are validated by experienced staff. Inconsistencies and errors are corrected and missing information is added in close cooperation with the depositor. This ensures that the data is complete, usable and interpretable which is an important prerequisite for all following preservation measures.

Generation of technical, administrative and descriptive metadata as well as indexing and the generation of further documentation material is based on national and international standards, classifications, thesauri and other controlled vocabularies where necessary and relevant.

Data is currently mostly archived in formats (SPSS and/or Stata) widely used by the designated community. Since this approach is not totally adequate with regard to long term preservation, we are currently in the process of migrating our holdings to CSV accompanied by setups for SPSS, Stata, MySQL and additionally we generate a structured generic text format. However, migrating data from nearly all different SPSS/Stata versions must be carried out with utmost caution. In case that accessibility and/or usability is no longer granted, format migrations will be carried out.

Through the archive agreement (4), the archive receives all necessary rights of (non-exclusive) use as laid down in German copyright law (especially §§16 and 19 UrhG). Thus the Data Archive receives permission from the data producers to carry out long-term preservation actions, e.g. migration to a different file format, as well as making

several copies of the data and their documentation for backup and distribution. In addition the data producer assures that legal obligations regarding personal or confidential data are adhered to.

Links (last accessed 03.08.2017):

(1) Preservation
Policy: https://www.gesis.org/fileadmin/upload/institut/wiss_arbeitsbereiche/datenarchiv_analyse/DAS_Preservation_Policy_eng_1.4.8.p
.

(2) Data Archive Website: http://www.gesis.org/en/services/archiving-and-registering/data-archiving/
(3) Schumann, Natascha; Mauer, Reiner: The GESIS Data Archive for the Social Sciences: A Widely Recognised
Data Archive on its Way. The International Journal of Digital Curation. Vol 8, No 2 (2013).
http://dx.doi.org/10.2218/ijdc.v8i2.285
(4) Archive
agreement: https://www.gesis.org/fileadmin/upload/institut/wiss_arbeitsbereiche/datenarchiv_analyse/Archivierungsvertrag_GESIS_Date

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

11. Data quality

Minimum Required Statement of Compliance:

0. N/A: Not Applicable.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

To ensure data quality the Data Archive has defined minimum requirements that the delivered data has to meet. It requests data producers to submit all materials necessary for a secondary analysis. This includes at least

- information about the primary researcher(s) and title of the study

- the data itself, prepared for direct use in statistical software packages if possible,

- the instrument or instruments used for data collection (e.g. questionnaire),

- a methodological description of the data collection and preparation procedures,

- publications or references to publications based on the respective data.

No materials are accepted that are subject to any copyright restrictions which may interfere with the use of the data as outlined in the archive agreement (e.g. copies of complete books).

After submission the data is checked according to its archiving level by domain experts who are able to assess quality and completeness of data and documentation.

GESIS makes broad information about a given study available to data consumers via web pages, data catalogues, variable reports and other publications so that secondary users can analyze the data and can assess its quality (1).

Most of the metadata are created by archive staff on the basis of the documentation delivered by the depositors. The metadata scheme (2) used by the Data Archive is compliant with the DDI Standard (3), as well as with the da|ra (4) and DataCite metadata schemes (5). Resource discovery metadata is available for download in DDI2 and DDI3 format under a CC0 license. Datasets of particular importance and important study collections falling into the Data Archive's core areas of collection are processed (cumulated, harmonized, standardized), documented, and enhanced in much greater depth – not only on study level, but on the level of individual questions and variables. This metadata is used for different purposes (e.g. production of codebooks/ variable reports; long-term preservation) and is made available online via portals like ZACAT (6). The deployed tools (e.g. Dataset Documentation Manager, CodebookExplorer (7)) produce DDI compliant metadata as well.

Automated validation is carried out for DDI metadata in XML format for different purposes like export from the data catalogue DBK and download, for harvesting via OAI-PMH or for archival storage.

Each study is assigned a Digital Object Identifier (DOI), a permanent, persistent identifier used for citing and linking electronic resources. GESIS not only assigns DOIs to its own holdings but as a member of the Data Cite Consortium and maintainer of the allocation agency da|ra also offers DOI services to a wider national and international community.

The user is obliged to quote all used documents according to scientific conventions (8).

Links (last accessed 03.08.2017):
(1) Terms of use/ usage regulations:
http://www.gesis.org/fileadmin/upload/dienstleistung/daten/umfragedaten/_bgordnung_bestellen/Usage_regulations.pdf

(2) Zenk-Möltgen, Wolfgang; Habbel, Norma (2012): Der GESIS Datenbestandskatalog und sein Metadatenschema, Version 1.8. GESIS Technical Reports, 2012/01. (in German):
http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis_reihen/gesis_methodenberichte/2012/TechnicalReport_2012-01.pdf
(3) DDI Codebook: http://www.ddialliance.org/Specification/DDI-Codebook/2.5/
(4) da|ra Metadata Schema: http://dx.doi.org/10.4232/10.mdsdoc.2.2.1
(5) DataCite Metadata Schema: http://schema.datacite.org/
(6) ZACAT: http://zacat.gesis.org/webview/
(7) Dataset Documentation Manager and CodebookExplorer: https://www.ddialliance.org/
(8) Bibliographic citation of research data and study related documents:
http://www.gesis.org/en/services/data-analysis/data-archive-service/citation-of-research-data/

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

12. Workflows

Minimum Required Statement of Compliance:

0. N/A: Not Applicable.

Applicant Entry

Statement of Compliance:

3. In progress: We are in the implementation phase.

Self-assessment statement:

Workflows within the archive are organised according to an archival life cycle, ranging from acquisition/ pre-ingest to dissemination of data. The central functions of the OAIS reference model can be mapped to the existing structure of the archive (1). For most parts of the corresponding workflows, procedures, standards and rules are in place. Even though internal documentation exists it currently is not complete and up to date for every activity. Significant steps to improve this situation were already taken, but there is still some work to be done. An overview about the different steps and processes (pre-ingest, ingest, study description, archiving, access) can be found on our website (2) and in publications (3).

All transformations made to data are documented. All significant corrections/changes of the data will be discussed with data depositors beforehand. General information about handling of the data is given on our website and as well during pre-ingest communication with depositors.

Standard procedures for changes applied to the data during ingest or at later stages are available (e.g. naming conventions, handling of missing data, versioning rules). Existing internal documentation is available to all staff members through a wiki.

Data may contain confidential information that may not be accessed either by the public or by staff that is not authorised. If not clarified before delivering data to the archive, employees who are trained in data protection issues have to decide at a very early stage, if a study contains confidential data that requires anonymisation or special protection. All further steps depend on this decision. To accommodate sensitive data, the existing workflows were expanded with additional measures. Access to confidential data is only given through our Secure Data Center (SDC). It provides controlled and secure access to data deserving special protection. Data protection legislation requires that the possibility of re-identifying individuals in data provided by GESIS must be avoided. The SDC offers restricted access to data which has not been fully anonymised.
For evaluating our processes we tested successfully the use of encrypted file containers. On the one hand it should be as secure as possible and on the other hand it should also be as simple in handling as possible. The results are summarised in a concept for secure data management that will be implemented in the course of 2017.

In cases where archiving with GESIS cannot be realized, we try to find alternative data centres (e.g. partners in the network of research data centres of the German Data Forum) and bring data producers into contact with them. If there is no alternative option, GESIS tries to act as (temporary) fallback option.

Links (last accessed 03.08.2017):
(1) Workflow:
https://www.gesis.org/fileadmin/upload/institut/wiss_arbeitsbereiche/datenarchiv_analyse/Workflow_Data_Archive.pdf

(2) Data Archive Website: http://www.gesis.org/en/services/archiving-and-registering/data-archiving/
(3) Schumann, Natascha; Mauer, Reiner: The GESIS Data Archive for the Social Sciences: A Widely Recognised Data Archive on its Way. The International Journal of Digital Curation. Vol 8, No 2 (2013). http://dx.doi.org/10.2218/ijdc.v8i2.285

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

13. Data discovery and identification

Minimum Required Statement of Compliance:

0. N/A: Not Applicable.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

The data catalogue (1) is the central access point to the holdings of the archive and comprises study descriptions for all archived studies, mainly including micro data from survey research and aggregate time series data. Further portals (2, 3, 4, 5, 6) allow for access to special holdings and/or offer particular additional services, e.g. the ZACAT (2)Online Study catalogue, which provides selected studies with extensive documentation on study and variable level, or HISTAT (3), a system giving access to historical time series data.

The metadata scheme used by the Data Archive is compliant with the DDI Standard (7), as well as with the da|ra (8) and DataCite metadata schemes (9).

In addition to the central systems for immediate download, users can order data that requires permission of the data depositors (access categories B and C (10)) from the archive's Data Service via a shopping cart system, by e-mail or telephone. They will receive this data (in customized form, if they wish) on a CD-ROM or DVD or via a secure download using Cryptshare. For this data service handling fees are charged (11).

Different tools to search, explore, and analyze studies (data sets and accompanying material) are in place (partly allowing searches down to the level of individual variables/ questions). Harvesting of metadata of our holdings published via ZACAT is in place as well as an OAI interface (OAI-PMH) (12) for our data catalogue (DBK). All metadata from GESIS DBK are available under the Creative Commons CC0 1.0 Universal Public Domain Dedication (13). However, GESIS requests that users actively acknowledge and give attribution to all metadata sources, such as the data providers and any data aggregators, including GESIS (14,15).

Each study is assigned a Digital Object Identifier (DOI) (16), a permanent, persistent identifier used for citing and linking electronic resources. GESIS not only assigns DOIs to its own holdings but as a member of the Data Cite Consortium and maintainer of the allocation agency da|ra also offers DOI services to a wider national and international community.

Links (last accessed 03.08.2017):

(1) Data Catalogue: https://dbk.gesis.org/dbksearch/index.asp?db=e

(2) ZACAT Online Study Catalogue: http://zacat.gesis.org/webview/
(3) Histat: https://www.gesis.org/angebot/daten-analysieren/daten-historischer-studien/datenbank-histat/
(4) sowiport: http://sowiport.gesis.org/
(5) European System of Social Indicators:
https://www.gesis.org/en/services/data-analysis/social-indicators/european-system-of-social-indicators/
(6) MISSY: http://www.gesis.org/en/missy/
(7) DDI Codebook: http://www.ddialliance.org/Specification/DDI-Codebook/2.5/

(8) da|ra: http://www.da-ra.de/en/home/
(9) DataCite Metadata Schema: https://schema.datacite.org/
(10) Access categories (Topic 3):
https://www.gesis.org/fileadmin/upload/dienstleistung/daten/umfragedaten/_bgordnung_bestellen/Usage_regulations.pdf

(11) Charges:
https://www.gesis.org/fileadmin/upload/dienstleistung/daten/umfragedaten/_bgordnung_bestellen/Charges.pdf

(12) Data Catalogue OAI-PMH: https://dbk.gesis.org/dbkoai/?verb=Identify
(13) Usage Guidelines for GESIS DBK Metadata: https://dbk.gesis.org/dbksearch/guidelines.asp?db=e
(14) Bibliographic citation of research data and study related documents:
https://www.gesis.org/en/services/data-analysis/data-archive-service/citation-of-research-data/
(15) Regulations for citation (in
German): http://www.gesis.org/fileadmin/upload/dienstleistung/daten/umfragedaten/_bgordnung_bestellen/Bibliographisches_Zitieren_vo

(16) DOI: http://www.doi.org/

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

14. Data reuse

Minimum Required Statement of Compliance:

0. N/A: Not Applicable.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

To make sure that social science research data can be reused by others it is not sufficient to archive just the data set. To understand the data additionally information is needed, e.g questionnaires, codebooks, method reports etc. Thus the Data Archive (1) requests data producers to submit all materials necessary for a secondary analysis. This includes at least

- information about the primary researcher(s) and title of the study

- the data itself, prepared for direct use in statistical software packages if possible,

- the instrument or instruments used for data collection (e.g. questionnaire),

- a methodological description of the data collection and preparation procedures,

- publications or references to publications based on the respective data (2,3).

Another important part of our approach is the creation of standardized metadata. Besides technical and administrative information, providing extensive context information is essential for future usability and interpretability of social science research data.

Most of the metadata are created by archive staff on the basis of the documentation delivered by the depositors. The metadata scheme used by the Data Archive is compliant with the DDI Standard, as well as with the da|ra and DataCite metadata schemes.

Data are usually delivered in formats that are well-established within the social science community, in particular statistical formats like SPSS or Stata. The Data Archive has published a list of recommended formats (4) on its website.

After a submission has been received, the data and all accompanying material are assessed with regard to content, structure and format. Incoming objects in non-compliant formats will be converted to defined archival formats suitable for long-term preservation. If that is not possible objects resp. studies are not accepted for archiving or, exceptionally, bit-level preservation is offered. However, to provide users with more comprehensive information on this matter we have published a list of preferred formats.

For each submission, archive staff checks whether the delivered material is complete, correct, and in a suitable technical condition (e.g. readable, virus free, etc.). Further checks concerning the plausibility, consistency, data weighting and data protection are carried out. This ingest control is carried out based on an internal checklist which at same time is used for documentation purposes.

Format transformations are currently done with a proprietary third-party tool (StatTransfer). However, to obtain more control over this process presently a tool is under developement. At the time being, it is limited to handling SPSS system files, but shall be expanded to cover more data formats (Stata coming next). It exports the table data to a tab separated file and the metadata (variable names, labels, format specifications, value labels...) to various formats including an in-house intermediary format, DDI-2.5, SPSS syntax files, and sql database definitions.

Undertaking extensive validation action during the ingest phase enables us to gain an exact overview about the data we have in the archive. We know the formats and can estimate the risks in terms of preservation. To avoid future obsolescence, we accept only those formats we know we are capable of preserving. Data submitted in other formats are converted during ingest into suitable ones. If data is in a format we cannot handle or convert, we direct data producers to a more suitable archive.

Supported by a constant monitoring of technology (storage technology and media, software and file formats) as well as a normalization of file formats during ingest, the Data Archive pursues a migration strategy to ensure long-term access to its holdings. Data and documentation are archived in well-defined, standardized file formats to ensure that efficient migration strategies can be developed when this becomes necessary. Syntax/ setup files documenting the changes between different versions are kept in addition.

While the refreshing of storage media takes place continuously, format migrations are undertaken only if the readability / interpretability of archive holdings is endangered by technological obsolescence, if they cannot be processed and used anymore in a state-of-the-art manner, and/or if a format migration brings considerable advantages with regard to user-friendliness and the work of the archive.

During format migrations utmost care is taken not to alter the significant properties of the archival objects. Being domain experts, all central aspects of social science data (e.g. characteristics of data matrices and variable definitions) and the special features of the different file formats are well-known to archive staff. However, we are currently approaching the topic of significant properties in a more systematic way with the objective of developing a detailed and testable definition (5).

Each migration is documented thoroughly in order not to compromise the archival objects' authenticity during the migration process.

Links (last accessed 03.08.2017):
(1) Data Archive Webpage, Ingest:
http://www.gesis.org/en/services/archiving-and-registering/data-archiving/ingest/
(2) Data Service Infrastructure for the Social Sciences and Humanities. "Report about Preservation Service Offers", Deliverable: D4.2, p 47-70:
http://dasish.eu/publications/projectreports/D4.2_-_Report_about_Preservation_Service_Offers.pdf
(3) Mauer, Reiner (2012): Das GESIS Datenarchiv für Sozialwissenschaften. In: Altenhöner, Reinhard; Oellers, Claudia (Hrsg.): Langzeitarchivierung von Forschungsdaten. Standards und disziplinspezifische Lösungen, Berlin: Scivero, S. 197-215, http://ratswd.de/dl/downloads/langzeitarchivierung_von_forschungsdaten.pdf

(4) Recommended formats:
http://www.gesis.org/en/services/archiving-and-registering/data-archiving/preparing-data-for-submission /
(5) Recker, Jonas A., and Stefan Müller. 2015. "Preserving the Essence: Identifying the Significant Properties of Social Science Research Data." New Review of Information Networking 20 (1-2): 231-237. doi: http://dx.doi.org/10.1080/13614576.2015.1110404

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

15. Technical infrastructure

Minimum Required Statement of Compliance:

0. N/A: Not Applicable.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

The GESIS data center was build up in 2011 based on criteria from the German Federal Office for Information Security (BSI) (1), M 1 IT-Infrastructure, concerning electricity supply, air conditioning, early fire detection, fire extinguishing system, access control system and intruder alarm system. The probability of occurrence of critical events was considered in the planning process. The whole data center is protected by an alarm system. In case of an alarm a special plan is in place. The accessibility of the data center is oriented at the Data center tiers.

All the servers and storage media are refreshed about every five years. Constant monitoring is carried out by our IT department based on reports from the Computer Emergency Response Teams (CERT) (2) from the Federal Office for Information Security and from German National Research and Education Network, DFN (3).

To gain a more structured overview of workflows and processes and to identify and close possible gaps a mapping between the Archive and the OAIS functional model, as well as an application of the concepts from the OAIS information model has been carried out for the most parts (4). The mapping is available on request. Further conceptual work is on the way and concentrates on relevant aspects for our archival work. Even though not all processes are carried out exactly in the way defined by the OAIS reference model, the GESIS Data Archive fulfils all responsibilities described in OAIS reference model.

Links (last accessed 03.08.2017):
(1) BSI:
https://www.bsi.bund.de/DE/Themen/ITGrundschutz/ITGrundschutzKataloge/Inhalt/_content/m/m01/m01.html
(2) CERT-Bund: https://www.bsi.bund.de/EN/Topics/IT-Crisis-Management/CERT-Bund/CERTBund.html

(3) DFN-CERT: https://www.dfn-cert.de/en.html
(4) Recker, Jonas A.; Schumann, Natascha. 2014: Evaluating and developing Ingest workflows with OAIS and PAIMAS at the GESIS Data Archive for the Social Sciences:
http://ist.publisher.ingentaconnect.com/content/ist/ac/2014/00002014/00000001/art00034;jsessionid=i55m2pm48n8q.victoria

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

16. Security

Minimum Required Statement of Compliance:

0. N/A: Not Applicable.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

The security and risk management is carried out in close co-operation with GESIS's IT department, which administers the servers and takes care of backups, media monitoring and refreshing. To protect the data, the following backup and access control procedures are in place to guarantee the (physical) safety of the digital archive holdings:
1) Physical protection measures:
a. The computing center and server rooms are secured against unauthorized access by means of an electronic access control system.
b. Smoke and water detectors are in place, temperatures in the computing center are monitored.
2) Redundant data storage in different locations (Cologne and Mannheim):
a. Frequent (up to daily) incremental and complete back-ups to onsite disk and tape libraries (tapes stored in suitable vault). In addition, frequent backups to offsite tape libraries.
3) Diversity of storage media (hard disk, tape) and frequent media refreshment.

The backup and storage procedures (1) (redundant and distributed storage) in place allow for fast and complete recovery / restoral of the archive holdings in case of a disaster.
An IT security concept regulates the protection of the infrastructure, data and services from GESIS. It contains the classification of data, services and network segments, network issues (WLAN, DMZ, Firewall), backup and storage, antivirus, encryption, user management, hardware, etc.

A concept for secure data management within the data archive was developed and will be implemented in 2017. Data may contain confidential information that may not be accessed either by the public or by staff that is not authorised. If not clarified before delivering data to the archive, employees who are trained in data protection issues have to decide at a very early stage, if a study contains confidential data that requires special protection. All further steps depend on this decision. To accommodate sensitive data, the existing workflows were expanded with additional measures. For this purpose we tested different encryption methods to find the best ones for our needs. On the one hand it should be as secure as possible and on the other hand it should also be as simple in handling as possible.
Access to confidential data is only given through our Secure Data Center (SDC) (2). It provides controlled and secure access to data deserving special protection. Data protection legislation requires that the possibility of re-identifying individuals in data provided by GESIS must be avoided. The SDC offers restricted access to data which has not been fully anonymised.

Links (last accessed 03.08.2017):
(1) Data Service Infrastructure for the Social Sciences and Humanities. "Report about Preservation Service

Offers", Deliverable: D4.2, p 47-70:
http://dasish.eu/publications/projectreports/D4.2_-_Report_about_Preservation_Service_Offers.pdf

(2) SDC: http://www.gesis.org/en/services/data-analysis/data-archive-service/secure-data-center-sdc/

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

17. Comments/feedback

Minimum Required Statement of Compliance:

0. N/A: Not Applicable.

Applicant Entry

Statement of Compliance:

0. N/A: Not Applicable.

Self-assessment statement:

No further comments.

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments: