

nestor for newbies

Das Konzept der Vertrauenswürdigkeit

Natascha Schumann
GESIS Datenarchiv für Sozialwissenschaften

- ..■ Hintergrund und Historie
- ..■ European Framework for Audit and Certification of Digital Repositories
 - ..■ Data Seal of Approval (DSA)
 - ..■ DIN 31644/nestor-Siegel
 - ..■ ISO 16363 (RAC) + ISO 16919
 - ..■ Kriterien und Beispiele DSA und nestor-Siegel
- ..■ Warum das Ganze?

- .. Seit ca. Mitte der 1990er Jahre Beschäftigung mit der Frage, wie ein digitales Archiv seine Vertrauenswürdigkeit unter Beweis stellen kann



Trust me by [Heirozen](#). [Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 License](#)

- ..■ Unterschiedliche Initiativen und Akteur*innen haben sich mit dem Thema beschäftigt und Publikationen, Guidelines, Kriterienkataloge etc. entwickelt
 - ..■ 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**

- .. Unterzeichnung eines Memorandum of Understanding (MoU) durch Vertreter dreier Initiativen (2010)



- ..■ DSA
- ..■ DIN 31644/nestor-Siegel
- ..■ ISO 16363

- ..■ Vergleich/Überblick



- ..■ Ursprünglich von DANS entwickelt, Internationales Board
- ..■ 16 Guidelines
- ..■ Unterteilung in 3 Bereiche:
 - ..■ Datengeber (Guidelines 1-3)
 - ..■ Datenrepository (Guidelines 4-13)
 - ..■ Datennutzer (Guidelines 14-16)
- ..■ Verschiedene Erfüllungsgrade



- ..■ Selbstevaluierung
- ..■ Review durch Mitglieder des Boards und erweiterter Kreis aus der DSA-Community
- ..■ Überarbeitung der Richtlinien und Anpassungen der Erfüllungsgrade pro Guideline: alle zwei Jahre
- ..■ 56 Archive haben inzwischen das DSA erhalten



..■ Neuerungen

- ..■ Kooperation im Rahmen der Research Data Alliance (RDA) mit dem ICSU World Data System (WDS)
- ..■ Entwicklung gemeinsamer Kriterien
- ..■ Verbindung der eher geistes- und sozialwissenschaftlichen Ausrichtungen des DSA mit der eher naturwissenschaftlichen des WDS
- ..■ Testphase gerade abgeschlossen und Einführung in diesem Jahr



| Statement of Compliance | Means | Comments and/or URLs |
|-------------------------|----------------|---|
| 0 | Not Applicable | Provide an explanation |
| 1 | No | Provide an explanation |
| 2 | Theoretical | Provide a URL for the initiation document |
| 3 | In Progress | Provide a URL for the supporting document |
| 4 | Implemented | Provide a URL for the supporting document |

- ..■ Entwickelt von der nestor-AG „Vertrauenswürdige Archive/ Zertifizierung“
- ..■ nestor-Kriterienkatalog Vertrauenswürdige digitale Langzeitarchive, 2006/2008

<http://nbn-resolving.de/urn:nbn:de:0008-2008021802>

- ..■ DIN 31644 Kriterien für vertrauenswürdige Archive, 2012 (Basis für nestor-Siegel)

http://www.langzeitarchivierung.de/Subsites/nestor/DE/nestor-Siegel/siegel_node.html

- ..■ 34 Kriterien
- ..■ Unterteilung:
 - ..■ Organisatorischer Rahmen
 - ..■ Umgang mit Informationsobjekten
 - ..■ Infrastruktur und Sicherheit
- ..■ Selbstevaluierung
- ..■ Review durch zwei Gutachter aus der nestor-AG Zertifizierung
- ..■ Zwei Archive haben das nestor-Siegel bislang erhalten

| Punkte | Erläuterung |
|--------|----------------------------|
| 0 | Noch offen, keine Konzepte |
| 3 | Konzept vorhanden |
| 6 | Ausgearbeitetes Konzept |
| 10 | Implementiert |

Bei der Bewertung der anwendbaren Kriterien K13 - K34 muss ein Durchschnitt von 7 Punkten erreicht werden.

- ..■ **ISO 16363: Audit and certification of trustworthy digital repositories (2012)**, basiert auf TRAC
- ..■ 50 Hauptkriterien, insgesamt 109
- ..■ Unterteilung:
 - ..■ Organisatorische Infrastruktur
 - ..■ Umgang mit digitalen Objekten
 - ..■ Infrastruktur und Risikomanagement
- ..■ Wenig Informationen verfügbar

- ..■ Zertifizierung nur erlaubt durch Organisationen/Personen, die nach ISO 16919 (Requirements for Bodies Providing Audit and Certification of Candidate Trustworthy Digital Repositories) zertifiziert sind
- ..■ Template:
http://www.iso16363.org/?smd_process_download=1&download_id=30

| | |
|---|---|
| | |
| 0 | Non-compliant or not started: The repository has not yet addressed the requirement or has not started the review of the requirement. |
| 1 | Slightly compliant: The repository has something in place, but has a lot of work to do in addressing the requirement. |
| 2 | Half compliant: The repository has partially addressed the requirement and has significant work remaining to fully address the requirement. |
| 3 | Mostly compliant: The repository can demonstrate that it has mostly addressed the requirement and is working on full compliance. |
| 4 | Fully compliant: The repository can demonstrate that it has comprehensively addressed the requirement. |

| | DSA | nestor-Siegel/DIN 31644 | ISO 16363 |
|-----------|--|--|--|
| Kriterien | 16 | 34 | 50 Hauptkriterien, insgesamt 109 |
| Aufbau | <ol style="list-style-type: none"> 1. Datengeber 2. Archiv 3. Datennutzer | <ol style="list-style-type: none"> 1. Organisatorischer Rahmen 2. Umgang mit Informationsobjekten 3. Infrastruktur und Sicherheit | <ol style="list-style-type: none"> 1. Organisatorische Infrastruktur 2. Umgang mit digitalen Objekten 3. Infrastruktur und Risikomanagement |
| Review | Self-Assessment, DSA-Board | Self-Assessment, AG-Zertifizierung | Zertifizierung nach 16919 |
| Kosten | - | 500€ | ? |

- ..■ 16 DSA Guidelines und Beispiel
- ..■ 24 Kriterien nestor-Siegel und Beispiel



1. The data producer deposits the research data in a data repository with **sufficient information** for others to assess the scientific and scholarly quality of the research data and compliance with disciplinary and ethical norms
2. The data producer provides the research data in **formats recommended** by the data repository
3. The data producer provides the research data together with the **metadata** requested by the data repository



4. The data repository has an **explicit mission** in the area of digital archiving and promulgates it
5. The data repository uses due diligence to ensure **compliance with legal regulations** and contracts including, when applicable, regulations governing the protection of human subjects.
6. The data repository applies **documented processes and procedures** for managing data storage
7. The data repository has a **plan for long-term preservation** of its digital assets
8. Archiving takes place according to **explicit workflows** across the data life cycle



9. The data repository assumes **responsibility** from the data producers for access and availability of the digital objects
10. The data repository enables the users to **utilize** the research data and refer to them
11. The data repository ensures the **integrity** of the digital objects and the metadata
12. The data repository ensures the **authenticity** of the digital objects and the metadata
13. The **technical infrastructure** explicitly supports the tasks and functions described in internationally accepted archival standards like **OAIS**



14. The data consumer complies with **access regulations** set by the data repository
15. The data consumer conforms to and agrees with any **codes of conduct** that are generally accepted in higher education and research for the exchange and proper use of knowledge and information
16. The data consumer respects the applicable **licenses** of the data repository regarding the use of the research data



1. The data producer deposits the data in a data repository with sufficient information for others to assess the quality of the data, and compliance with disciplinary and ethical norms.

Applicant Entry

Self-assessment statement:

Statement of Compliance:

Links:

Reviewer Entry

Accept

1. The data producer deposits the data in a data repository with sufficient information for others to assess the quality of the data, and compliance with disciplinary and ethical norms.

Minimum-Required-Statement-of-Compliance:
3. In progress: We are in the implementation phase.

Applicant-Entry

Statement-of-Compliance:
4. Implemented: This guideline has been fully implemented for the needs of our repository.
Self-assessment-statement:

Sufficient information on depositing and archiving of social science research data is provided on the GESIS web pages. Remaining questions and details are clarified during the pre-ingest phase.

Pre-ingest communication with data producers adheres to important steps of the Preliminary and Formal Definition Phases as outlined in the Producer-Archive Interface Methodology Abstract Standard (PAIMAS) standard (CCSDS-2004). The Preliminary Phase is designed to gather necessary information, establish contact with data producers, and to provide the data producers with all necessary information – in particular about the archiving services offered by GESIS, the minimal requirements that have to be fulfilled for GESIS to accept data for archiving (see below), as well as licensing options and the archive agreement. During the Formal Definition Phase, agreements and schedules sketched before are discussed in more detail and on a more concrete level. In this phase, subject-related, technical, and legal questions are clarified, and agreements made with regard to delivery formats, schedules, etc. This phase concludes with the signing of the archive agreement. The Data Archive 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).

Comprehensive (general) information on data management, data archiving, data sharing, and re-use, as well as specific information on how to deposit with the GESIS Data Archive is given on the GESIS web pages. Beyond that GESIS actively promotes and supports all aspects of research data management, including e.g. long-term preservation, re-use, citation, through own research & development, publications, presentations, as well as active participation in a wide range of initiatives and projects.

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.

Links:

Data-Archive-Website: <http://www.gesis.org/en/services/archiving-and-registering/data-archiving/>

PAIMAS standard (CCSDS-2004): <http://public.ccsds.org/publications/archive/651x0m1.pdf>

Archive-agreement: http://www.gesis.org/fileadmin/upload/institut/wiss_arbeitsbereiche/datenarchiv_analyse/Archivierungsvertrag_GESIS_Datenarchiv_v9_englisch.pdf

General Advice on Data Management & Archiving: <http://www.gesis.org/en/archive-and-data-management-training-and-information-center/training-center-home>

Jensen, Uwe (2012): Leitlinien zum Management von Forschungsdaten. Sozialwissenschaftliche Umfragedaten. GESIS Technical Reports, 2012/07: http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis_reihen/gesis_metho_denberichte/2012/TechnicalReport_2012-07.pdf

daorium: <https://datorium.gesis.org/xmlui/?locale-attribute=en>

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

Data Catalogue and other portals: <http://www.gesis.org/en/services/research/>

Example of a study description (European Values Study 2008: Germany (EVS 2008)) in the Data Catalogue: <https://dbk.gesis.org/dbksearch/SDesc2.asp?no=4753&tab=0&il=10¬abs=&af=&nf=1&search=evs&search2=&db=E>

All-URLs accessed 07-April-2014

Reviewer-Entry

Accept or send back to applicant for modification:

Accept

Comments:



- 1. The data producer deposits the data in a data repository with sufficient information for others to assess the quality of the data, and compliance with disciplinary and ethical norms.**

Minimum Required Statement of Compliance:

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

Applicant Entry

Statement of Compliance:

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



Sufficient information on depositing and archiving of social science research data is provided on the **GESIS web pages**. Remaining questions and details are clarified during the pre-ingest phase.

Pre-ingest communication with data producers adheres to important steps of the Preliminary and Formal Definition Phases as outlined in the Producer-Archive Interface Methodology Abstract Standard (PAIMAS) standard (CCSDS 2004). The Preliminary Phase is designed to gather necessary information, establish contact with data producers, and to provide the data producers with all necessary information – in particular about the archiving services offered by GESIS, the minimal requirements that have to be fulfilled for GESIS to accept data for archiving (see below), as well as licensing options and the archive agreement. During the Formal Definition Phase, agreements and schedules sketched before are discussed in more detail and on a more concrete level. In this phase, **subject-related, technical, and legal questions are clarified, and agreements made with regard to delivery formats, schedules**, etc. This phase concludes with the **signing of the archive agreement**.



The Data Archive 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).

Comprehensive (general) **information on data management, data archiving, data sharing and re-use, as well as specific information on how to deposit with the GESIS Data Archive is given on the GESIS web pages**. Beyond that GESIS actively promotes and supports all aspects of research data management, including e.g. long-term preservation, re-use, citation, through own research & development, publications, presentations, as well as active participation in a wide range of initiatives and projects.

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.



- .. Data Archive Website: <http://www.gesis.org/en/services/archiving-and-registering/data-archiving/>
- .. PAIMAS standard (CCSDS 2004): <http://public.ccsds.org/publications/archive/651x0m1.pdf>
- .. Archive agreement:
http://www.gesis.org/fileadmin/upload/institut/wiss_arbeitsbereiche/datenarchiv_analyse/Archivierungsvertrag_GESIS_Datenarchiv_v9__englisch.pdf
- .. General Advice on Data Management & Archiving: <http://www.gesis.org/en/archive-and-data-management-training-and-information-center/training-center-home/>
- .. Jensen, Uwe (2012): Leitlinien zum Management von Forschungsdaten, Sozialwissenschaftliche Umfragedaten. GESIS Technical Reports, 2012/07:
http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis_reihen/gesis_methodenberichte/2012/TechnicalReport_2012-07.pdf
- .. datorium: <https://datorium.gesis.org/xmlui/?locale-attribute=en>
- .. Usage regulations: <http://www.gesis.org/en/services/data-analysis/data-archive-service/usage-regulations/>
- .. Data Catalogue and other portals: <http://www.gesis.org/en/services/research/>
- .. Example of a study description (European Values Study 2008: Germany (EVS 2008)) in the Data Catalogue:
<https://dbk.gesis.org/dbksearch/SDesc2.asp?no=4753&tab=0&ll=10¬abs=&af=&nf=1&search=evs&search2=&db=E>

All URLs accessed 07 April 2014

- K1 Auswahl der Informationsobjekte und ihrer Repräsentationen
- K2 Verantwortung für den Erhalt
- K3 Zielgruppen
- K4 Zugang
- K5 Interpretierbarkeit
- K6 Rechtliche und vertragliche Basis
- K7 Rechtskonformität
- K8 Finanzierung
- K9 Personal
- K10 Organisation und Prozesse
- K11 Erhaltungsmaßnahmen
- K12 Krisen-/Nachfolgeregelung

- K13 Signifikante Eigenschaften
- K14 Integrität: Aufnahmeschnittstelle
- K15 Integrität: Funktionen der Archivablage
- K16 Integrität: Nutzerschnittstelle
- K17 Authentizität: Aufnahme
- K18 Authentizität: Erhaltungsmaßnahmen
- K19 Authentizität: Nutzung
- K20 Technische Hoheit
- K21 Transferpakete
- K22 Transformation der Transferpakete in
Archivpakete

K23 Archivpakete

K24 Interpretierbarkeit der Archivpakete

K25 Transformation der Archivpakete in Nutzungspakete

K26 Nutzungspakete

K27 Identifizierung

K28 Beschreibende Metadaten

K29 Strukturelle Metadaten

K30 Technische Metadaten

K31 Protokollierung der Langzeiterhaltungsmaßnahmen

K32 Administrative Metadaten

K33 IT-Infrastruktur

K34 Sicherheit

K1 Auswahl der Informationsobjekte und ihrer Repräsentationen

Erfüllungsgrad: implementiert

Punkte: 10

Umsetzung: Grundlage der Sammlung ist das Gesetz über die Deutsche Nationalbibliothek (DNBG). Der Sammelauftrag ist präzisiert in der Pflichtablieferungsverordnung (PflAV). Die gesetzlichen Bestimmungen zum Sammelauftrag werden durch die Sammelrichtlinien weiter erläutert und präzisiert.

- .. Dokumentation 1: DNB-Gesetz
- .. Dokumentation 2: Pflichtablieferungsverordnung
- .. Dokumentation 3: Sammelrichtlinien

- ..■ Vertrauen schaffen durch Transparenz
 - ..■ Interessierte können sich ein Bild darüber machen, wie das Archiv die digitalen Objekte behandelt
 - ..■ Vertrauen einwerben bei den unterschiedlichen Stakeholdern: Zuwendungsgeber, Datengeber, Datennutzer
 - ..■ Das Archiv kann darstellen, dass es die grundlegenden Funktionalitäten kennt und angemessen berücksichtigt



- ..■ Qualitätssicherung

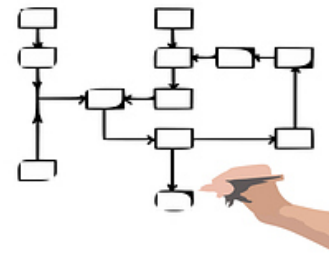
- ..■ Ordnung schaffen:

Überblick über vorhandene und fehlende Dokumente/Dokumentationen

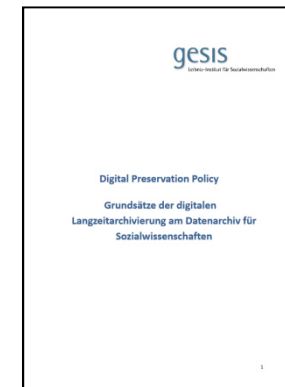


- ..■ Aufdecken der Stärken, Schwächen und Lücken

- ..■ Strukturierte und standardisierte Überprüfung der Workflows



- ..■ Implementierung von neuen Workflows (z.B. Monitoring der Prüfsummen)
- ..■ Überarbeitung der Außendarstellung (z.B. Webseite)
- ..■ Erstellung neuer oder angefangener Dokumente (z.B. Preservation Policy, Versionierungsregeln)



Ausgehend davon, dass Langzeitarchivierung kein abgeschlossener Prozess ist, macht es Sinn, die eigenen Workflows und Maßnahmen regelmäßig zu überprüfen und anzupassen. Da sich die Standards ebenfalls an neuen Entwicklungen orientieren, können sie diese Prozesse unterstützen.

- ..■ European Framework for Audit and Certification of Digital Archives:
<http://www.trusteddigitalrepository.eu/Welcome.html>
- ..■ Data Seal of Approval:
<http://www.datasealofapproval.org/en/>
- ..■ nestor-Siegel:
http://www.langzeitarchivierung.de/Subsites/nestor/DE/nestor-Siegel/siegel_node.html
- ..■ ISO 16363: <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>
- .. Internationale Task Force: Entwicklung von TRAC: **Trustworthy Repositories Audit & Certification**, 2007
<http://www.repositoryaudit.eu/about/>
- .. Repositories Audit und Certification RAC, Fortsetzung TRAC, Consultative Committee for Space Data Systems (CCSDS)
<http://public.ccsds.org/publications/archive/652x0m1.pdf>
- .. Digital Repository Audit Method Based on Risk Assessment (DRAMBORA) -> Drambora Online Toolkit
<http://www.repositoryaudit.eu/about/>

Natascha Schumann

GESIS

Datenarchiv für Sozialwissenschaften

www.gesis.org

natascha.schumann@gesis.org