Data Seal of Approval

Certification for sustainable and trusted data repositories

Data Seal of Approval: Certification for sustainable and trusted data repositories

Ingrid Dillo and Lisa de Leeuw

Data Archiving and Networked Services (DANS)

Postbox 93067
2509 AB The Hague
The Netherlands
+31 70 349 44 50
info@dans.knaw.nl
www.dans.knaw.nl

This article was previously published in the Handbook Information Science and in the IWA-base, both B + B Vakmedianet.nl.

Data Seal of Approval

- Website: http://www.datasealofapproval.org
- Contact: info@datasealofapproval.org
- Current Seal:

Introduction

If we want to share data, the long-term storage of those data in a trusted digital archive is a *sine qua non*. Data created and used by scientists should be managed, curated and archived in order to preserve the initial investment in collecting them. Researchers must be certain that the data provided by the archives remain useful and meaningful, even in the long term. In addition, the archives should have sustainable business models themselves. The concept of sustainability involves many challenging aspects in many areas: organizational, technical, financial, legal, etc. Certification can be an important contribution to ensuring the reliability and durability of digital archives and hence the possibilities for sharing data over a long term.

Definition of research data
What exactly do we mean by the term 'research data'? It certainly does not refer to the scientific data generated by the major facilities in the beta sciences only. The concept of research data is to be interpreted broadly.

Many definitions can be found online (http://ands.org.au/guides/what-is-research-data.html). Griffith University (Australia) uses the following definition: *'Research data are factual records, which may take the form of numbers, symbols, text, images or sounds, which used as primary sources for research, which are commonly accepted in the research community as necessary to validate research findings.'* Research data are data which are produced, collected and/or used by researchers. Another definition comes from the University of Minnesota. It accommodates the different processing levels of data: *'Research data are data in any format or medium that relate to or support research, scholarship, or artistic activity. They can be classified as:*
• Raw or primary data: information recorded as notes, images, video footage, paper surveys, computer files, etc.
• Processed data: analyses, descriptions, and conclusions prepared as reports or papers.
• Published data: information distributed to people beyond those involved in data acquisition and administration.'

From these definitions it becomes clear that the certification of digital archives is not only important for scientific archives of primary research data, but also for cultural heritage institutions such as public libraries, museums and archives.

Sharing data and the concept of trust
In recent years, sharing data has become an important issue in Europe. The famous European Commission (EC) report *Riding the Wave. How Europe can gain from the rising tide of scientific data, published in 2011, already stressed 'the critical importance of sharing and preserving reliable data produced during the scientific process'.*

The following year, EC Vice-President Neelie Kroes (see box p. 4) encouraged researchers to grant open access to their data as

much as possible. Kroes was convinced that *'sharing data, and having the forum to openly use and build on what is shared, are essential to science. They fuel the progress and practice of scientific discovery.*' For Europe, data are the new gold.

Neelie Kroes

• Neelie Kroes, Vice-President of the European Commission responsible for the Digital Agenda. Opening Science Through e-Infrastructures: European Federation of Academies of Sciences and Humanities Annual Meeting – 'Open infrastructures for Open Science' Rome, Italy, 11 April 2012. European Commission – SPEECH/12/258 11/04/2012 http://europa.eu/rapid/press-release_SPEECH-12-258_en.htm?locale=en

Closed data

• Closed data culture is breeding ground for scientific fraud. Peter Doorn and Ingrid Dillo conclude that the prevalent culture in academic psychology is not to make research public. In order to reduce the susceptibility to fraud that culture should be abolished. Data must always be shared. Their full article can be found at: http://www.refdag.nl/opinie/gesloten_datacultuur_voedingsbodem_voor_wetenschappelijke_fraude_1_589330

This paved the way for the *Recommendation on Access to and Preservation of Scientific Information*, which the European Commission published in 2012. In this recommendation, the Commission encouraged a European open access policy.

Meanwhile, the new European research framework program, Horizon 2020, includes an open data pilot. In the United States, too, the government promotes an open data policy. President Obama has published the Executive Order '*Making Open and Machine Readable the New Default for Government Information*', including research data.

Why is increasing importance being attached to research data? First, sharing data makes science more transparent. It facilitates replication and validation of research. This will enhance research quality. In view of the many cases of fraud in Dutch research in recent years, this is certainly an important argument in our national context (see box).

Another benefit of sharing data is that data can be reused by researchers who did not generate those data themselves. This reuse will lead to greater efficiency in research. It offers researchers the ability to combine datasets and use them across disciplines. Furthermore, open data can also be used for economic and social interests beyond science. Eventually, data sharing will lead to a higher return on the initial investment.

Although data sharing has clear advantages for science and society as a whole, it is certainly not common practice. In 2011, DANS (Data Archiving and Networked Services) conducted a national survey: 'The Dutch data landscape in 32 interviews and a survey' (see box). In this survey 400 researchers from all disciplines were asked where they store their data. More than 70 percent responded that they kept their data on their own or their institute's computer.

A recent study by the Norwegian Research Council also shows that 80 percent of the researchers surveyed stress the importance of sharing data, but at the same time 85 percent report that they still keep their data on a private or institutional computer.

Why is sharing data still threatening to so many researchers? One of the arguments mentioned is that data generated elsewhere are not reliable. This has everything to do with trust. To refute this

argument, we must ensure that we build the element of trust into the digital archiving services that provide researchers with sustainable access to third parties' data.

Trust is the basis of storing and sharing data. That trust must be present in various stakeholders. The data depositors want the assurance that their data in the digital archive are safe and will remain accessible, usable and meaningful. Data users have questions like: have the data been well kept, have they retained their authenticity and integrity, are the data of good quality, do the identifiers refer to the appropriate objects? The funders have other concerns. They want to be certain that their investment in data production yields optimum returns, i.e. that the data will be available for long-term reuse.

What characteristics make digital archives reliable? First, a digital archive's mission should be to give reliable long-term access to the digital data under their care, now and in the future. Second, there should be permanent monitoring, planning and maintenance. The threats and risks within their systems must be understood. Finally, there should be a regular audit and certification cycle in place. Reliability is not something you achieve once and can then take for granted.

Certification can make an important contribution to the confidence of various stakeholders. The Data Seal of Approval (DSA) has sixteen guidelines for data repositories and enables basic certification. In this article we will discuss the DSA standard extensively.

The Dutch data landscape

• The Dutch data landscape in 32 interviews and a survey. This publication shows how Dutch researchers think about data sharing in their field. In addition to the survey DANS interviewed a large number of top researchers, creating a picture of the full breadth of Dutch science: How are date being handled, what could be improved, and how should this be done? For more info: http://www.dans.knaw.nl/content/categorieen/ publicaties/dutch-data-landscape-32-interviews-and-survey

DANS

• Data Archiving and Networked Services (DANS).
DANS promotes sustained access to digital research data. For this purpose, DANS encourages researchers to archive and reuse data in a sustained manner, e.g. through the online archiving system EASY. DANS also provides access, via NARCIS.nl, to thousands of scientific datasets, e-publications and other research information in the Netherlands. In addition, the institute provides training and advice, and performs research into sustained access to digital information. Driven by data, DANS ensures that access to digital research data keeps improving, through its services and by taking part in national and international projects and networks. DANS is an institute of the Royal Netherlands Academy of Arts and Sciences(KNAW) and the Netherlands Organisation for Scientific Research (NWO). See also: http:// www.dans.knaw.nl/

Data Archiving and Networked Services
DANS

Data Seal of Approval

When DANS (see box) was established by the two main Dutch science organizations, KNAW and NWO, they assigned it the task of developing a Seal of Approval for digital data to ensure that archived data can still be found, understood and used in the future. In 2008 the first edition of Data Seal of Approval: Quality guidelines for digital research data was presented at an international conference. The seal was initially developed for use in the Netherlands, but it was soon found to be very useful in an international context too. In 2009 the Data Seal of Approval was therefore transferred to an international body, the DSA Board, which has managed and further developed the guidelines and the peer review process ever since.

The objectives of the Data Seal of Approval are to safeguard data, to ensure high quality and to guide reliable management of data for the future without requiring the implementation of new standards, regulations or heavy investments.

The Data Seal of Approval:
• Gives researchers the assurance that their data will be stored in a reliable manner and can be reused;
• Provides funding bodies with the confidence that research data will remain available for reuse;
• Enables researchers to assess in a reliable manner the repositories that hold the data which they want to reuse;
• Supports data repositories in the efficient archiving and distribution of data.

The 16 guidelines
The Data Seal of Approval involves 16 guidelines for applying and verifying quality aspects concerning the creation, storage, use and reuse of digital data. The guidelines have been designed with a focus on scientific materials, but they can be applied to all types of digital information. The guidelines serve as the basis for awarding the Data Seal of Approval by the DSA Board (see box).

The criteria for awarding the Data Seal of Approval to data repositories are in accordance with national and international guidelines for digital data archiving such as the *Kriterienkatalog vertrauenswürdige digitale Langzeitarchive* developed by NESTOR, the Digital Repository Audit Method Based on Risk Assessment (DRAMBORA) published by the Digital Curation Centre (DCC) and DigitalPreservationEurope (DPE), and *Trustworthy Repositories Audit & Certification (TRAC): Criteria and Checklist* of the Research Library Group (RLG). The following publications have also been taken into account: *Foundations of Modern Language Resource Archives* by the Max Planck Institute, and *Stewardship of Digital*

DSA board

- The following organizations are members of the DSA Board:

Research Data: A Framework of Principles and Guidelines by the Research Information Network.

The DSA guidelines can be seen as a minimum set distilled from the above proposals.

Fundamental to the guidelines are five principles that together determine whether or not the digital data may be considered as sustainably archived:
- The data can be found on the Internet.
- The data are accessible, while taking into account relevant legislation with regard to personal information and intellectual property.
- The data are available in a usable format.
- The data are reliable.
- The data can be referred to (persistent identifiers).

These principles are integral to the guidelines, which focus on three stakeholders:
- The *data producer*, who is responsible for the quality of the digital data;
- The *data repository*, who is responsible for the quality of storage and availability of the data (data management);
- The *data consumer*, who is responsible for the quality of use of the data.

The basic assumption is that the *data repository* is responsible for enabling and supporting data producers' and data consumers' compliance with the guidelines.

A data repository is designated a *Trusted Digital Repository* (TDR) if it complies with Guidelines 4 to 13 and if it enables data producers and data consumers to comply with Guidelines 1 to 3 and 14 to 16.

Other guidelines

- Drambora: http://www.dcc.ac.uk/resources/ repository-auditand-assessment/ drambora
- Trac: http://www.crl.edu/archiving-preservation/digital-archives/metrics-assessing-and-certifying
- Foundations of Modern Language Resource Archives – an article by Peter Wittenburg, Daan Broeder, Wolfgang Klein, Stephen Levinson and Laurent Romary: http://pubman.mpdl.mpg.de/pubman/item/ escidoc:58934:4/component/escidoc:58935/Wittenburg_2006_foundations.pdf
- Stewardship of Digital Research Data: A Framework of Principles and Guidelines: http://www.rin.ac.uk/our-work/data-management-andcuration/stewardship-digital-research-data-principles-and-guidelines

Guidelines for data producers

The quality of the digital research data is determined by:
• Their intrinsic value to their sector (designated community): scientific, scholarly, business, etc.;
• The format in which the data and supporting information are stored;
• The documentation (metadata or contextual information) supporting the data.

Guideline 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.
Sector-specific quality criteria indicate to what degree the data are of interest to consumers. The assessment by experts and colleagues in a field is the main deciding factor for the quality of data. Transparency in terms of adherence to ethical norms in relevant disciplines facilitates the assessment of data content. Therefore, it is the responsibility of the data producer to provide sufficient information to enable data consumers to assess the data.

Guideline 2:
The data producer provides the data in formats recommended by the data repository.
The bits which make up a digital object are ordered according to the rules of a specific data format. Various data formats exist for digital objects, and all formats run the of becoming obsolete and rendering data objects unusable. For storage of data objects preferred formats are used. Preferred formats are those that a data repository can reasonably assure will remain readable and usable. Typically, these are the de facto standards employed by a particular discipline.

Dublin Core

• The Dublin Core Metadata Element Set is a vocabulary of fifteen properties for use in resource description. The name 'Dublin' is due to its origin at a 1995 invitational workshop in Dublin, Ohio; 'core' because its elements are broad and generic, usable for describing a wide range of resources. http://dublincore.org/documents/dces/

Guideline 3:
The data producer provides the data together with the metadata requested by the data repository.
It is the responsibility of the data producer to provide the data with information about the context of the data (metadata). There is a distinction between descriptive, structural and administrative metadata. These must be provided in accordance with the guidelines of the data repository.
• *Descriptive metadata* consist of information required to find data and add transparency to their meaning (definition and value) and importance. Examples of descriptive metadata are the data elements of the Dublin Core Element Set (see box), with fields such as creator, type and date.
• *Structural metadata* indicate how different components of a set of associated data relate to one another. These metadata are needed to be able to process the data. When data are encoded, the codebook will be a component of the structural metadata.

• *Administrative metadata* are required to enable permanent access to the data. This concerns the description of intellectual property, conditions for use and access and the preservation metadata needed for durable archiving of the data.

The data repository specifies the level of producer-created metadata required and provides the tools for their effective capture.

Guidelines for data repositories

The *data repository* is responsible for access and preservation of digital data in the long term. Two factors, in particular, determine the quality of the data repository:
• The quality of the organizational framework in which the data repository is incorporated (*organization and processes*);
• The quality of the *technical infrastructure* of the data repository.
Organizations that play a role in digital archiving and are establishing a Trusted Digital Repository shall possess a sound financial, organizational and legal basis in the long term.

Guideline 4:
The data repository has an explicit mission in the area of digital archiving and promulgates it.

Guideline 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.

Guideline 6:
The data repository applies documented processes and procedures for managing data storage.

Guideline 7:
The data repository has a plan for long-term preservation of its digital assets.

Guideline 8:
Archiving takes place according to explicit work flows across the data life cycle.

Guideline 9:
The data repository assumes responsibility from the data producers for access and availability of the digital objects.

Guideline 10:
The data repository enables the users to discover and use the data and refer to them in a persistent way.

Guideline 11:
The data repository ensures the integrity of the digital objects and the metadata.

This Guideline relates to the information contained in the digital objects and metadata and whether it is complete, whether all changes are logged and whether intermediate versions are present.

Guideline 12:
The data repository ensures the authenticity of the digital objects and the metadata.
This Guideline refers to the degree of reliability of the original and to the provenance of the data, including relationships between the original data and those disseminated, and whether or not existing relationships between datasets and metadata are maintained.

Guideline 13:
The technical infrastructure explicitly supports the tasks and functions described in internationally accepted archival standards like OAIS.
The technical infrastructure constitutes the foundation of a Trusted Digital Repository. The OAIS reference model, an ISO standard, is the de facto standard for using digital archiving terminology and defining the functions that a data repository fulfils.

Guidelines for data consumers

The *data consumer* uses the digital data in compliance with the guidelines below.

Guideline 14:
The data consumer complies with access regulations set by the data repository.

Guideline 15:
The data consumer conforms to and agrees with any codes of conduct that are generally accepted in the relevant sector for the exchange and proper use of knowledge and information.

Guideline 16:
The data consumer respects the applicable licences of the data repository regarding the use of the data.
The quality of the use of data is determined by the degree to which the data can be used without limitation by the various target groups, while complying with applicable codes of conduct. The open and free use of data takes place within the relevant legal frameworks and the policy guidelines as determined by relevant national authorities.
With regard to accessing information, the data consumer is bound by relevant national legislation. The data repository may have separate access regulations, which include restrictions imposed by the laws of the country in which the data repository is located. Access regulations should be based on relevant international access standards (e.g. Creative Commons) as much as possible.
Most nations have legal frameworks relating to the ethical use and reuse of data. These frameworks range from statutory codes – which protect the privacy of individuals – to formal codes of

conduct which inform ethical issues. Repositories must be aware of these local legal frameworks and ensure that they are taken into account when providing data for reuse.

Procedures

Self-assessment and peer-review process
The starting point for obtaining the Data Seal of Approval is the website www.datasealofapproval.org, where an application form can be submitted. Once the form is received by the DSA Board, a self-assessment is made available in the DSA online tool. The self-assessment is meant to supply evidence that the applicant data repository meets the 16 DSA guidelines and the relevant level of compliance. A description of the context of the data repository is also required.
Since the DSA is used in an international environment, the language of communication is English in order to increase transparency.

After the submission of the self-assessment by the data repository, the DSA Board appoints a peer reviewer who is given two months' time in which to evaluate the self-assessment. The peer reviewer will either confirm the evidence or require additional information depending on the adherence to the guidelines and the level of compliance. Resubmission of the modified application and requesting for additional information by the peer reviewer will continue until the reviewer has seen enough evidence to award the DSA. In the event of a dispute, the applicant data repository can contact the DSA Board.

As long as a self-assessment is in the application process, it will not be made public. The self-assessment, including all evidence, will only be published on the websites of the DSA and the applicant data repository after the DSA has been awarded.

Since approved applications, including any evidence and peer review comments, are publicly available on the DSA website, they can be used as references or samples.

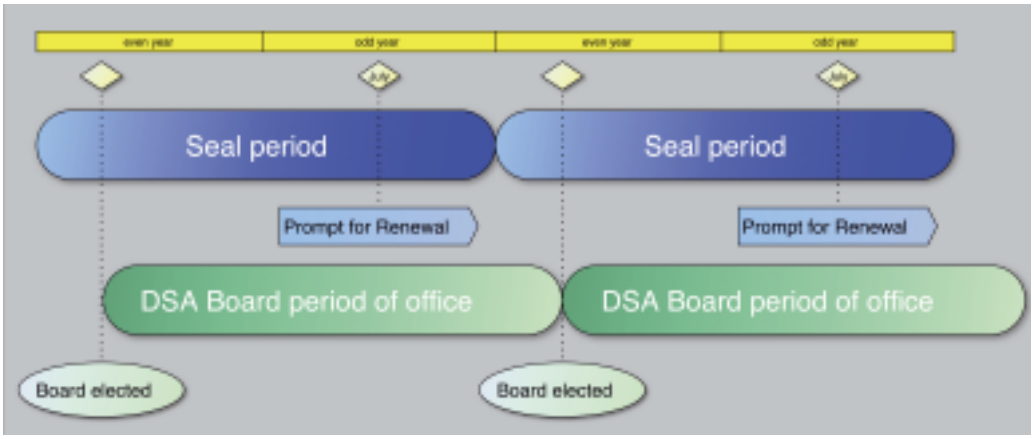
Displaying the Data Seal of Approval
After the Data Seal of Approval is awarded by the DSA Board, the DSA logo may be displayed on the repository's website. The Board will provide appropriate HTML code, which includes the DSA logo and a link to the organization's assessment.

At the same time, the DSA Board will post the approved assessment of the new DSA repository on the DSA website, using the name of the specific repository and a logo if provided.

Renewing the Data Seal of Approval
A Data Seal of Approval for a given period can be displayed indefinitely but will need to be updated periodically if the repository wants to stay compliant with newly released standards and receive the latest DSA logo. DSA-certified repositories will be

Data Seal of Approval life cycle

- The DSA Board meets formally once a year. Its members are elected in an even year for a two-year term.
- The new DSA Board is appointed immediately after the General Assembly meeting at which the election results are announced. Interim Board members, elected by an extraordinary General Assembly meeting, are appointed immediately after their election, their term ending with the full Board's term.
- The Board considers amendments to the DSA guidelines and Regulations during the term for which it is elected. The changes will be introduced at the end of the current Seal period.
- The DSA guidelines remain valid for a period of two calendar years: the Seal period.
- The Data Seal of Approval is awarded for a Seal period.
- The DSA logo shows the two years of the Seal period.
- Each data repository displaying the DSA will be informed well in advance of the expiry of the Seal period and asked to renew their application.
- The data repository then has the choice of either (a) updating its self-assessment for the new period and submitting an application for the most recent Seal, including the latest version of the guidelines, or (b) continuing to display the outdated logo on their website.



contacted automatically when an update is available. The current Seal is the one issued according to version 2 of the guidelines and displaying the years 2014-2015 (see box).

Data Seal of Approval Community and Regulations
The Data Seal of Approval is driven by the voluntary involvement of all stakeholders. The organization of the DSA is established by Regulations, which are available on the DSA website. The Regulations define the various rights and duties of the DSA Community.

The world of DSA is made up of a number of components:
- The DSA Community comprises all of the organizations with one or more DSA-certified repositories.
- The DSA General Assembly is the governing body of the DSA Community. The General Assembly elects the DSA Board and

provides the Board with advice when needed. General Assembly members commit to conducting a maximum of three peer reviews a year to ensure that the DSA remains community-driven and sustainable.
• The DSA Board is drawn from and elected by the General Assembly representatives. The Board conducts the daily business of the DSA Community, manages and monitors the DSA assessment procedure, convenes meetings of the General Assembly and informs the DSA Community about all DSA activities.
• Peer reviewers belong to one of the organizations in the General Assembly and have completed at least one self-assessment which resulted in the award of the latest DSA. They review and assess evidence in a timely, complete and impartial manner, ensuring that DSA applications remain confidential until the DSA is awarded.

DSA online tool
An online tool has been developed to make the DSA application process easier and more transparent. It is an online system that guides the applicant and the peer reviewer from the application to the award of the DSA. When a DSA is applied for on the DSA website, the DSA Board forwards the information to the administrator of the online tool. The administrator enters all the details of the organization and the associated data repository into the system and then adds the peer reviewer appointed by the Board. The applicant will receive a user name, password and login link to gain access to the online tool. The tool is located in a secure environment that can be recognized by the 'https' preceding the URL.

Once the applicant logs into the online tool, he is shown the assessment guide. It gives a brief overview of the purpose of a DSA, some completion instructions and an explanation of the terms used in the online tool. In addition it also shows a list of 16 guidelines and the minimum level of compliance required. When the user indicates that he has read the assessment guide, the applicant is guided through the self-assessment in which the 16 guidelines are discussed one by one. In order to further facilitate the review of the data repository the user is also asked to describe the context in which the data repository operates. During the completion of the evidence and the compliance level help texts are available for structuring the evidence.

Once the applicant has completed the evidence and the level of compliance for the data repository, the self-assessment can be submitted through the system. The applicant will receive an automatically generated confirmation email.

The reviewer assigned to the application will receive an automatic email from the system, notifying that a new application has been submitted. It also states the deadline for completing the assessment.

The reviewer of a submitted self-assessment has two months to

evaluate it. As long as a reviewer has not completed the evaluation he will receive automatic reminders at specified intervals prior to the two-month deadline. The online tool administrator also receives these reminders so that he can notify the Board, if necessary, and they can take action to keep the process going. The system guides the reviewer through the evidence and the chosen level of compliance. Help texts are available with each Guideline to assist the reviewer of the self-assessment.

Once the reviewer has completed the review in the system, the applicant is automatically informed of the outcome. This will be either a request for more information to enable the reviewer to better assess the application, or confirmation that the DSA is awarded.

If the former is the case, the applicant is notified that the self-assessment has been reviewed but that additions are needed before the DSA can be awarded. The applicant is invited to log into the online tool again to complete the self-assessment. When the applicant logs into the system the numbers of the guidelines that need supplementing are highlighted in red to make it obvious where changes are needed. Also, any comments added by the reviewer will be visible. If the applicant is satisfied with his additions he will resubmit the self-assessment and the system will automatically inform the reviewer, who can continue the review. This process can be repeated several times before the DSA is awarded.

In the latter case, the applicant receives a notification from the system that the DSA has been awarded. It also contains the HTML code with the DSA logo and a link to the self-assessment to be included on the applicant's website. The system also ensures that the self-assessment and the review will be published on the DSA website.

Experiences of applicant organizations
During the annual Data Seal of Approval (DSA) conferences case studies are presented of data repositories that applied for a DSA. The following conclusions can be drawn from their experiences.

General:
• Performing DSA self-assessment does not take much time; on average, two to four days. This mainly depends on the level of documentation and its disclosure.
• Although most documentation is intended to be publicly accessible, an exception can be made for documentation containing privacy-sensitive and confidential information, such as a long-term vision.
• The DSA process is very useful as an evaluation of internal procedures, which can be reviewed and updated where necessary. The current state of affairs, which can also serve for future accreditation, is made visible. Additionally, the procedures and documentation are evaluated, tested and approved by an external professional and the DSA is very helpful in determining strengths and weaknesses.

• The DSA reaffirms the necessity and usefulness of succession/long-term planning and helps to get these issues higher on the agenda of management.
• The DSA contributes to a reliable image. It can be used to improve reputation, but also as a benchmark for comparison. It clarifies what constitutes a digital archive and its business, and it creates transparency for the community in the area of sustainability.
• The DSA increases the confidence of users: it shows that standards are being used, just like the ones being used by traditional museums or archives.
• The DSA helps to build a community: 'we' all work according to the same standards.
• The DSA emphasizes the need to conform to the OAIS standards.
• Interaction with the peer reviewer is perceived as significant.
• The DSA logo makes it easy to recognize what the seal stands for.
• The DSA guidelines are sufficiently generic to be applied to publications as well as scientific data.
• Because of its general approach the DSA is perceived as a less 'threatening', detailed and time-consuming procedure than more comprehensive standards, such as ISO or TRAC. The focus is on increasing awareness and transparency; DSA takes a community's and peer reviewer's point of view rather than a top-down approach.
• The DSA is a solid foundation for applying for DIN 31644 certification.
• By renewing the DSA the data repository will show its progress.

Practical:
• The DSA self-assessment is a static document; keep in mind that links can change or may no longer be accessible.
• It is better not to include technological details that are subject to change.
• It may be useful to draw up a schedule for reviewing the DSA self-assessment, regardless of reminders of the need to renew the DSA.

ADS

• ADS have made an analysis of their experiences when applying for a DSA. This case study can be a useful source of information. http://www.dcc.ac.uk/resources/case-studies/ads-dsa

ads ARCHAEOLOGY DATA SERVICE

DIN, NESTOR and ISO

- DIN 31644 standard: Information und Dokumentation – Kriterien für vertrauenswürdige digitale Langzeitarchive. http://www.nabd.din.de/cmd?level=tpl-art-detailansicht&committeeid=54738855 &artid=147058907&languageid=de
- NESTOR is a competence network for long-term preservation of digital data in Germany. It is composed of twelve libraries, archives and museums who bundle together standardisation activities and provide standards for user communities. http://www.langzeitarchivierung.de/Subsites/nestor/EN/Home/ home_node.html
- DIN 31644 certification can be applied for through NESTOR: http://www.langzeitarchivierung.de/Subsites/nestor/EN/nestorSiegel/siegel_node.html
- ISO standard 16363: Audit and Certification of trustworthy digital repositories http://www.iso.org/iso/iso_catalogue/catalogue_tc/ catalogue_detail.htm?csnumber=56510
- ISO standard 16919: Requirements for bodies providing audit and certification of candidate trustworthy digital repositories http://www.iso.org/iso/catalogue_detail.htm?csnumber=57950
- ISO 16363 certification can be applied for through: http://www.iso16363.org/

Related initiatives and the future

Three ways of evaluating a digital archive
For some time there has been demand for a way to evaluate the reliability of digital archives. The last few years a number of evaluation guidelines have become available.

The Data Seal of Approval (DSA) offers the possibility of basic certification.

The DIN standard provides a second set of guidelines. The 34 criteria were developed by the German organization NESTOR (a consortium of museums, archives and libraries) and formalized as the DIN 31644 standard.

One of the NESTOR working groups has developed a certification process based on DIN 31644: Kriterien für vertrauenswürdige digitale Langzeitarchive. This DIN standard is essentially a catalogue of criteria which digital archives should satisfy. In 2014 the first DIN-based audits are being conducted.

The third way to evaluate a digital archive is provided by ISO standard 16363. This standard is based on the OAIS model that provides a framework for understanding archival concepts needed for the preservation of and long-term access to digital information. Based on this model, the American organizations RLG, OCLC and NARA developed TRAC: the Trustworthy Repository Audit and Certification Criteria and Checklist, the forerunner of the current ISO standard.

The standard is very detailed and contains more than one hundred criteria for different aspects of a digital archive. They focus on organizational infrastructure, digital object management, and infrastructure and risk management. In 2011, six test audits were performed: three in Europe and three in the US. The ISO standard is based on a formal external audit of the archive, formalized in ISO 16919: *Requirements for bodies providing audit and certification of candidate trustworthy digital repositories*.

The European Framework for Audit and Certification of digital Repositories
How do these standards fit together? In 2010, a Memorandum of Understanding (MoU) was signed by the parties involved. The purpose of the MoU was to set up a comprehensive multi-level framework for the certification of digital archives. This European Framework for Audit and Certification of Digital Repositories offers three evaluation levels of increasing reliability.

Basic Certification is granted to repositories qualifying for the DSA.
Extended Certification is granted to repositories with

Basic Certification that perform an additional 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, pass a full external audit and certification based on ISO 16363 or DIN 31644.

The DSA focuses on smaller organizations within the scientific data domain. The two more formal standards, DIN and ISO, are more demanding but also give more guarantees regarding reliability.

Collaboration with the World Data System (WDS)
In 2013, a number of funding organizations (the European Commission, the National Science Foundation and the Australian government) promoted the establishment of the International Research Data Alliance (RDA, see box).

Work within the RDA is carried out by working groups and so-called interest groups. Recently the working group Repository Audit and Certification: DSA WDS Partnership was launched (see box). In this group, the DSA Board collaborates with the scientific committee of ICSU/WDS. The World Data System (WDS) is a body of the International Council for Science (ICSU), whose data archives can be members. The WDS required some categories of members to go through an accreditation process. This accreditation is very similar to DSA. It was therefore decided to explore the possibilities of collaboration in this working group. Collaboration could lead to more efficiency and more certifications in the future. The experience of DSA lies mainly within the social sciences and humanities, while the WDS has focused in the earth and space sciences.

Memorandum of Understanding

- Memorandum of Understanding to create a European Framework for Audit and Certification of Digital Repositories
http://www.trusteddigitalrepository.eu/Site/Trusted%20Digital%20Repository.html

RDA

- The RDA 'builds the social and technical bridges that enable open sharing of data. The RDA vision is researchers and innovators openly sharing data across technologies, disciplines, and countries to address the grand challenges of society.'
https://rd-alliance.org/about.html

WDS

- The World Data System (WDS) is a body of the International Council for Science (ICSU): https://www.icsu-wds.org/organization/ intro-to-wds
RDA working group aimed at Repository Audit and Certification:
https://rd-alliance.org/group/ repository-audit-and-certification-ig-dsa%E2%80%93wds-partnership-wg.html

Infrastructures

- Consortium of European Social Science Data Archives (CESSDA): http://www.cessda.net/

- Common Language Resources and Technology Infrastructure(CLARIN): http://clarin.eu/

- Digital Research Infrastructure for the Arts and Humanities (DARIAH): http://www.dariah.eu/

- European Data Infrastructure (EUDAT): http://eudat.eu/

DSA business model

DSA is doing well. The DSA Community is growing and thriving. Today (2014), more than 31 Seals have been awarded and nearly 30 digital archives are working on their DSA self-assessment.

The added value of the DSA process is not only recognized by individual repositories. Within the European research infrastructures, building confidence in the services offered is considered increasingly important. In this context infrastructures such as CESSDA, CLARIN and DARIAH are looking at the DSA guidelines. CLARIN has already made DSA certification mandatory for all its centres.

CESSDA is working to integrate the DSA guidelines with their own infrastructure and DARIAH is considering the adoption of the guidelines. In the proposal for the continuation of the European EUDAT project, DSA will also play a significant role.

At the same time, the DSA's success provides the challenge to further professionalize the DSA organization in the coming years in order to enable its community to continue to grow.

Ingrid Dillo

• Ingrid Dillo studied history and wrote a PhD thesis at Leiden University on the Dutch East India Company. Ingrid is a generalist who over the last twenty-five years has mainly been active in the field of policy development. After a period as policy researcher at Research voor Beleid in Leiden she worked for many years as a senior policy advisor at the Dutch ministry of Education, Culture and Science and at the Koninklijke Bibliotheek, the National Library of the Netherlands. Ingrid is now director policy at DANS (Data Archiving and Network Services). Among her areas of interest are research data management and the certification of digital repositories. Internationally Ingrid is active in the Research Data Alliance, the ICSU Word Data System and the Knowledge Exchange.

Lisa de Leeuw

• Lisa de Leeuw has worked for ABN AMRO Bank N.V. from 1995 to 2008 of which the last six years she was a management assistant to several Executive Vice Presidents of international departments. Per August 2008 she started at DANS (Data Archiving and Networked Services) as management assistant to the deputy director and project assistant for the preparation of DARIAH (Digital Research infrastructure for the Arts and Humanities). In 2009 she took on the secretariat of the Data Seal of Approval. Currently she is coordinator of the DANS projectburo, part of the DARIAH CIO team and management assistant.

Guidelines

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.
2 The data producer provides the data in formats recommended by the data repository.
3 The data producer provides the 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 work flows 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 discover and use the data and refer to them in a persistent way.
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 the relevant sector for the exchange and proper use of knowledge and information.
16 The data consumer respects the applicable licences of the data repository regarding the use of the data.