gesis Leibniz Institute for the Social Sciences

Data Seal of Approval

Natascha Schumann, GESIS Data Archive for the Social Sciences

Member of the Leibniz Association

Headline & Content

- A question of trust: Brief overview about different initiatives
- European Framework for Audit and Certification of Digital Repositories
- New common requirements DSA/WDS
- Guidelines

Trusted digital repository

- "A trusted digital repository is one whose mission is to provide reliable, long-term access to managed digital resources to its designated community, now and in the future"

(RLG/OCLC Working Group: Research Libraries Group, 2002, p. i).

gesis Leibniz Institute for the Social Sciences

Trusted digital repository

- TDR has to ensure that the digital objects it preserves are not corrupted by accident or intentionally
- TDR has to ensure that access is given – not only physically, but also in appropriate digital formats
- Demonstration of know-how in digital preservation
- Transparency is very important in the context of trust: All stakeholders should have the opportunity to ascertain the statements made by the institution.

Member of the Leibniz Association

General aspects

- Criteria have to be as open as possible to meet the needs of a wide range of different repositories and archives
- Trust not only based on technical issues but also on organisational ones
- Repositories and archives have to demonstrate that they are aware of the challenges of digital curation and digital archiving

DATA HANDLING

History

- Since the mid-1990s different approaches and initiatives are concerned with "trust" within the context of digital archives
- How to ensure the integrity and authenticity of digital objects?
- How to develop standards for processes and procedures?

gesis Leibniz Institute for the Social Sciences

Different stakeholders

- Selection of publications, guidelines, criteria:
  - Task Force on Archiving of Digital Information (Commission on Preservation and Access and Research Libraries Group (RLG))
  - OCLC
    -> **TRAC Trustworthy Repositories Audit & Certification**
  - Consultative Committee for Space Data Systems (CCSDS)
    -> **RAC Repositories Audit & Certification**
  - Digital Curation Centre (DCC) und DigitalPreservationEurope (DPE)
    -> **DRAMBORA**
  - nestor AG Vertrauenswürdige Archive/ Zertifizierung
    -> **DIN 31644/nestor-Siegel**
  - Data Archive and Networked Services (DANS)
    -> **Data Seal of Approval**
  - Primary Trustworthy Digital Repository Authorisation Body (ISO-PTAB)
    -> **ISO 16363**

gesis Leibniz Institute for the Social Sciences

European Framework for Audit & Certification

- To harmonise the different approaches a Memorandum of Understanding with 3 initiatives was signed in 2010

-> European Framework for Audit & Certification of Digital Repositories

- Three different tiers that build upon each other

gesis Leibniz Institute for the Social Sciences

Framework

gesis Leibniz Institute for the Social Sciences

Theory and…

- … practice
  - Extented certification is only possible by DIN 31644
  - Formal certification only by ISO 16363

- Reflections about an evaluation of the framework with regard to these developments and the new common requirements
- Project proposal „Core Certification for Data Repositories in an Open Science Infrastructure" (Trust4Data) was rejected (Horizon2020)

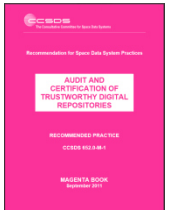
gesis Leibniz Institute for the Social Sciences

Involved initiatives

- Data Seal of Approval

- DIN 31644/nestor-Siegel

- ISO 16363 (ISO 16919)

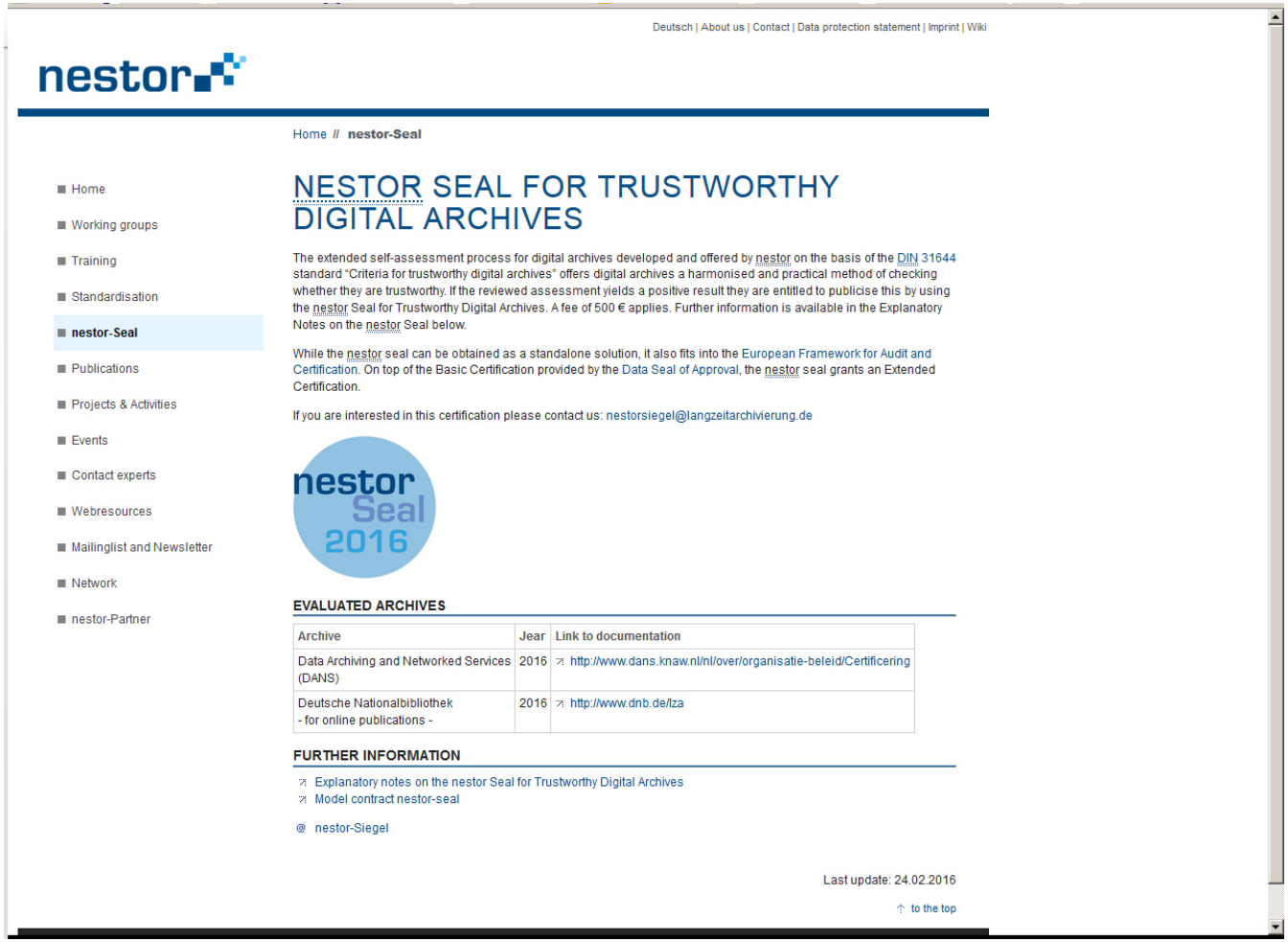


- ICSU/WDS

Member of the Leibniz Association

gesis Leibniz Institute for the Social Sciences

Website nestor



Member of the Leibniz Association

gesis Leibniz Institute for the Social Sciences

DIN 31644/nestor-Siegel

- Started in 2004 as a nestor Working group
  - nestor: German competence network for digital preservation
- Publication: "Catalogue of Criteria for Trusted Digital Repositories" (2006/2009)
- Since 2012: DIN 31644 (German Institute for Standardisation)
- Since 2013: nestor Seal, based upon DIN 31644
- No formal board, but nestor working group with currently 11 members from different institutions

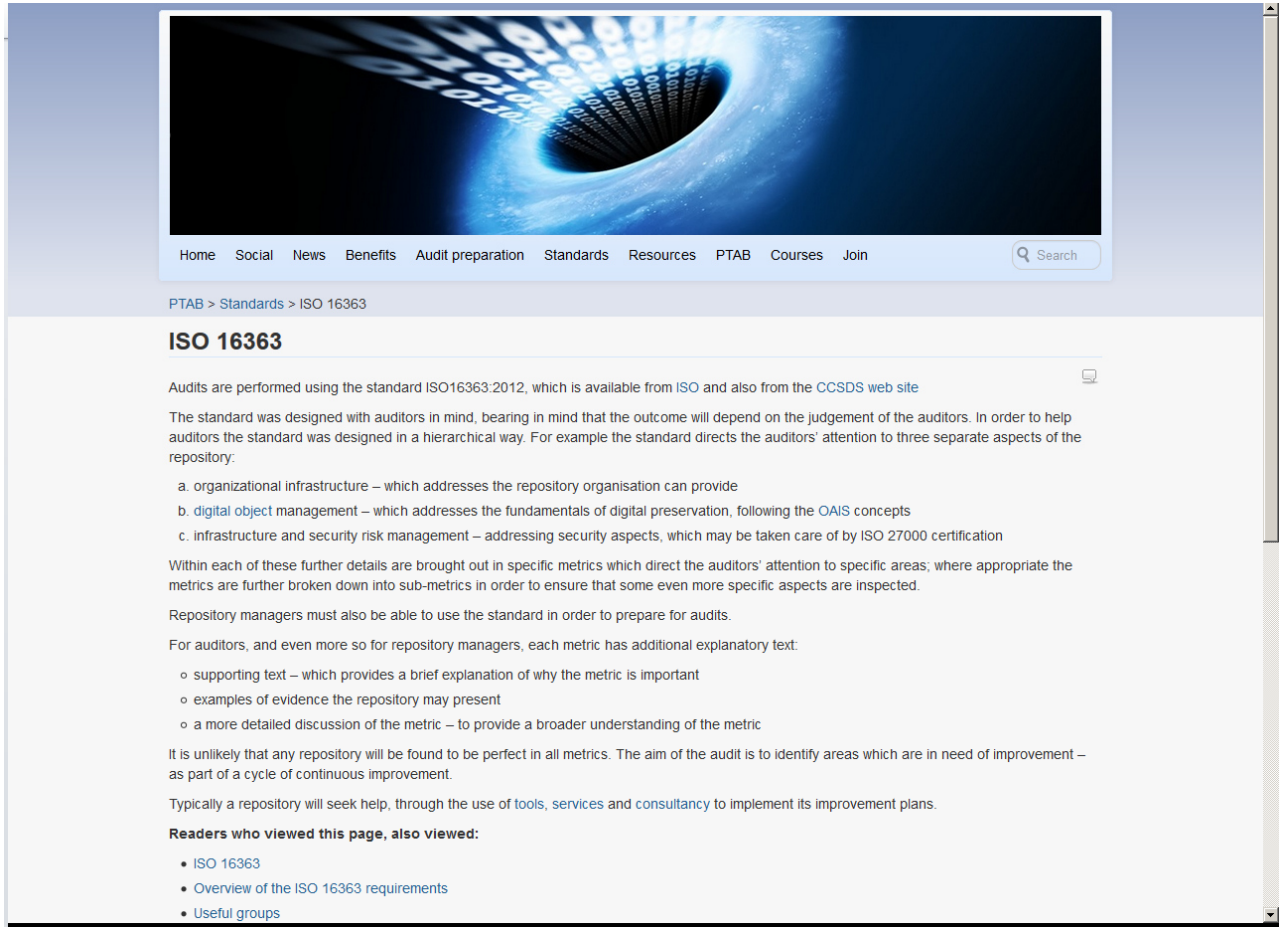
gesis Leibniz Institute for the Social Sciences

nestor Seal

- 34 criteria, divived in three parts:
  1. Organisational framework
  2. Object management
  3. Infrastructure and security
- Different levels of compliance
- Self assessment
- 500 € fee
- Currently developing tools for the assessment
- Review by two reviewers
- Currently 2 evaluated archives

gesis Leibniz Institute for the Social Sciences

Website PTAB



Member of the Leibniz Association

gesis Leibniz Institute for the Social Sciences

ISO 16363

- ISO 16363: Audit and certification of trustworthy digital repositories (2012), based on TRAC
- More than 100 criteria, divided in three parts
  1. Organisational infrastructure
  2. Object management
  3. Infrastructure and risk management
- PTAB (Primary Trustworthy Digital Repository Authorisation Body)
- Different compliance levels

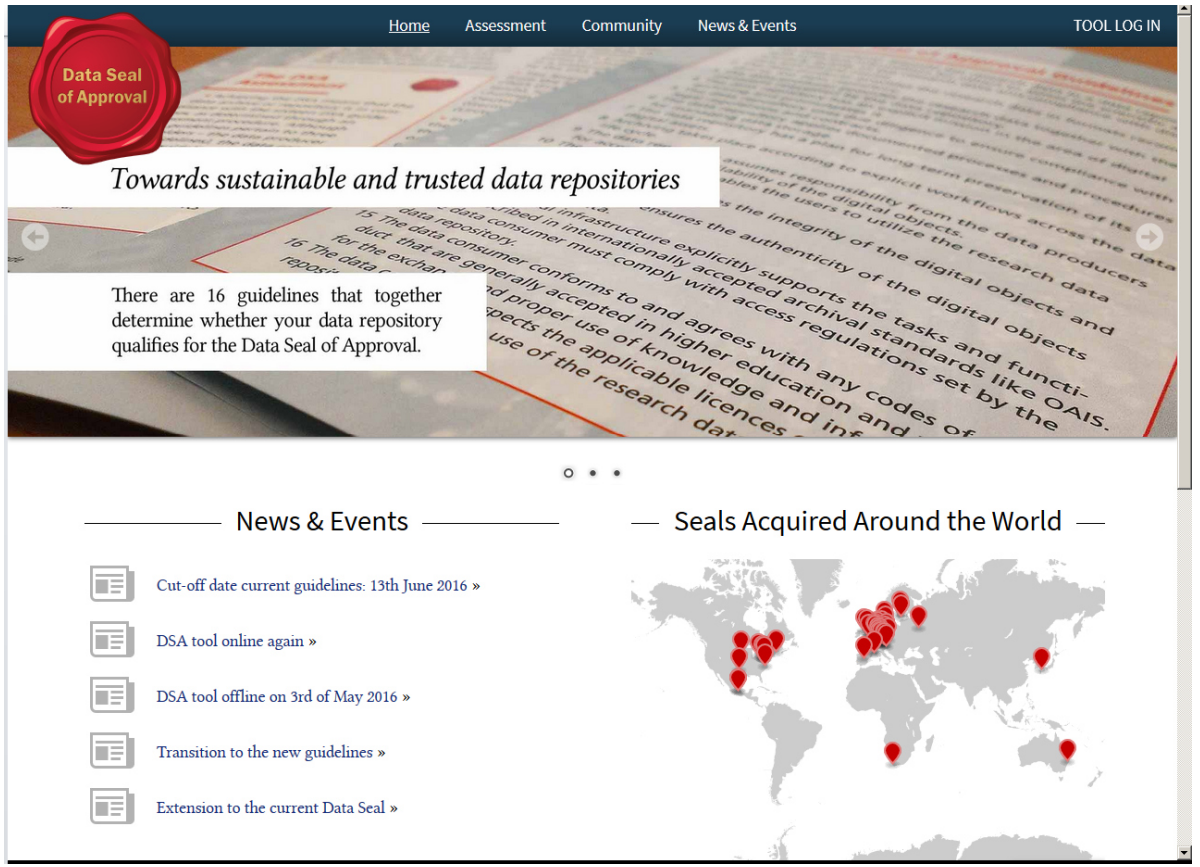
gesis Leibniz Institute for the Social Sciences

ISO 16363

- Full external audit
- Certification only allowed by organisations/persons who are certified by ISO 16919 (Requirements for Bodies Providing Audit and Certification of Candidate Trustworthy Digital Repositories)
- It was created for this purpose
- No formal certifications yet
- No information about fees and costs

gesis Leibniz Institute for the Social Sciences

Website DSA



Member of the Leibniz Association

gesis Leibniz Institute for the Social Sciences

Data Seal of Approval

- Developed in 2008 by DANS
- Lightwight approach
- Since 2009 International Board

Member of the Leibniz Association

gesis Leibniz Institute for the Social Sciences

Current International Board Members

- Ingrid Dillo, DANS
- Francoise Genova, Strasbourg Astronomical Data Center
- John Howard, University College Dublin
- Mari Kleemola, Finnish Data Archive (FSD)
- Herve L'Hours, UKDA (Chair)
- Marion Massol, CINES
- Natascha Schumann, GESIS
- Paul Trilsbeek, MPI

gesis Leibniz Institute for the Social Sciences

DSA

- 16 Guidelines, **originally** divided in three parts/perspectives:
  - Data producer
  - Data repository
  - Data user
- Different levels of compliance
- Self assessment
- Review through Board and DSA community members

gesis Leibniz Institute for the Social Sciences

DSA

- Online tool for assessment and review
- Every 2 years review of guidelines and compliance levels
- Currently 62 assessments
- New developments: DSA and World Data System (ICSU/WDS) started a common working group under the umbrella of RDA (Research Data Alliance) in 2012

gesis Leibniz Institute for the Social Sciences

Website WDS

Member of the Leibniz Association

gesis Leibniz Institute for the Social Sciences

ICSU WDS

- "As an ICSU Interdisciplinary body, the mission of the World Data System is to support ICSU's vision by promoting **long-term stewardship** of, and universal and equitable **access** to, quality-assured scientific data and data services, products, and information across a range of disciplines in the natural and social sciences, and the humanities. ICSU-WDS aims to facilitate scientific research under the ICSU umbrella by coordinating and supporting **trusted scientific data services** for the provision, use, and preservation of relevant datasets, while strengthening their links with the research community."

gesis Leibniz Institute for the Social Sciences

ICSU WDS Core Certification

- Mandatory certification for (different types of) members
- 17 criteria:
  1. WDS general requirements and policies (organization specific requirements)
  2. Organizational framework
  3. Management of data, products and services
  4. Technical infrastructure
- Based on self-assessment, reviewed by the WDS Scientific Committee
- Focus on (earth) sciences

gesis Leibniz Institute for the Social Sciences

DSA-WDS

- DSA more focused on social sciences and humanities and WDS more on physics and natural sciences
- DSA mainly in Europe, while WDS more common in US and Asia
- Despite the difference: similar approaches
- DSA and World Data System (ICSU/WDS) started a common working group under the umbrella of RDA (Research Data Alliance) in 2012

gesis Leibniz Institute for the Social Sciences

Common Working Group DSA-WDS

- Members from both organisations involved
- Developing a Common Catalogue of Requirements
- Developing common procedures for an assessment
- Simplifying assessments
- Motivation for more assessments
- Creation of a shared testbed for the new requirements

Common Requirements

- Structure is now oriented at the other standards
  - Context
  - Organisational infrastructure
  - Digital object management
  - Technology
  - Additional information and applicant feedback

Changes DSA – DSA-WDS

- No change of contents
- New structure: Like the standards within the framework
- Easier to go through
- Stronger emphasis on documented procedures and plans
- Clear labeling of the requirements

gesis Leibniz Institute for the Social Sciences

Comparison

- Structure is now (with the new DSA-WDS requirements) very similar
  - Organisational aspects
  - Management of digital objects
  - Technical aspects
- Different numbers of requirements
- Different kinds of review procedure
- Different costs

gesis Leibniz Institute for the Social Sciences

A walk through all Common Requirements

2014 DSA 2017

ICSU WORLD DATA SYSTEM

Member of the Leibniz Association

gesis Leibniz Institute for the Social Sciences

Compliance levels

- 0 – Not applicable
- 1 – The repository has not considered this yet
- 2 – The repository has a theoretical concept
- 3 – The repository is in the implementation phase
- 4 – The guideline has been fully implemented in the repository

gesis Leibniz Institute for the Social Sciences

Context

R0. Please provide context for your repository

- Repository Type. Select from:
  - Domain or subject-based repository
  - Institutional repository
  - National repository system, including governmental
  - Publication repository
  - Library/Museum/Archives
  - Research project repository
  - Other (Please describe)
- Brief Description of the Repository's Designated Community
- Level of Curation Performed. Select from:
  - Content distributed as deposited
  - Basic curation – e.g., brief checking, addition of basic metadata or documentation
  - Enhanced curation – e.g., creation of new formats, enhancement of documentation
  - Data-level curation – as in C above, but with additional editing of deposited data for accuracy
- Outsource Partner

gesis Leibniz Institute for the Social Sciences

I. Mission/Scope

R1. The repository has an explicit mission to provide access to and preserve data in its domain.

- Explicit statements of this role within the organization's mission and provide links.
- The level of approval within the organization that such a mission statement has received (e.g., approved public statement, roles mandated by funders, policy statement signed off by governing board).

gesis Leibniz Institute for Social Sciences

GESIS Constitution [only in German]

Satzung vom 05.12.2014

§ 1 Name und Sitz

Der Verein führt den Namen "GESIS – Leibniz-Institut für Sozialwissenschaften e.V". Er hat seinen Sitz in Mannheim und ist dort im Vereinsregister eingetragen.

§ 2 Zweck

(1) Der Verein dient der Förderung der sozialwissenschaftlichen Forschung. Er erbringt grundlegende, überregional und international bedeutsame forschungsbasierte Dienstleistungen für die Sozialwissenschaften. Er hat die Aufgabe, durch Grundlagenforschung sozialwissenschaftliche Untersuchungsansätze und Forschungsinstrumente zu entwickeln und zu verbessern.

(2) Der Vereinszweck wird insbesondere durch die Erfüllung der folgenden Aufgaben verfolgt:

a) kontinuierliche, interdisziplinäre Forschung und Entwicklung im Zusammenhang mit den in den Buchstaben b) bis g) dieses Absatzes genannten Aufgabenbereichen,

b) Beschreibung und Erklärung gesellschaftlicher Entwicklungen in nationaler, international vergleichender und historischer Perspektive einschließlich der Datenerhebung, statistischen Modellierung und Dauerbeobachtung,

c) Archivierung, Dokumentation und Langzeitsicherung sozialwissenschaftlicher Daten, einschließlich ihrer Erschließung sowie qualitativ hochwertigen Aufbereitung besonders relevanter Daten für Sekundäranalysen,

d) Aufbereitung von Literatur- und Forschungsinformationen,

e) Schaffung eines benutzerfreundlichen und hochqualitativen Zugangs zu allen für die empirische Sozialforschung relevanten Informationen und Daten einschließlich des Aufbaus und der Pflege sozialwissenschaftliche Portale und Kommunikationsnetzwerke, Entwicklung effektiver Instrumente für die Recherche, Aufbereitung, Auswertung, Sicherung und Archivierung der relevanten Informationen,

"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"

Member of the Leibniz Association

gesis Leibniz Institute for the Social Sciences

II. Licenses

R2. The repository maintains all applicable licenses covering data access and use and monitors compliance.

- License agreements in use.
- Conditions of use (distribution, intended use, protection of sensitive data, etc.).
- Documentation on measures in the case of noncompliance with conditions of access and use.

gesis Leibniz Institute for the Social Sciences

Usage regulations

Usage regulations - GESIS Dept. Data Archive and Data Analysis
(Valid as of April, 24th 2007)

Please note: Neither the depositor (individual(s), institute(s) etc.) nor GESIS bear any responsibility for the analysis or interpretation of the data which is supplied by GESIS.

For further information (e.g. on access categories, exact charges for special studies) please turn to:

- Oliver Watteler, M.A., Tel. 0221- 47694-76, <oliver.watteler@gesis.org>

1. Introduction

Data Archive and Data Analysis, in short "Data Archive", is a department of GESIS. The Data Archive preserves scientific data and document, especially from survey research, and makes this material available for further use.

The archive holdings include the data depositor's original data and documents as well as additional Data Archive material, which is the result of standardized documentation and processing (e.g. codebook, adjusted data sets, etc.).

Data and documents are made available exclusively on the basis of these regulations for use.

In the context of the international archive convention, users living outside Germany are asked to direct their requests regarding data to their home archive first.

2. General access conditions

As far as not explicitly indicated differently, the Data Archive makes data and documents available only for scientific analysis carried out in academic research and teaching. Institutes and individuals outside academic research and teaching can apply for provision in written form.

The acquisition of data and documents is done within the limits of access categories (please see 3.). The respective data depositor determines access limitations.

Independent of these access limitations for the use of data and documents, everybody is allowed to take a look at the documents – as far as not decided differently.

3. Access categories

The provision of data and documents is regulated by the following access categories. They are indicated in the respective study description in the Data Catalogue.

Category 0	Data and documents are released for everybody.
Category A	Data and documents are released for academic research and teaching.
Category B	Data and documents are released for academic research and teaching, if the results won't be published. If any publications, or any further work on the results is planned, a permission must be obtained by the Data Archive.
Category C	Data and documents are only released for academic research and teaching after the data depositor's written authorization. For this purpose the Data Archive obtains a written permission with specification of the user and the analysis intention.

4. Request and provision of material

To order material from the archive holding, forms are provided in which all necessary information for the request processing can be entered. The material is only made available for the purpose provided to the Data Archive or the data supplier. Passing on material to a third party is not allowed.

5. Completion of the project

The user is responsible to inform the Data Archive about the completion of the project for which the material was used. To prevent misuse the data has to be deleted and the medium carrying the data has to be made unreadable after completion of the project. In case further usage is intended, the user must apply to the Data Archive for new use permission.

6. Obligation to quote, specimen copy

The user is obliged to quote all used documents according to scientific conventions and to send two specimen copies of his/her publication to the Data Archive.

7. Charges

The charging regulations specify fees for the provision of data and documents from the Data Archive.

8. Final clause

Changes of the regulations for use and the scale of charges become effective from 24.04.2007.

gesis Leibniz Institute for the Social Sciences

III. Continuity of access

R3. The repository has a continuity plan to ensure ongoing access to and preservation of its holding.

- The **level of responsibility** undertaken for data holdings, including any guaranteed preservation periods.
- The plans in place to ensure the continued availability and accessibility of the data
  - **response to rapid changes** of circumstance
  - **long-term planning** should be described
  - indicating options for **relocation**

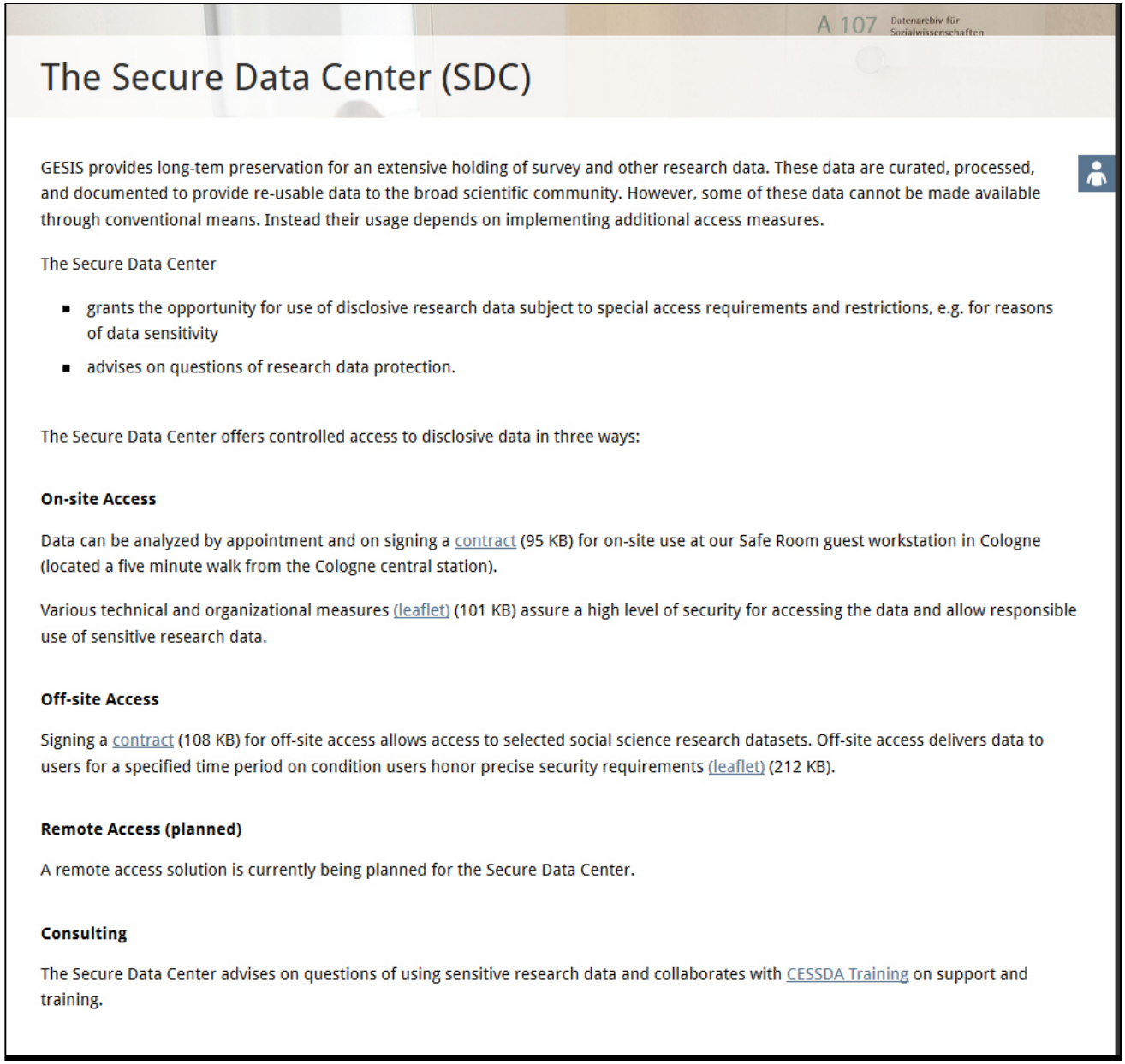
gesis Leibniz Institute for the Social Sciences

IV. Confidentiality/Ethics

R4. The repository ensures, to the extent possible, that data are created, curated, accessed, and used in compliance with disciplinary and ethical norms.

- How does the repository comply with applicable **disciplinary norms**?
  - confirmation that data collection was carried out in **accordance with legal and ethical criteria**
- Are special procedures applied to **manage data with disclosure risk**?
  - Storage and distribution of data with disclosure risk
- Are **procedures in place to review disclosure risk** in data, and to take the necessary steps?
- Are **staff trained** in the management of data with disclosure risk?

gesis Leibniz Institute for the Social Sciences



A 107 Datenarchiv für Sozialwissenschaften

The Secure Data Center (SDC)

GESIS provides long-tem preservation for an extensive holding of survey and other research data. These data are curated, processed, and documented to provide re-usable data to the broad scientific community. However, some of these data cannot be made available through conventional means. Instead their usage depends on implementing additional access measures.

The Secure Data Center

- grants the opportunity for use of disclosive research data subject to special access requirements and restrictions, e.g. for reasons of data sensitivity
- advises on questions of research data protection.

The Secure Data Center offers controlled access to disclosive data in three ways:

On-site Access

Data can be analyzed by appointment and on signing a contract (95 KB) for on-site use at our Safe Room guest workstation in Cologne (located a five minute walk from the Cologne central station).

Various technical and organizational measures (leaflet) (101 KB) assure a high level of security for accessing the data and allow responsible use of sensitive research data.

Off-site Access

Signing a contract (108 KB) for off-site access allows access to selected social science research datasets. Off-site access delivers data to users for a specified time period on condition users honor precise security requirements (leaflet) (212 KB).

Remote Access (planned)

A remote access solution is currently being planned for the Secure Data Center.

Consulting

The Secure Data Center advises on questions of using sensitive research data and collaborates with CESSDA Training on support and training.

Member of the Leibniz Association

gesis Leibniz Institute for the Social Sciences

V. Organizational infrastructure

R5. The repository has adequate funding and sufficient numbers of qualified staff managed through a clear system of governance to effectively carry out the mission.

- The repository is hosted by a **recognized institution** appropriate to its Designated Community.
- The repository has **sufficient funding**
- The repository ensures that its **staff** have access to ongoing training and professional development.

gesis Leibniz Institute for the Social Sciences

VI. Expert guidance

R6. The repository adopts mechanism(s) to secure ongoing expert guidance and feedback (either in-house, or external, including scientific guidance, if relevant)

- Does the repository have in-house **advisers**, or an external **advisory committee** that might be populated with technical members, data science experts, and disciplinary experts?
- How does the repository **communicate** with the **experts** for advice?
- How does the repository **communicate** with its **Designated Community** for feedback?

gesis Leibniz Institute for the Social Sciences

Rules of internal procedure for the Scientific Advisory Board of GESIS

§ 1 Legal basis

The legal basis of these rules of internal procedure is the GESIS constitution, in particular Section 11, in which it is stated that "The Scientific Advisory Board shall hereby establish the rules of internal procedure, requiring the approval of the Board of Trustees."

§ 2 Duties of the Scientific Advisory Board

The Scientific Advisory Board shall take a position concerning the long-term development of the association and advise the President and the Board of Trustees in the matter of recommendations of the senate of the Leibniz-Gemeinschaft for the fulfilment of its duties. In particular, it shall have the following tasks (see Section 11 of the constitution):

(1) Voicing its position concerning long-term research and development planning and the programme budget
(2) Voicing its position in respect of the proposed appointments for the offices of President and Scientific Director
(3) Performing department audits
(4) Reporting to the Board of Trustees.

§ 3 The chairman

The Scientific Advisory Board elects a Chairman and a Deputy Chairman by a simple majority vote for a term of four years. One-term and immediate re-election are possible (see constitution, Section 11, Paragraph 3).

(1) The Chairman shall convene and preside over the meetings of the Scientific Advisory Board. Furthermore, the Chairman shall compile an agenda together with the Deputy Chairman.
(2) The Chairman shall represent the Scientific Advisory Board externally.

§ 4 The members

The Board of Trustees shall appoint the members of the Scientific Advisory Board for a term of four years. Appointment may be extended by one term. The President of GESIS and the members of the Scientific Advisory Board can submit proposals for the appointment of the members.

(1) The Scientific Advisory Board shall be comprised of up to 12 external, internationally recognised and currently active scientists.
(2) The composition shall consider the main areas of activity of GESIS.
(3) At least two scientists shall belong to the Scientific Advisory Board who are active at institutions outside of Germany.
(4) In the last meeting of the member's term of office, proposals for new appointment and re-appointment shall be adopted and then made known to the Board of Trustees.

...

Member of the Leibniz Association

gesis Leibniz Institute for the Social Sciences

VII. Data integrity and authenticity (I)

R7. The repository guarantees the integrity and authenticity of the data.

- Description of **checks** to verify that a digital object has not been altered or corrupted (i.e., fixity checks).
- Documentation of the **completeness** of the data and metadata.
- Details of how all **changes** to the data and metadata are logged.
- Description of **version control** strategy.
- Usage of appropriate international **standards** and conventions (which should be specified).

gesis Leibniz Institute for the Social Sciences

VII. Data integrity and authenticity (II)

- Does the repository have a **strategy for data changes**? Are data producers made aware of this strategy?
- Does the repository maintain **provenance data** and related audit trails?
- Does the repository maintain **links to metadata** and to other datasets? If so, how?
- Does the repository **compare** the essential properties of different versions of the same file? How?
- Does the repository check the **identities** of depositors?

gesis Leibniz Institute for the Social Sciences

Versioning ▲

Data stored in the archive undergo revision and changes even after their publication. For example, subsequently discovered errors are corrected, or the data is augmented by additional variables or interviews. Assigning version numbers guarantees datasets used for publications are identifiable together alongside their study number, allowing for unique referencing and citation.

A persistent identifier (DOI name) assigned to each version also makes the data easier to locate. DOI names link the user directly to the study description in the DBK.

Changes are documented on three levels: Major.Minor.Revision (e.g. 2.1.0):

1. Position – Major:

- Addition of one or more new samples (usually countries) in an integrated or cumulative data set
- Addition of one or more new waves in a cumulative data set
- Addition/deletion of one or more variables in a data set
- Addition/deletion of one or more cases in a data set
- Enhanced processing for a higher class (usually class 1)

2. Position – Minor

- Changes relevant to the meaning of a variable, or completing in the data set (label, recoding, data formats…)

3. Position – Revision

- Changes that do not affect the meaning of a variable (e.g. correction of spelling mistakes)
- Simple revision of labels without change in meanin

Example:

A spelling mistake in a data set with version 1.2.3 is corrected (→1.2.4), a variable is recoded (→1.3.0), and a variable is added (→ 2.0.0). If all the changes are made at once, version number 2.0.0 is assigned. If only the first mentioned two changes have been carried out, version number 1.3.0 is assigned.

Member of the Leibniz Association

gesis Leibniz Institute for the Social Sciences

VIII. Appraisal

R8. The repository accepts data and metadata based on defined criteria to ensure relevance and understandability for data users.

- Does the repository use a **collection development policy**?
- Does the repository have **quality control checks** to ensure the completeness and understandability of data?
- Does the repository have procedures in place to determine that the **metadata** required to interpret and use the data are provided?
- Does the repository publish a list of **preferred formats**?

gesis Leibniz Institute for the Social Sciences

Recommended formats

To support the long-term preservation, interpretability, and accessibility of data, choosing suitable file formats is of particular importance. Just like hardware, software constantly evolves. For example, new functions are added to software programs, or software is adapted to new operating systems. Both can lead to changes in the file format. In consequence, digital data is constantly at risk from changes in the hard- and software environment. This risk can be mitigated if suitable file formats are used.

The GESIS Data Archive recommends using the following formats for the most important object classes:

Submitted datasets should be usable in one of the widely used statistical packages (SPSS, Stata or SAS). More specifically, data can be submitted in the following forms:

1. As so-called system files in the proprietary formats of common statistical packages (e.g. SPSS System File).

2. In software-specific portable file formats (e.g. SAS Transport File).

3. As text files (comma-, tab-delimited formats) with the required setup or syntax files to enable importing into statistical packages.

Type of data	Preferred formats	Acceptable formats
Dataset (statistical file formats)	▪ SPSS Portable (*.por) ▪ STATA (*.dta) ▪ SAS Transport (*.sas) ▪ Widely used (proprietary) formats of statistical packages, e.g. SPSS (*.sav), Stata (*.dta), SAS (*.sas7bdat) ▪ Tab-, comma-, delimited text files ("csv") with setup file (setup, command or syntax file for SPSS, Stata, SAS, etc.) and the respective data definitions (variable names and labels, missing values, etc.). Alternatively, data definitions can be submitted as DDI-XML file.	▪ OpenDocument table format (*.ods) , MS Excel (*.xls, *.xlsx), MS Access (*.mdb, *.accdb) ▪ CSV formats without data definition files (setup, syntax, command file) ▪ Column binary format (a standard making it possible to represent data as images of punch cards) or card image format.
Documentation (texts)	▪ PDF/A (*.pdf) ▪ Text formats (ASCII, ANSI, etc.)	▪ OpenDocument text (*.odt) ▪ PDF (*.pdf) ▪ MS Word (*.doc, .docx) ▪ RichText (*.rtf) ▪ WordPerfect (*.wpd, *.cwp, *.vwp) ▪ HTML (*.htm)
Images	▪ TIFF Version 6 uncompressed (*.tif)	▪ JPEG 2000 ▪ JPEG, PNG, GIF, BMP ▪ PDF/A, PDF (*.pdf)

The Data Archive will accept additional formats, especially for data, which (upon consultation) can be converted into preferred formats for preservation. Regardless of the specific file format datasets should always be structured in a manner allowing third parties to read and understand them.

Data may therefore not be encrypted. In addition, functions such as printing or copying should not be disabled.

Member of the Leibniz Association

gesis Leibniz Institute for the Social Sciences

IX. Documented storage procedures

R9. The repository applies documented processes and procedures in managing archival storage of the data.

- How are relevant **processes and procedures** documented and managed?
- What **levels of security** are required, and how are these supported?
- Are **backup strategy** and **data recovery** in place?
- Are **risk management** techniques used to inform the strategy?

gesis Leibniz Institute for the Social Sciences

X. Preservation plan

R10. The repository assumes responsibility for long-term preservation and manages this function in a planned and documented way.

- Is the '**preservation level**' for each item understood? How is this defined?
- Does the **contract between depositor and repository** provide for all actions necessary to meet the responsibilities?
- Does the repository have the **rights** to copy, transform, and store the items, as well as provide access to them?
- Are **actions** relevant to preservation specified in documentation, including custody transfer, submission information standards, and archival information standards?

gesis Leibniz Institute for the Social Sciences

Archive agreement

§ 2 Property and usage rights

By providing (delivering, making available) the study, the data provider transfers to the Archive the rights to use the content – i.e. data and text(s) – stipulated in this agreement, in particular

1. to systematically archive and prepare the data and text(s) for the purpose of long-term physical storage and further evaluation. In this the Archive may utilize all technical means, formats and methods indicated to serve this purpose;
2. to digitize the text(s) belonging to the study(ies) in the event that they are not provided in digital form, and to make them publicly available for downloading as part of the online offerings of the Archive, if there are no prohibiting agreements or arrangements;
3. to make publicly available to Archive users in accordance with the terms of use known to the data provider and attached to this agreement representing the following set access categories:

Category		✓
0 (zero)	Data and documents are released for everybody	
A	Data is released for academic research and teaching.	
B	Data is released for academic research and teaching, if the results won't be published. If any publications or any further work on the results is planned, permission must be obtained from the data depositor.	
C	Data is only released for academic research and teaching after the data depositor's written authorization. For this purpose the Data Archive obtains a written permission with specification of the user and the analysis intention.	

Unless there is an agreement regarding the access category (above) the material delivered is to be considered accessible under the terms of Category 0 (i.e. default category).

Classification into category B or C takes place for a period from ..... year(s) following initial archiving. In the event that no time period is specifically named for limited access, the study will automatically fall into category A at the end of one year. When not explicitly stated the classification into category B or C applies solely to the datasets, not the accompanying text(s).

The data provider transfers to the Archive all necessary usage rights for the stated purpose, in particular rights to reproduction (§16 UrhG) as well as the right of public accessibility (§ 19 a UrhG). Granting of usage rights is unlimited in terms of place and time period.

The data provider agrees to publication of the study's metadata.

The Archive takes responsibility for storing the delivered materials in original or digitized form, free of charge, to be made available to the data provider upon simple request.

In the context of its work, the Archive is allowed to make usage of assistance from third parties.

2

Member of the Leibniz Association

gesis Leibniz Institute for the Social Sciences

XI. Data quality

R11. The repository has appropriate expertise to address technical data and metadata quality and ensures that sufficient information is available for end users to make quality-related evaluations.

- The approach to **data and metadata quality** taken by the repository.
- The ability of the **Designated Community** to comment on, and/or rate data and metadata.
- Whether **citations** to related works or links to citation indices are provided.

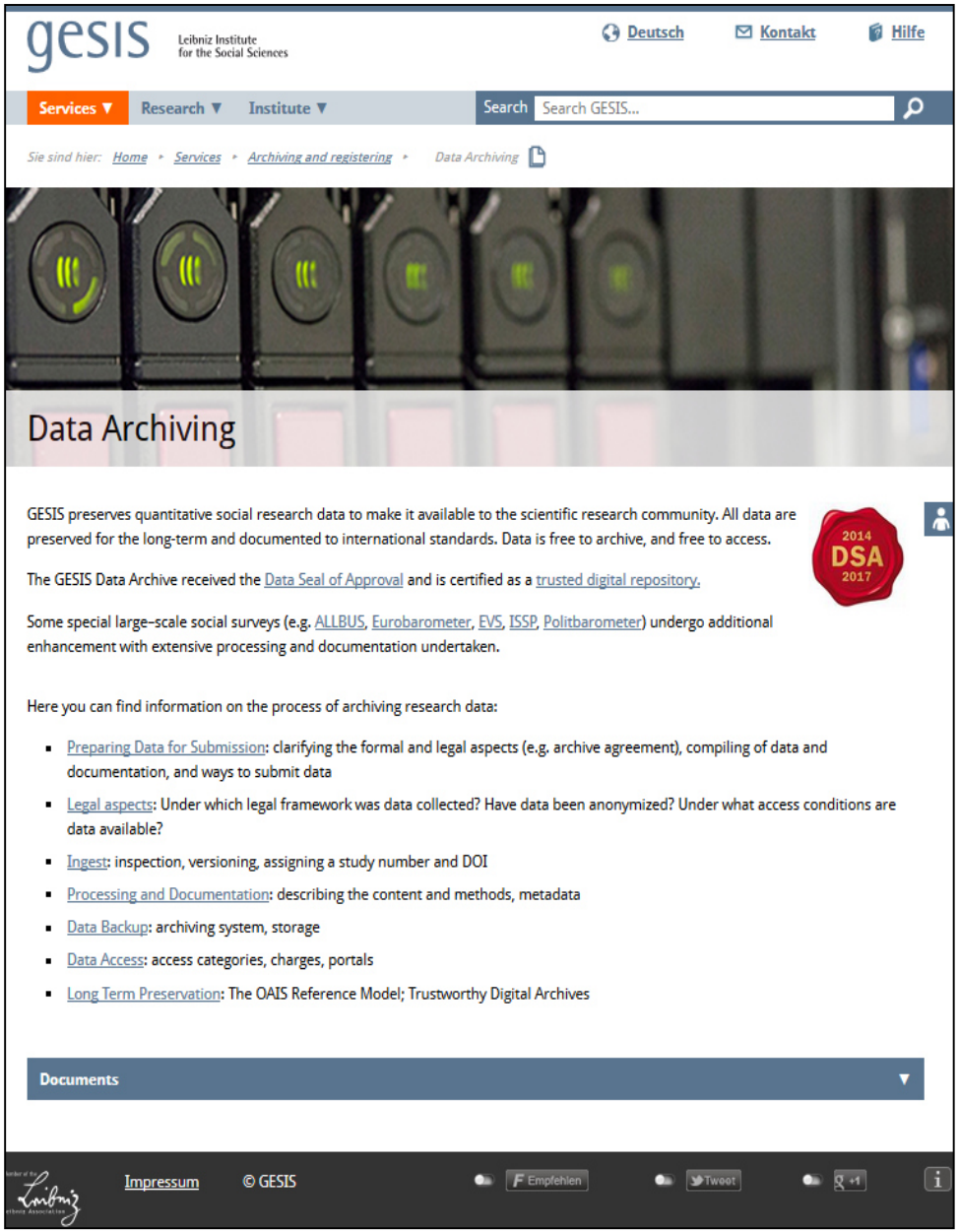
gesis Leibniz Institute for the Social Sciences

XII. Workflows

R12. Archiving takes place according to defined workflows from ingest to dissemination.

- **Workflows/business process** descriptions.
- **Levels of security** and impact on workflows (guarding privacy of subjects, etc.).
- **Appraisal and selection** of data.
- Approaches towards **data that do not fall within the mission/collection profile.**
- **Change management** of workflows.

gesis Leibniz Institute for the Social Sciences



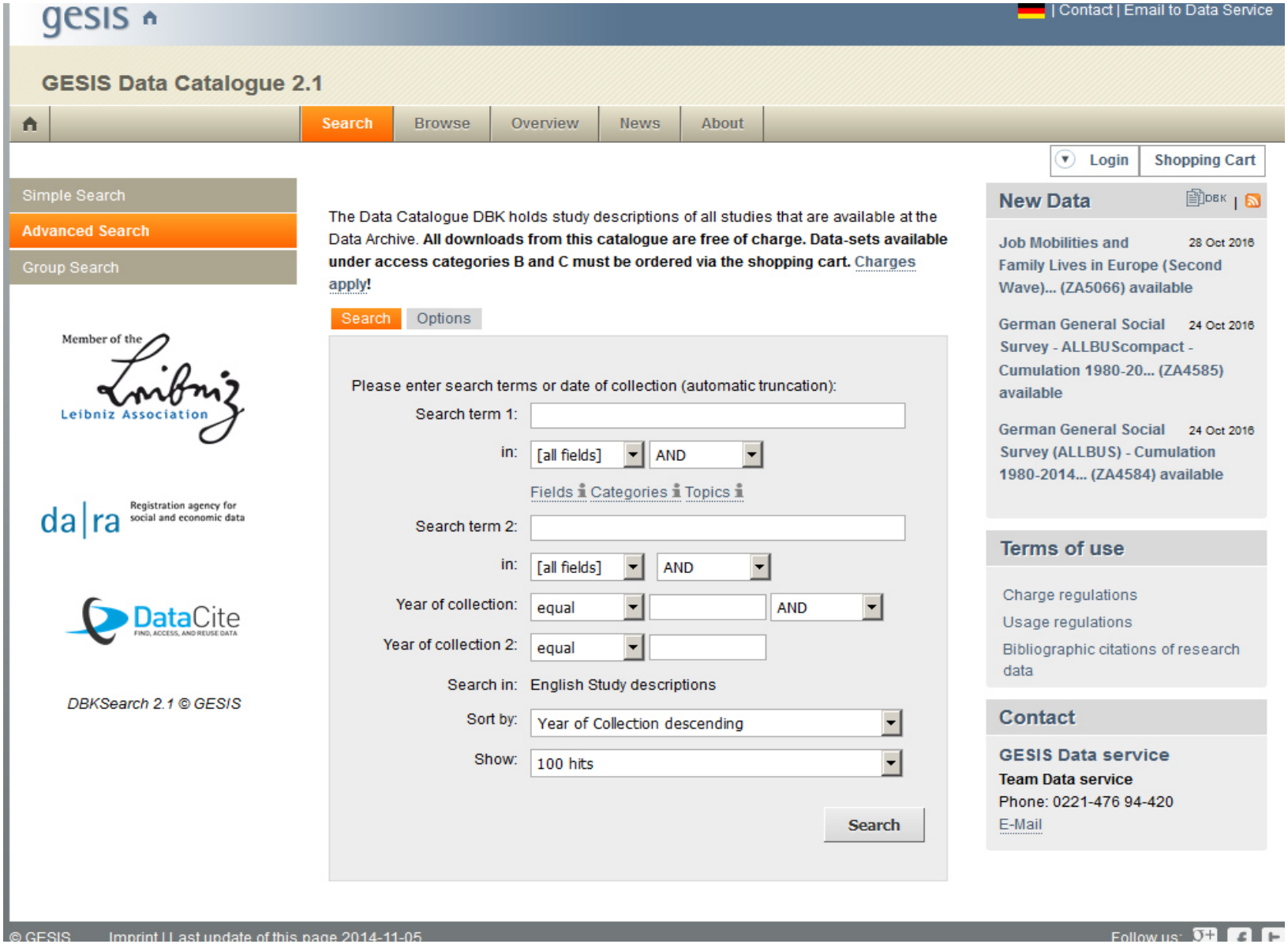
gesis Leibniz Institute for the Social Sciences

XIII. Data discovery and identification

R13. The repository enables users to discover the data and refer to them in a persistent way through proper citation.

- **Search facilities**
- **Searchable metadata**? Harvesting?
- Is the repository **included** in **registries**?
- Does the repository offer **recommended data citations**?
- Does the repository offer **persistent identifiers**?

gesis Leibniz Institute for the Social Sciences



gesis Leibniz Institute for the Social Sciences

XIV. Data reuse

R14. The repository enables reuse of the data over time, ensuring that appropriate metadata are available to support the understanding and use of the data.

- Which **metadata are required** by the repository?
- Are data provided in **formats** used by the Designated Community? Which formats?
- Are measures taken to account for the possible **evolution of formats?**
- Are plans related to **future migrations** in place?
- How does the repository **ensure understandability** of the data?

gesis Leibniz Institute for the Social Sciences

XV. Technical infrastructure

R15. The repository functions on well-supported operating systems and other core infrastructural software and is using hardware and software technologies appropriate to the services it provides to its Designated Community.

- What **standards** does the repository use for reference? Are these international and/or community standards? How often are these reviewed?
- How are the **standards implemented**?
- Does the repository have a plan for **infrastructure development**?
- Is a **software inventory** maintained and is system documentation available?

gesis Leibniz Institute for the Social Sciences

XVI. Security

R16. The technical infrastructure of the repository provides for protection of the facility and its data, products, services, and users.

- Procedures and arrangements are in place to provide **swift recovery** or backup
- Your **IT security system**, disaster plan, and business continuity plan

gesis Leibniz Institute for the Social Sciences

XVII. Additional information

R17. Any other relevant information you wish to provide on your repository.

- The repository may **add any extra information** that is not covered in the above Requirements but that may be helpful to the reviewers in making their assessment.
- For example:
  - The usage and impact of the repository data holdings (citations, use by other projects, etc.).
  - A national, regional, or global role that the repository serves.
  - Any global cluster or network organization that the repository belongs to.

Phew!

Benefits of "doing" the DSA

- Prerequiste for CESSDA membership
- Quality assurance
- Trust against stakeholders
  - Data producers
  - Data users
  - Funders
- For new archives: Guideline and support for build up the archive
- For „old“ ones: evaluation of established procedures and processes

gesis Leibniz Institute for the Social Sciences

Some handy tips

- Creation of lists (in case they are online: links)
  - all relevant documents already available
  - all documents not yet completed
  - all documents still needed

- Possible arrangements
  - Policies/Mission statement
  - Contracts and SLA
  - General information/Websites
  - Technical documentation
  - Publications

Some handy tips

- Colleguages to be involved
- Time schedule
- Regular meeting with a core group
- Creation of missing documents
- Evaluation of website

gesis Leibniz Institute for the Social Sciences

Publications and Links

- Common requirements DSA-WDS: https://docs.google.com/document/d/1_DPwSA5P8LpK9Q34BhxJmX8So2GKL7eSLa-G-z5JvVg/edit
- nestor Catalogue of Criteria for Trusted Digital Repositories: http://nbn-resolving.de/urn:nbn:de:0008-2010030806
- PTAB: http://www.iso16363.org/
- Preserving Digital Information. Report of the Task Force on Archiving of Digital Information (Commission on Preservation and Access and Research Libraries Group (RLG)), 1996 http://oclc.org/content/dam/research/activities/digpresstudy/final-report.pdf
- RLG u. OCLC: Trusted Digital Repositories: Attributes and Responsibilities, 2002 http://www.oclc.org/content/dam/research/activities/trustedrep/repositories.pdf

gesis Leibniz Institute for the Social Sciences

Publications and Links

- Internationale Task Force: Trustworthy Repositories Audit & Certification, 2007 http://www.repositoryaudit.eu/about/
- Repositories Audit und Certification RAC, Consultative Committee for Space Data Systems (CCSDS) http://public.ccsds.org/publications/archive/652x0m1.pdf

Natascha Schumann

natascha.schumann@gesis.org

Picture: Private