Tried and Trusted

Experiences with Certification Processes at the GESIS Data Archive by Natascha Schumann[1]

GESIS

Abstract

The need to prove and improve trustworthiness is an issue not only for new archives but also for those who have been doing this job successfully for a considerable time. But what happens when an established data archive faces the challenges of certification and audit processes? Founded in 1960, the GESIS Data Archive for the Socials Sciences has been "in the business" for more than 50 years and is one of the oldest archives in Germany to preserve electronic resources for the long term. Driven by a growing awareness of the needs of its stakeholders, who have to be sure that the data they produce, use, or fund is treated according to common standards, the Data Archive started a process of audit and certification within the European Framework for Audit and Certification.

After giving an overview of the GESIS Data Archive and the European Framework for Audit and Certification, this article describes how existing workflows were evaluated with regard to the requirements of the chosen level of certification. While the workflows themselves are already in place, in some cases the evaluation process showed a lack of appropriate documentation. Suitable documents have to be created and made available to the public. In some cases, this process has to be accompanied by discussions within the institution about the mission and goals of the archive. In our experience, these can be very productive and lead to a common understanding and an improvement of services[2].

Keywords: Digital preservation, trusted digital repositories, certification.

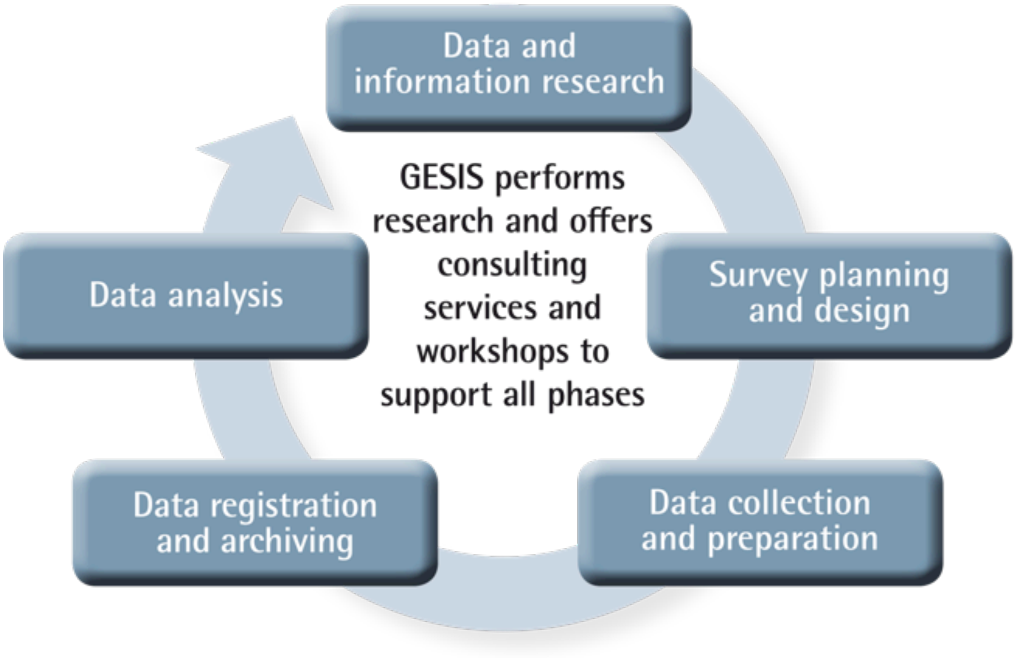
Data and information research
Survey planning and design
Data collection and preparation
Data registration and archiving
Data analysis
GESIS performs research and offers consulting services and workshops to support all phases

Figure 1: Research Data Life Cycle

The GESIS Data Archive and its organizational context

As an infrastructure institution for the social sciences, GESIS not only carries out research but provides services in all phases of the research data life cycle (see figure 1), from the conception to the archiving and re-use of social science research. Among others, it provides consultation in methodology, develops software tools for research, and offers information services.

The GESIS Data Archive for the social sciences is one of five departments of GESIS and has been providing comprehensive data services for national and international comparative surveys for several decades. One of its main tasks is to make research data available for re-use. To support this goal, the data archive has an explicit mission for long-term preservation, which is also laid down in GESIS's by-laws. Accordingly, among GESIS'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" (GESIS Constitution § 2).3

Workflows within the archive are organized according to an archival life cycle, ranging from pre-ingest (incl. acquisition) to ingest and processing to archival storage up to the dissemination of data. The central functions of the OAIS reference model (CCSDS, 2012) can be mapped to the existing structure of the archive (see Schumann and Recker, 2013).

However, although the GESIS data archive already has working procedures and processes in place to ensure the preservation of its data, there is still a need for further activities. For example, in some cases documentation of workflows and defined interfaces between different steps of the preservation process are lacking, and some definitions of information packages are not up to date. By addressing these issues in a systematic fashion, the archive aims to further increase its trustworthiness.

A need for trust

A definition of a trusted digital repository is given by the RLG/OCLC Working Group: "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" (Research Libraries Group, 2002, p. i). But what does that mean in detail? Audit and certifications standards such as the nestor Catalogue of Criteria for Trusted Digital Repositories on the one hand add aspects from an IT security perspective – for example, "authenticity, integrity, confidentiality and availability" (nestor, 2009, p. 1). These technical aspects are relevant issues for trust, but beyond that organizational aspects are just as important. As pointed out in Audit and Certification of Trustworthy Repositories (CCSDS, 2011), "[c]onstant monitoring, planning, and maintenance, as well as conscious actions and strategy implementation will be required of repositories to carry out their mission of digital preservation" (p. 2-1). These quotations show that building trust depends on more than one factor. A trusted digital repository has to ensure that the digital objects it preserves are not corrupted by accident or intentionally, and that access is given – not only physically, but also in appropriate digital formats. Another criterion of trust is if and how the organization demonstrates its know-how in digital preservation and, for example, if succession plans exist for the case that the institution ceases to exist. Thus, transparency is very important in the context of trust. All stakeholders should have the opportunity to ascertain the statements made by the institution.

Accordingly, to appear trustworthy, the GESIS data archive has to provide stakeholders – data depositors, data users and funders – with sufficient information to demonstrate that their data is treated according to the agreed standards of the social sciences and digital preservation communities. Because existing certification standards and audit tools support archives in the building of trust, we decided to start a process of audit and certification within the European Framework for Audit and Certification (see below). Our decision to do so coincided with similar efforts initiated by the Council of European Social Science Data Archives (CESSDA), of which GESIS is a member. As CESSDA has been transformed into a new organization and legal form, CESSDA AS, and is on its way to becoming a European Research Infrastructure Consortium (ERIC), it is necessary that all member institutions agree on the same standards regarding trustworthiness. To start off this process, during 2013 all archives carried out a self-assessment based on the guidelines of the Data Seal of Approval (DSA; see below).

All of this illustrates that the need to prove trustworthiness is not only an issue for new players, but also for established ones like the GESIS Data Archive. However, the challenges such "established players" face are somewhat different from those that new archives have to deal with: It is a different kind of procedure to set up a completely new service or to conduct a certification process in an existing system. Thus, when setting up a new archive it is possible to take into account the requirements for trusted digital repositories from the outset. What is more, new archives can benefit from other institutions and their experiences and avoid mistakes. In contrast, an existing archive may have gained a lot of expertise and know-how over time, but it can be very complex and challenging to adapt established workflows to new requirements.

European Framework for Audit and Certification of Trusted Repositories

Over the years, many different approaches and standards have been developed in the field of audit certification for trusted digital repositories. The most established among them are

- the Data Seal of Approval (DSA), originally initiated by DANS,
- the nestor Catalogue of Criteria for Trusted Digital Repositories, which became a German DIN standard (DIN 31644) in 2013 and will also be available in English, and
- the Trusted Repository Audit Checklist (TRAC), which is also an ISO standard (ISO 16363).

To achieve greater harmonization between these different initiatives and criteria catalogues, a Memorandum of Understanding (MoU) was signed in 2010 for a European Framework for Audit and Certification[4]. This process was accompanied by the European Commission and the Alliance for Permanent Access to the Records of Science (APARSEN) [5].

The MoU defines three levels of certification (see figure 2):

1. Basic Certification is granted by obtaining the DSA.
2. Extended Certification requires completing the DSA and an externally reviewed self-audit based either on ISO 16363 or DIN 31644.
3. Formal Certification requires completing the DSA and a full external certification based either on ISO 16363 or DIN 31644.

Data Seal of Approval

The target audience of the DSA are repositories committed to long-term preservation. Working from the assumption that data quality is dependent on "aspects related to the creation, storage and (re-)

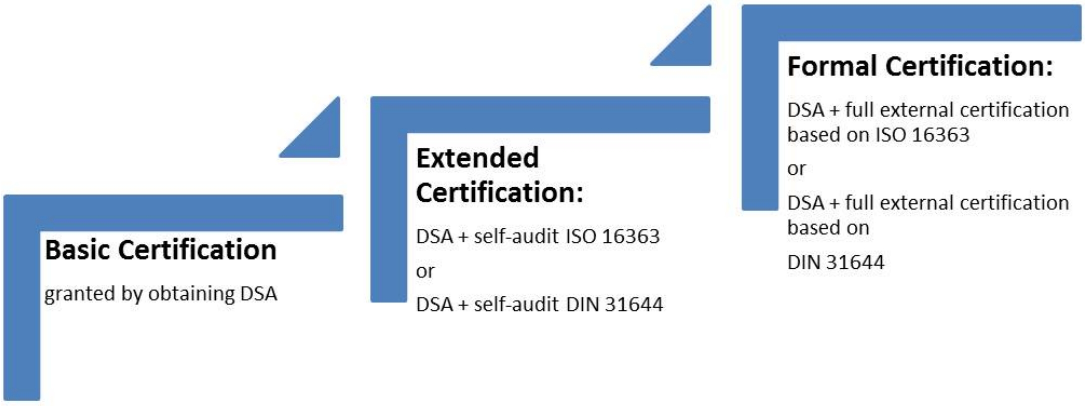
Basic Certification
granted by obtaining DSA
Extended Certification:
DSA + self-audit ISO 16363
or
DSA + self-audit DIN 31644
Formal Certification:
DSA + full external certification based on ISO 16363
or
DSA + full external certification based on
DIN 31644

Figure 2: Three Levels of Certification within the European Framework for Audit and Certification

use of digital data" (DSA, 2013, p. 5), the DSA contains 16 Guidelines reflecting different roles: Data producers, data repository and data users. Although the main focus of the DSA is on data repositories, it is open to other digital archives as well. There are different levels of compliance for each guideline:

0	Not Applicable
1	No. We have not considered it yet
2	Theoretical. We have a theoretical concept
3	In Progress. We are in the implementation phase
4	Implemented. This guideline has been fully implemented for the needs of our repository

To be awarded the DSA, the minimum compliance level as stated in the DSA guidelines has to be reached by the applicant (see DSA, 2013, p. 6). The Archaeology Data Service has published a best practice report to support other institutions in obtaining the DSA (Mitcham and Hardman, 2011).

nestor Seal/ DIN 31644

The DIN 31644/nestor Seal is based on the nestor Catalogue of Criteria for Trusted Digital Repositories (2009). In 2012 it was accepted as the German Standard DIN 31644. It contains 34 criteria covering the following thematic areas: organizational framework, handling of information objects and their representations, infrastructure and security. The level of compliance is measured on the following scale:

0	No concepts are in place
3	A concept exists
6	Well-elaborated concept
10	Implemented

RAC/ISO 16363

The Repositories Audit Checklist (RAC) was developed from the Audit Checklist for the Certification of Trusted Digital Repositories (2005) and Trustworthy Repositories Audit and Certification: Criteria and Checklist (2007). It became an ISO standard in 2011 and information about the standard can be found at the Primary Trustworthy Digital Repository Authorisation Body (ISO-PTAB)[6]. RAC consists of 50 main criteria and has 109 criteria in total. Their structure is orientated towards the nestor Catalogue and accordingly RAC criteria cover the following areas: organizational infrastructure, digital object management, infrastructure and security risk management.

Our Approach

As stated above, the GESIS Data Archive decided to commence its certification activities with the DSA. Several reasons contributed to this decision. First of all, the DSA is a good starting point for certification activities because it addresses all basic aspects of trust but is not as detailed as either DIN 31644 or ISO 16363. Thus the DSA is an immensely helpful tool to gain an overview of the processes within our archive. It therefore helps us lay the groundwork for follow-up activities in audit and certification within the context of the European Framework.

Getting started

As the DSA is a self-assessment, the applicant completes a self-assessment statement for each of the guidelines including links to the relevant documentation or evidence (see DSA, 2013, p. 6). This will then be reviewed by a peer reviewer appointed by the DSA-Board.

As a first step we evaluated the existing workflows with regard to the requirements stated in the guidelines. In this manner we obtained an overview of those workflows already in place and supported by sufficient documentation. This initial evaluation showed that the majority of our workflows comply with the DSA guidelines. However, a lack of appropriate documentation, especially on our website, became apparent.

In consequence, the main tasks at this stage were:

- The detection and subsequent creation of missing documentation and documents, e.g. policies or recommendations.
- The revision of existing documentation and documents if those were not up to date or not yet ready for publication. E.g. a versioning policy had to be updated.

An example for a newly created document is our preservation policy. It contains information on the organizational context of the GESIS Data Archive, states our mission, and describes the main principles of our approach to the digital preservation of data (see also Friese in this issue). The process of developing the policy not only meant creating the content, it was also a process requiring a good deal of coordination: staff members from different teams had to be involved as well as the head of the department. As it is insufficient for the certification process to publish the policy only in German, an English translation was also created.

As explained above, an important requirement for trustworthy digital repositories is transparency: all relevant information should be available to the public. In our case this meant redesigning the archive's website and deciding which information it should include. In this process, we reconsidered the whole structure of the website. It is now organized along the steps of the data lifecycle and refers to the functional entities in the OAIS reference model. Creating the website content was another challenge and involved more than simply adding some documents. We had to answer the question of how to make the website helpful for the different stakeholders and user groups: data producers, data users, funders, and other interested persons – both in terms of content and the language used.

Experiences and Benefits

One of the (first) benefits of the preparatory work on the DSA was that we gained an overview of what we already have in place and what will have to be amended or improved. It became clear that we would have to reconsider some of our workflows.

Traditionally, the focus of the GESIS Data Archive has been the indexing and processing of empirical social research data. Over the past years, more and more digital preservation issues became relevant – for example, questions of how to implement procedures to ensure authenticity and integrity, or the need to define different archival packages referring to the OAIS model etc.. But not only did the archive have to update workflows; an important part of this process (which is not completed yet) was the development of a common understanding of digital preservation: Is it an "an added value" that we somehow create on top of everything else that the archive does? Or is it not rather the sum of everything we do in the archive: the bundle of measures, that is, that we employ to ensure access and use of the data in the long term, including, for example, the creation of metadata and documentation, DOI registration, etc.?

In addition, the strengthened focus on digital preservation has consequences not only with regard to processes and procedures but also for the level of transparency. Stakeholders have an increasing interest in learning how their data is curated and preserved, and the archive has to provide the respective evidence in order to maintain its stakeholders' trust.

However, the DSA application not only required us to address external stakeholder communication; it also prompted a review and evaluation of our internal communication and documentation procedures. For example, the archive staff uses a wiki for internal documentation purposes. It was built and filled with content over the last years, but our review in relation to the DSA application showed that it was not up to date – neither with regard to the contents nor to the structure of our workflows. We are now in the process of restructuring and updating it, which also includes agreeing on the current procedure of maintaining the wiki and keeping it up to date.

The processes set in motion by our decision to apply for the DSA have helped to create awareness of the capabilities and strengths as well as of weaknesses or gaps within our archive. The systematical compilation of existing and relevant information and documentation required by the DSA is a helpful step in itself, and this gap analysis has helped us gain a more concrete idea of our workflows and their documentation instead of the vague feeling that "surely, everything works as it is supposed to."

Preparing the DSA application served as an incentive to continue some projects – e.g. for new services or the adoption of standards – that had been planned for some time but had been neglected in the face of "more pressing" problems arising in our day to day work. Some of the required measures are easily created and implemented, but others need more time and discussion to be realized. The fact that so many activities are linked to each other entails that it may take some time to implement new processes: the necessary changes concern different applications or workflows cutting across different teams and cannot be made without implications for other parts of the system.

The process of (preparing) an audit or certification is very time consuming. But in our experience, the process of preparing the DSA self-assessment statement produced many valuable effects in that it helped us establish a common understanding for the mission and goals of our archive. The first steps of complying with the guidelines of the DSA have been made and we have gained an overview of our capabilities as well as existing gaps. Accordingly, we now know where we stand and what our tasks for the near future are. Our next step will be to hand in the DSA application. After this is completed and the DSA will have been granted, we are already planning to take the next step in the European Framework for Audit and Certification: the extended certification, which will take the form of conducting a self-audit for the nestor Seal, based on DIN 31644.

References

CCSDS, 2011. Audit and Certification of Trustworthy Digital Repositories. [pdf] Available at: <http://public.ccsds.org/publications/archive/652x0m1.pdf> [Accessed 03 December 2013]

Consortium of European Social Science Data Archives (CESSDA-AS). Homepage. [online] Available at: <http://www.cessda.net/> [Accessed 03 December 2013]

Data Seal of Approval: Homepage. [online] Available at: <http://www.datasealofapproval.org/en/> [Accessed 03 December 2013]

DIN, 2012. DIN 31644: Kriterien für vertrauenswürdige digitale Langzeitarchive. Available at: <http://www.nabd.din.de/cmd?level=tpl-art-detailansicht&committeeid=54738855&artid=147058907&languageid=de&bcrumblevel=3> [Accessed 03 December 2013]

Friese, Y., 2014. How to develop a preservation policy? Guidelines from the nestor working group. [pdf] In: IASSIST Quarterly [include information for this issue].

Mitcham, J. and Hardman, C., 2011. ADS and the Data Seal of Approval – case study for the DCC. [online] Available at: <http://www.datasealofapproval.org/en/news-and-events/news/2011/9/14/ads-and-data-seal-approval-case-study-dcc/> [Accessed 15 November 2013]

nestor, 2009. nestor criteria: Catalogue of Criteria for Trusted Digital Repositories, Version 2. nestor materials 8. [pdf] Available at: <http://nbn-resolving.de/urn:nbn:de:0008-2010030806> [Accessed 03 December 2013]

Schumann, N. and Recker, A., 2013. De-mystifying OAIS compliance: benefits and challenges of mapping the OAIS reference model to the GESIS Data Archive. [pdf] IASSIST Quarterly, 36 (2), pp. 6-11. Available at: <http://www.iassistdata.org/downloads/iqvol36_2_recker_0.pdf> [Accessed 03 December 2013]

Research Libraries Group, 2002. Trusted Digital Repositories: Attributes and Responsibilities. An RLG-OCLC Report. [pdf] Available at <http://oclc.org/content/dam/research/activities/trustedrep/repositories.pdf> [Accessed 03 December 2013]

CRL and OCLC, 2007. Trustworthy Repositories. Audit & Certification: Criteria and Checklist. [pdf] Available at: <http://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf> [Accessed 03 December 2013]

Yakel, E., Faniel, I., Kriesberg, A., and Yoon, A., 2013. Trust in Digital Repositories. [pdf] The International Journal of Digital Curation, 8.1 (2013), pp. 143-156. Available at: <doi:10.2218/ijdc.v8i1.251> [Accessed 03 December 2013]

NOTES

1. Natascha Schumann is affiliated at the Data Archive for the Social Sciences at the GESIS Leibniz Institute for the Social Sciences in Cologne. The main focus of her work is on digital curation of social science research data and audit and certification in this area. Her contact email is natascha.schumann@gesis.org.
2. This paper is an updated version of a presentation given at the IASSIST 2013 conference in Cologne.
3. GESIS Constitution (in German): http://www.gesis.org/das-institut/der-verein/satzung/
4. http://www.trusteddigitalrepository.eu/Site/Welcome.html
5. http://www.alliancepermanentaccess.org/index.php/aparsen/
6. Primary Trustworthy Digital Repository Authorisation Body (ISO-PTAB): http://www.iso16363.org/