Improving the Quality of Digital Preservation Using Metrics

by Mari Kleemola[1]

Abstract

The Finnish Social Science Data Archive (FSD) is dedicated to preserving digital research data in the social sciences and to providing good quality services to the research community. Since the research data landscape is perpetually changing, digital repositories need to stay alert and be ready to review their policies and procedures, continuously. In FSD's case, it is critical that our operations and procedures are up-to-date and consistent with relevant standards and best practices, and that our stakeholders trust us.

In this paper we outline FSD's venture into the world of digital preservation standards and assessments. We started by exploring the key standard for long term preservation of digital data, the Open Archival Information System (OAIS). We then proceeded to conduct a self-assessment within the framework of the Audit and Certification of Trustworthy Digital Repositories (TDR) Checklist, followed by the CESSDA Trust Process which resulted in FSD's successful application for the Data Seal of Approval certification. The process has been fruitful and we can safely say that in the case of using metrics to improve quality, it is the journey that matters as much, if not more, than the destination.[2]

Keywords: digital preservation, certification, OAIS, trusted digital repositories, assessment, DSA

The Finnish Social Science Data Archive FSD

The Finnish Social Science Data Archive (FSD) is a national resource centre that promotes open access to research data as well as transparency, accumulation and efficient reuse of scientific research data. FSD was established in 1999 to archive and disseminate digital, quantitative data for social science research, teaching and learning. Over the years, it has broadened its services to include qualitative data archiving as well as provision of guidance on research ethics and data management.

In many ways the turn of the century proved to be an excellent time to set up a national data archive. In the late 1990s the Internet was already a daily working tool for social scientists, and standards like the Data Documentation Initiative were emerging. Most importantly, the international social science data archiving community was well established and networked, having become a part of the research scene in the 1950s and 1960s. The various social science data archives were actually the very first institutions to handle and preserve digital material (Doorn and Tjalsma 2007). Our new archive was privileged to learn from the practices and experiences of the pioneers of digital data archiving.

Early on we realised that preservation, management and dissemination of datasets as well as building high-quality knowledge-based services needed to be done in a transparent way, at the most appropriate time possible and without over-burdening researchers. It was also clear that for the outcome to be successful, many organisational issues and pieces needed to be in place, including policies, procedures and sustainable resources. Consequently, we paid attention to documenting our processes and procedures from the

very beginning: our first Internal Handbook was created during the first two years of operation and our first Archives Formation Plan, or AMS, emerged in 2003. The AMS is FSD's highest-ranking document concerning provision of data services. It is based on guidelines set by the National Archives of Finland and it describes tasks and processes, selection criteria, preservation periods and forms, confidentiality and legislative issues, responsibilities, and data systems and security.

Nowadays, FSD is an acknowledged and active national centre of expertise in the areas of preserving and providing access to digital research data. FSD's core user community consists of researchers, teachers and students from Finland and abroad. Most FSD's services, such as the Data Management Guide, are openly and freely available via our website and in 2014 the number of successful web page requests reached 1.2 million. Access to FSD's data holdings is provided via the Aila Data Portal. In June 2015, Aila contained 1200 datasets and had 1300 registered users. FSD is funded by the Ministry of Education and Culture and operates as a separate unit at the University of Tampere. FSD is Finland's Service Provider for the pan-European Research Infrastructure, CESSDA[3].

Measuring Trustworthiness of Digital Preservation

Hedstrom (1998) defines digital preservation as 'the planning, resource allocation, and application of preservation methods and technologies necessary to ensure that digital information of continuing value remains accessible and usable'. Digital preservation is therefore an active, and in the optimal case, even a pro-active process that does not include extended periods of inactivity. Digital preservation is also something that needs to be done presently for the future; we need to think about a time period long enough to be concerned with the impacts of changing technologies or a changing user community (CCSDS 2009; Giaretta 2011).

The demands these definitions make are powerful and illustrate well the difficulties in providing long-term access to digital data. Preserving digital data is a challenge. For the outcome to be successful, many organisational and practical issues need to be in place. Key aspects in demonstrating trustworthiness include: transparency, documentation, adequacy and information security. In addition, one has to keep in mind that one needs to evaluate trust into the future. (Dobratz et al. 2010; Giaretta 2011.)

For quite some time it has been evident that methods are needed to assess the trustworthiness of a digital repository. In 1996, the Commission on Preservation and Access and the Research Libraries Group called for a certification programme for repositories claiming to serve an archival function, and since then several stakeholders have explored certification issues (see Dobratz et al. 2010). The need to be able to test a repository's claims about digital preservation was also one of the key drivers of the Open Archival Information System (OAIS) Reference model (Giaretta 2011, 461). In the European data archive world assessment and trust issues came up roughly ten years ago. In 2006, The Council of European Social Science Data Archives (CESSDA) Research Infrastructure was identified as an existing pan-European RI recommended for a major upgrade by the European Strategy Forum on Research Infrastructures Roadmap. As a direct result of this, CESSDA launched the Preparatory Phase Project (PPP) in 2008. The project recommended, amongst other things, that adherence to standards should be a part of CESSDA ERIC's membership criteria since common use of standards is necessary for compatibility. As possible tools, the project brought forward the OAIS model in particular, and the DSA criteria. (Dusa et al. 2010.)

The OAIS model (ISO14721:2003) is the international key standard for archival systems and for organisations engaged in long-term preservation of digital data (see, for example, Spence 2006; Lavoie 2004; Giaretta 2011). Prior to the CESSDA PPP project, the OAIS model had been explored by the UK Data Archive (Beedham et al. 2005) and the ICPSR (Vardigan and Whiteman 2007). A Finnish version of the standard was published in March 2010 by the Finnish Standards Association SFS[4].

The OAIS model introduced important concepts and paved the way to a certification standard for digital repositories. OAIS conformance is necessary for trustworthiness but not sufficient since the OAIS model is very general and does not cover, for example, financial aspects. Metrics were therefore needed. The Trustworthy Repositories Audit and Certification Checklist (TRAC) was released in 2007, and its revised version, entitled the Trusted Digital Repository (TDR) Checklist, in 2011. The Checklist derives from the OAIS, and the ISO 16363:2012 standard[5] is based on it. The first edition of the Data Seal of Approval, DSA, emerged in 2008. It was initially developed for use in the Netherlands, but already in 2009 an international DSA Board was established. The DSA contains 16 guidelines, and it is the first step in the three-level European Framework for Audit and Certification of Digital Repositories[6] that was developed in 2010. Other certification initiatives include, for example, the Nestor Catalogue of Criteria (DIN 31644) and the DRAMBORA toolkit (for an overview of key guidelines and frameworks, see Kvalheim et al. 2013).

All these developments prompted us to take a critical look at our functions at FSD. We wanted to find out if they were adequate and up-to-date and hoped to find ways to rationalise our processes. We also wished to demonstrate that we can be trusted by our key stakeholders: data depositors, users and funders. Furthermore, an important goal was to ascertain that FSD would able to fulfill the CESSDA membership criteria. To achieve all this, we decided to use suitable standards and metrics to assess our systems, policies and procedures.

Step One: Start by OAIS

Our first step was to familiarise ourselves with the OAIS model in 2010. An OAIS compliant organisation has to support the information model of the standard and fulfil the six minimum requirements. Thus, the organisation must:

- negotiate for and accept appropriate information from information producers,
- obtain sufficient control of the information provided to the level needed to ensure long-term preservation;
- determine, either by itself or in conjunction with other parties, which communities should become the designated community and, therefore, should be able to understand the information provided;
- ensure that the information to be preserved is independently understandable to the designated community;
- follow documented policies and procedures which ensure that the information is preserved against all reasonable contingencies, and which enable the information to be disseminated as

authenticated copies of the original, or as traceable to the original; and
- make the preserved information available to the designated community. (CCSDS 2009.)

Fulfilling these general requirements proved straightforward. FSD negotiates with researchers when they are depositing their data, and FSD's designated community is stated in the regulations as social science researchers, teachers and students. FSD's Archives Formation Plan (AMS) contains specific information on FSD's tasks (including selection criteria, archival process and data protection practices) and the internal manual contains detailed practical instructions. Archived research data are processed and described in compliance with international standards and formats, and the research community is able to find information about the data on the FSD website.

The OAIS standard also contains a functional model consisting of six main entities: ingest, archival storage, data management, administration, preservation planning, and access. Additionally, an organisation has to provide common services, such as various technical support services. Exploring these functional entities in detail provided us with a new perspective about our processes. All the entities were identifiable in FSD's operations and luckily we were not able to find any serious defects. This in turn resulted in a strengthened conception that we had been doing the right things in the right way. However, it became clear that our processes could be further clarified and improved and that we should, for example, collect and store more extensive preservation information and to better structure it.

The OAIS compliance was expected for many reasons. First, FSD has been established to function as a repository for digital data, and also FSD's operational model has been based on the examples of well-established social science data archives. All in all, FSD emphasises functions slightly differently in comparison to the OAIS model. While the OAIS describes ingest only briefly and practically omits acquisition, they are central processes in the FSD. On the other hand, the OAIS presentation of administration and preservation planning is more complicated than FSD's procedures, mainly because FSD is still a relatively small archive. As FSD grows, it must be prepared to adopt more sophisticated administrative processes.

Our experiences about OAIS were very similar to those reported by other data archives, for example by the UK Data Archive (Beedham et al. 2005) and the ICPSR (Vardigan and Whiteman 2007). Also GESIS has carried out a mapping of the OAIS functional model to GESIS's operations (Schumann and Recker 2012).

Step Two: TDR Checklist

The results of our OAIS exercise were encouraging. However, the OAIS model is very high-level and thus did not provide enough concrete details to support the assessment of our day-to-day practices, which we thought would help us ensure the quality of our practices and processes and consequently the quality of our data services. Therefore, in spring 2012, we continued our assessment exercise with the help of the TDR Checklist. This decision was influenced by the UK Data Archive's draft audit against the emerging ISO 16363 standard (see Woollard, 2011). At the time of our self-assessment, the TDR Checklist was at its final stages of approval as ISO 16363 and was available as CCSDS 652.0-M-1 Magenta Book (CCSDS 2011). We used both the Magenta Book and the preliminary version of the TDR Checklist in Excel format provided by the Center for Research Libraries. The Checklist is designed to cover all the aspects necessary to demonstrate that an organisation can be trusted by its stakeholders. The main sections include organisational infrastructure, digital object management and infrastructure, and security risk management.

For each of the 100+ TDR clauses, we identified and described the written evidence and assessed our compliance using the scale of 0 (not compliant) to 3 (fully compliant). We found FSD to be well compliant (3) or almost compliant (2) with 66% of the clauses and not compliant (0) or only somewhat compliant (1) with 10% of the clauses. Six clauses were thought to be out of scope for FSD. In 18% of the clauses we could not make a straightforward assessment because we did not understand them completely. This is probably at least partly due to the fact that in many cases FSD relies on a manual process and a set of systems whereas the TDR Checklist describes a larger and more automated system. Our estimation is that FSD would be not compliant (0) or only somewhat compliant (1) with most of the "unclear" criteria. However, not all criteria have the same importance or level of risk.

Examples of FSD's full compliance:
- mission statement exists
- collection policy exists
- preservation policies exists
- short and long term business planning in place
- appropriate deposit agreements exist
- minimum information requirements specified
- minimum descriptive information captured

Examples of FSD's non-compliance:
- no formal succession plan
- no commitment to regular self-assessment or external certification
- no documented process for testing for an understanding of the AIP Content Information
- no systematic analysis of security risk factors

The good news was that we found no major failures or risks. However, the analysis revealed several weak points. For example, early on in the process it became evident that FSD needed to clarify several processes and to especially improve the documentation that describes the technical infrastructure and security risk management. As a result of the self-audit, FSD made an action plan for various changes and improvements. Some minor adjustments were made immediately and major changes became subject to discussions and have been – or will be – implemented gradually.

Our motivation for the self-assessment was to improve our practices and to plan for the development of our processes. For this, the TDR Checklist worked very well, although the learning curve was rather steep and understanding what constitutes adequate evidence was sometimes difficult. The self-assessment was also rather time-consuming and therefore we decided not to go too deep into the clauses and evidence. Since the goal of this TDR self-assessment was not a certification we did not analyze in detail which of the non-conformities would be acceptable and which were not. Our Checklist assessment is also in draft state (for example, some texts are in English, some in Finnish) and while it is sufficient for internal use, it has not been published. So even though the self-assessment provided us with a lot of insight into

our processes and helped to recognise and correct deficiencies, due to the lack of transparency it is not very useful for building trust with our key stakeholders.

Step Three: CESSDA Trust Process

In 2013, CESSDA archives carried out a two-phase self-assessment exercise also known as the "CESSDA Trust Process" that aimed to progress the CESSDA archives in the area of Trusted Digital Repository status. As CESSDA is on its way to becoming an European Research Infrastructure (ERIC), all its Service Providers need to meet the membership criteria.

In the CESSDA Trust Process, the Data Seal of Approval (DSA) was selected as a reference point because it is the base-level in the European Framework for Audit and Certification of Digital Repositories. The 16 guidelines of the DSA allow for verifying quality aspects concerning the creation, storage, use and reuse of digital data, and they can be seen as a minimum set distilled from proposals like Nestor, Drambora and TDR. The guidelines focus on three stakeholder groups: the data producers, the data repositories, and the data consumers. A repository is designated a Trusted Digital Repository if it complies to the ten guidelines for repositories and if it enables data producers and data consumers to comply with their three guidelines (Data Seal of Approval 2014).

At the beginning of the CESSDA Trust Process, the DSA criteria were mapped to the CESSDA member obligations. After the mapping, all the European data archives undertook a DSA self-audit. The process was coordinated by an Expert Panel consisting of four people, representing the British UK Data Archive, the Dutch DANS, the German GESIS and the FSD. The Expert Panel counselled the archives during the self-audit, reviewed all self-assessments and provided feedback and a gap analysis. All in all, European data archives appeared to have good practices in terms of, for instance, data reusability. Long-term preservation of data is also well managed, although the documentation of practices was inadequate across many data archives.

During the CESSDA Trust Process, FSD gained valuable knowledge about the DSA as well as of certification in general. Discussions with other data archives clarified the meaning of the guidelines and helped to understand and identify the relevant evidence ie. the documentation that was needed to demonstrate that we meet the DSA criteria. Feedback from the self-assessment and reviews led us to further improve our documentation.

Step four: DSA Application

The CESSDA Trust Process confirmed that FSD was in a good position to apply for the DSA. In May 2014 we implemented major changes in our data services that took us from paper-based data ordering to an advanced online data download system, so we decided to postpone our DSA application until August 2014. This ensured that we had time to revise both our internal and external documentation to reflect the changed service model. Before we sent our application, we also translated some key documents, like our Archives Formation Plan (AMS), into English.

Our actual DSA application process was very straightforward and not time-consuming. It took only a day. This somewhat contrasts with the GESIS' experience (Schumann 2012) and reflects the fact that we already had many pieces in place following the OAIS, TDR and CESSDA Trust exercises. For example, we had all the necessary documentation readily available to support our assertions, most of it in both Finnish and English. It is also worth mentioning that FSD had perceived good documentation as a cornerstone of its operations from the beginning so our documentation is the result of years of collective, continuous, and innovative work by FSD experts.

Our documentation consists of three main components: our website, the Archives Formation Plan (AMS) and the Internal Manual. FSD's website contains detailed information on archived data, archiving and ordering data, and on managing research data. The Archives Formation Plan includes, for example, our preservation plan, ingest criteria, and data protection practices, as well as the Internal Manual contains very detailed practical guidelines for all processes. The extensive and well-maintained documentation ensures, for example, that if any two data managers would process the same data according to the instructions, the resulting Archival Information Packages would be substantially similar.

On September 23rd, 2014 we were awarded the 2013 DSA Certificate[7], and are planning to renew it regularly since we view it as a tool that can be used in preparing for the changes and challenges we will inevitably face. FSD was the first Finnish organisation to acquire the DSA and we have received several inquiries about our certification. We have been very pleased that, for its own part, FSD's trust process has raised awareness about trust and long-term preservation of digital data in Finland.

Conclusion

Over the last few years, FSD has successfully used metrics to improve the quality of its processes and operations. As was to be expected, FSD conforms to the OAIS model. The TDR exercise confirmed that FSD is doing things in the right way, the CESSDA Trust Process allowed us to compare FSD with other archives, and the DSA certification increased trust with stakeholders.

The use of models and metrics to assess our procedures and policies have raised our awareness about the challenges of digital preservation, revealed existing and possible problems and weaknesses as well as strengths, steered and initiated minor and major changes in our operations, and resulted in improved documentation. As a consequence, many of our processes are now better and more efficient or, they will be better – some of the bigger changes will take time to implement. We are also able to better manage risks, provide more trustworthy services for the research community, and demonstrate FSD's trustworthiness to our stakeholders. In addition, we are in a good position to meet the CESSDA membership criteria.

Our journey into the world of standards and measurement has so far taken four years, and we see metrics such as the TDR and DSA as tools to be used continually for improving processes and services. The actual intensive work related to OAIS, TDR and DSA has required in total about six person months' worth of effort which we see as resources well spent. We are also considering taking the next step in the European Framework for Audit and Certification, which is a structured, externally reviewed and publicly available self-audit based on ISO 16363 or DIN 31644. The research landscape and thus the research data landscape is changing rapidly. As a digital repository we need to stay alert, be ready to review our policies and procedures methodically and critically, and build and share our competence continuously.

Funding Note

In order to be successful in any project, commitment from management as well as sufficient resources are needed. The work described in this paper has been part of the FSD Upgrade and VERIC projects[8], both funded by the Academy of Finland and both aimed at strengthening the Finnish national service provision as part of the CESSDA ERIC process.

References

Beedham H, Missen J, Palmer M and Ruusalepp R (2005). Assessment of UKDA and TNA compliance with OAIS and METS standards. Joint Information Systems Committee (JISC), United Kingdom. Available at: http://www.esds.ac.uk/news/publications/oaismets.pdf [Accessed 4.6.2015]

CCSDS (2011). Audit and Certification of Trustworthy Digital Repositories. Magenta Book. Issue 1. September 2011. Available at: CCSDS 652.0-M-1. http://public.ccsds.org/publications/archive/652x0m1.pdf [Accessed 10.4.2012]

CCSDS (2009). Reference Model for an Open Archival Information System (OAIS) Draft Recommended Standard. CCSDS 650.0-P-1.1 (Pink Book), Issue 1.1, August 2009. Available at: http://public.ccsds.org/sites/cwe/rids/Lists/CCSDS%206500P11/Attachments/650x0p11.pdf [Accessed 17.2.2010]

Center for Research Libraries (2012). TDR Checklist in Excel Format (Preliminary version). Available at: http://www.crl.edu/sites/default/files/attachments/pages/TDR_Checklist_Self_audit.xls [Accessed 10.4.2012]

Data Seal of Approval (2014). DSA Overview article. Available at: http://datasealofapproval.org/media/filer_public/2014/10/03/20141003_dsa_overview_defweb.pdf [Accessed 4.6.2015]

Dobratz, Susanne, Peter Rödig, Uwe M. Borghoff, Björn Rätzke and Astrid Schoger (2010). The Use of Quality Management Standards in Trustworthy Digital Archives. International Journal of Digital Curation 2010, Vol. 5, No. 1, pp. 46-63. DOI:10.2218/ijdc.v5i1.143

Doorn, Peter and Heiko Tjalsma (2007). Introduction: archiving research data. Archival Science 7(1), 1-20. DOI:10.1007/s10502-007-9054-6

Dusa A, Krejčí J, Štebe J, Fábián Z, Hegedus P, Hausstein B (2010). WP6 Final Report: Strenghtening the CESSDA RI (D6.1). CESSDA PPP. Available at: http://ppp.cessda.net/doc/WP6_Final_Report.pdf [Accessed 4.6.2015]

Giaretta, David (2011). Advanced Digital Preservation. Heidelberg: Springer-Verlag Berlin.

Hedstrom, Margaret (1998). Digital Preservation: A Time Bomb for Digital Libraries. Computers and the Humanities 31(3): 189-202. DOI 10.1023/A:1000676723815

Kvalheim V, Kiberg D, Kvamme T, Balster E, de Bruijne M, Wijnant A, Recker A, Lenkiewicz P, Widdop S, and Wloka B (2013). Roadmap for Preservation and Curation in the SSH. DASISH report D4.1. Available at: http://dasish.eu/publications/projectreports/D4.1_-_Roadmap_for_Preservation_and_Curation_in_the_SSH.pdf/ [Accessed 4.6.2015]

Schumann, Natascha and Recker, Astrid (2012). De-mystifying OAIS compliance: Benefits and challenges of mapping the OAIS reference model to the GESIS Data Archive. Iassist Quarterly Summer 2012, 6-11. Available at: http://www.iassistdata.org/downloads/iqvol36_2_recker_1.pdf [Accessed 4.6.2015]

Schumann, Natascha (2012). Tried and Trusted. Experiences with Certification Processes at the GESIS Data Archive. Iassist Quarterly Fall Winter 2012, 23-27. Available at: http://www.iassistdata.org/downloads/iqvol36_34_schumann_0.pdf [Accessed 4.6.2015]

Vardigan, Mary and Whiteman, Cole (2007). ICPSR meets OAIS: applying the OAIS reference model to the social science archive context. Archival Science 7:73-87. Available at: http://deepblue.lib.umich.edu/bitstream/2027.42/60440/1/Vardigan.Whiteman.Applying%20OAIS.pdf [Accessed 4.6.201]

Woollard, Matthew (2011). Standards-based approach to preservation planning. Presentation at the Aligning National Approaches to Digital Preservation Conference, Tallinn, Estonia, 23-25 May 2011. Available at: http://www.data-archive.ac.uk/media/276922/woollard_20110524.pdf [Accessed 4.6.2015]

Notes

1. Mari Kleemola is Information Services Manager at the Finnish Social Science Data Archive. She can be reached by email: mari.kleemola@uta.fi.
2. This paper is an updated version of a presentation given at IASSIST 2012: Kleemola, Mari (2012). Improving Operations Using Standards and Metrics: Self-Assessment of Long-Term Preservation Practices at FSD. A presentation at the 38th Annual IASSIST Conference, Washington DC, June 06, 2012. Available at: http://www.iassistdata.org/conferences/2012/presentation/3328
3. Consortium of European Social Science Data Archives, http://www.cessda.net
4. The Finnish version is called SFS 5972 Viitemalli pitkäaikaissäilytysarkistolle.
5. CCSDS 652.0-M-1 -- Audit and Certification of Trustworthy Digital Repositories (September 2011) contains the final draft standard submitted to ISO for review and approval and is freely available from the CCSDS website as a Recommend Practice document: http://public.ccsds.org/publications/archive/652x0m1.pdf [3.6.2015]
6. http://www.trusteddigitalrepository.eu/
7. https://assessment.datasealofapproval.org/assessment_109/seal/pdf/ [4.6.2015]
8. More information about FSD's projects: http://www.fsd.uta.fi/en/news/projects.html