Tried and trusted: Experiences with certification processes at the GESIS Data Archive

Natascha Schumann
GESIS Data Archive
GESIS

- Infrastructure institution for the Social Sciences
- GESIS provides services in the fields of
  - Data: archiving, research, social monitoring
  - Methodology: consultation, developing complex methods
  - Information: social science literature and research activities
Research life cycle

GESIS performs research and offers consulting services and workshops to support all phases.
GESIS Data Archive

- provides excellent data service for national and international comparative surveys from the fields of social and political science research
- keep social science research data available for data sharing and data re-use
- Explicit mission for archiving and long-term preservation
Archiving workflow
Why certification activities?

• Ensuring stakeholders that the data they are using, producing or funding is treated according to established standards and norms

• Need to prove trustworthiness is not only an issue for new players but also for established ones like the GESIS Data Archive

• Trustworthiness among European Archives (CESSDA)
Choosing a standard
European Framework for Audit and Certification

- CCSDS/ISO Repository Audit and Certification Working Group (RAC)
- Data Seal of Approval (DSA)
- DIN/nestor Working Group “Trustworthy Archives – Certification”

signed a Memorandum of Understanding
MEMORANDUM OF UNDERSTANDING
TO CREATE A
EUROPEAN FRAMEWORK FOR AUDIT AND CERTIFICATION OF DIGITAL
REPOSITORIES

THIS AGREEMENT is made between:

- David Giaretta in his capacity as chair of the CCSDS/ISO Repository Audit
  and Certification Working Group (RAC),
- Henk Harmsen in his capacity as Chair of the Data Seal of Approval (DSA)
  Board and
- Christian Ketel in his capacity as Chair of the DIN Working Group
  "Trustworthy Archives – Certification"

WHEREAS the parties to this Memorandum of Understanding all lead separate
groups aiming at certifying digital repositories, they wish to put in place mechanisms
to ensure that the groups can collaborate in setting up an integrated framework for
auditing and certifying digital repositories.

The framework will consist of a sequence of three levels, in increasing
trustworthiness:

- **BASIC CERTIFICATION** is granted to repositories which obtain DSA certification;

- **EXTENDED CERTIFICATION** is granted to Basic Certification repositories which
  in addition perform a structured, externally reviewed and publicly available
  self-audit based on ISO 16363 or DIN 31644;

- **FORMAL CERTIFICATION** is granted to repositories which in addition to Basic
  Certification obtain full external audit and certification based on ISO 16363 or
  equivalent DIN 31644.

This process is supported by and coordinated with the help of the European
Commission.

DURATION OF THIS AGREEMENT:
This agreement shall commence upon the signing of this agreement by all the parties
and will end on mutual agreement of the parties.

SIGNED on behalf of the parties by:

1. [Signature]
   date: 07/07/2010
   David Giaretta

2. [Signature]
   date: 07/07/2010
   Henk Harmsen

3. [Signature]
   date: 07/07/2010
   Christian Ketel
Our approach

1. Data Seal of Approval within the context of CESSDA-AS: Agreement that every archive should make a self-assessment based on the DSA

2. Data Seal of Approval

3. nestor/DIN 31644

4. ...
Repositories that are committed to archiving and providing access to data in a sustainable way.

16 Guidelines reflecting different roles:

• Data producers
• Data repository
• Data users
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.
DSA: level of compliance

- 0 - Not Applicable
- 1 - No
- 2 - Theoretical
- 3 - In Progress
- 4 - Implemented
Getting started

• Evaluation of existing workflows with regard to the requirements of the chosen level of certification
• Detection of missing documentation (internal as well as external)
• Detection of missing documents (e.g. policies, recommendations etc.)
Getting started

• Workflows themselves are already in place (and work), but there is a lack of appropriate documentation, especially on our website

• Many documents are there but not up to date or not yet ready for publication
Documents…

… have to be created and made available to the public, e.g. a **Preservation Policy**

Challenge: provide enough information without becoming too detailed as the policy should not be changed frequently
Redesign of website

More (detailed) information
• About the archiving workflow
• About legal aspects
• About concepts and standards
• …
→ Work in progress
Experiences

• Get an overview where we are standing
• Reconsider workflows
• Update processes with regard to digital preservation
  – Traditional scope of our work on curating empirical social research data – no emphasis on digital preservation
Experiences

• Become aware of the capabilities, strengths and most important weaknesses or gaps within the institution
• Time intensive
• Many processes and activities at the same time depending on each other
Benefits

• Valuable effect of establishing a common understanding for the mission and goals of an institution

• Ensure fulfillment of the requirements of being a trusted digital repository and display this to the public
Any questions?
natascha.schumann@gesis.org