



Repository Audit and Certification Interest Group DSA–WDS Partnership Working Group Case Statement

Working Group Charter

Data Sharing

Data sharing is a hot topic. Sharing enables the reuse of data by researchers who did not generate these data themselves; leading to greater efficiencies and more research. It offers researchers the possibility to combine datasets and to use datasets in other disciplines. Data sharing also stimulates the usage of data beyond research in academia. Ultimately, data sharing leads to a higher return on investment.

Data sharing furthermore makes science more transparent and facilitates replication of research by others. Validation and replication are important elements in guaranteeing the integrity of research as part of the scholarly record.

Audit & Certification

To ensure the quality and usability of shared data, the long-term preservation of these data in sustainable digital repositories is a *sine qua non*. Data that are created and used by science and scholarship need to be managed, curated, and archived so that the substantial investments in preparing and presenting the content and tools will not be lost. Certification is fundamental in guaranteeing the trustworthiness of digital repositories, and thus in sustaining the opportunities for long-term data sharing and corresponding services.

In recent years, a number of certification standards and accreditation procedures have been developed worldwide: Data Seal of Approval (DSA)¹, Network of Expertise in long-term Storage and Accessibility of Digital Resources in Germany (NESTOR) seal / German Institute for Standardization (DIN) standard 31644², Trustworthy Repositories Audit and Certification (TRAC) criteria / International Organization for Standardization (ISO) standard 16363³, and the International Council for Science⁴ World Data System (ICSU-WDS) certification of WDS Members⁵.

The DSA and WDS certifications both offer a basic certification standard for trusted digital repositories (see Page 4 for the different certification levels). Their catalogues of requirements and

¹ <http://datasealofapproval.org>

² NESTOR: www.langzeitarchivierung.de, DIN: www.nabd.din.de, www.trusteddigitalrepository.eu

³ ISO standard 16363: www.iso16363.org

⁴ International Council for Science : www.icsu.org

⁵ WDS: <http://icsu-wds.org/>

their review procedures are based on the same principles of openness and transparency, and of striking the right balance between simplicity and robustness of the work and effort involved.

Up to this point, the two standards have evolved and operated independently. The primary focus of DSA has been on digital repositories in the Humanities and Social Sciences. For historical reasons, the focus of ICSU-WDS has been on Earth and Space Sciences. Both initiatives, however, have fully multidisciplinary missions.

There are differences between the two organizations that will need to be addressed. For example, ICSU-WDS has a membership focus, and includes not only data centres (mainly repositories) but also data services. Many of these data services consist of a number of components (data centres, analysis centres, product centres, etc.) and have their own organisational structure (e.g., central bureau or governing board). Moreover, the data services often have close linkages with the ICSU Scientific Unions and their Associations. The partnership between DSA and ICSU-WDS will need to handle this complexity.

Under the umbrella of the RDA/WDS Interest Group (IG) on Certification of Digital Repositories, this Working Group (WG)—consisting of representatives from the DSA and WDS communities and beyond—aims to explore and develop a DSA–WDS partnership with the objectives of realizing efficiencies, simplifying assessment options, stimulating more certifications, and increasing impact on the community. The output from this WG is envisioned as a possible first step towards developing a common framework for certification and a service of trusted data repositories.

Value Proposition

Who will benefit:

Repository certification is important because it promotes trust and confidence in the usability and persistence of shared data resources. It also helps repositories improve their practices and procedures. However, the value of certification is not apparent to all communities; more work needs to be done to clarify the problem that certification solves and to demonstrate that certification is worth the associated effort.

By reconciling two basic and lightweight certification mechanisms, we seek to simplify the array of certification options and to show the value to be gained from a certification procedure requiring relatively low investment of time and effort. Four community stakeholder groups will directly benefit from this combined certification standard for trusted digital data repositories and services:

Researchers, who want to be confident that the integrity and authenticity of data in digital repositories are protected; that those data remain accessible, usable, and meaningful over time; and that the data services provided can be safely used.

Funders, who want reassurance that their investment in the production of valuable research data is not wasted, and will continue to stimulate research into the future. Certification will make it easier for funders to assess repositories and to make informed investment decisions. Indeed, some funders in the Netherlands, for example, require the deposit of data they fund in trusted repositories.

Data repositories and services, for which certification provides a quality indicator to show to funders and users. In addition, the certification process will provide them with advice on where improvements are needed and useful.

Science publishers, who want to redirect article-related data and other supplementary materials to trustworthy data repositories.

More generally, a combined standard will benefit the larger community of scientific data users because more repositories will be certified; leading to greater trust in these institutions and more data sharing.

Impact:

By combining forces, DSA and ICSU-WDS bring together much expertise and experience in the area of certification over a broad range of disciplines and with a global reach. Collaboration will create an opportunity to profit from the best elements of the two organizations. It will create an economy of scale, as well as the required critical mass of stakeholders to offer a certification service that is of high quality, efficient, agile, and able to serve the scientific community in an age in which the demand for data sharing will only grow.

The activities of this WG are furthermore of great value to the overall certification goals. The DSA–WDS partnership will work under the umbrella of the RDA/WDS IG on Certification of Digital Repositories. The long-term goal of the IG is implementation of certification as a common service in order to stimulate the development of a global network of trusted digital repositories that meet international certification standards. This long-term goal calls for an incremental and iterative approach; consisting of a number of limited, targeted, relatively short-term and sequenced steps.

The first of these steps will be the DSA–WDS partnership: DSA and ICSU-WDS do not fully address the same targets. The envisioned work examining the commonalities and differences will therefore highlight the issues that arise from having different approaches, with the goal of possibly bringing together the two certification methods.

Secondly, developing a common organizational framework for ICSU-WDS and DSA will serve as an exemplar, and make it possible to clarify the requirements for opening up this authority at a later stage to other certification bodies.

Finally, mutual work performed by DSA and ICSU-WDS in this WG will be the starting point for a second stage encompassing other organisations that deal with data repository certification. Dissemination of the results of DSA–WDS collaboration, and preliminary discussions with other potential groups, will be managed by the Certification IG. In the initial instance, we are focusing on two community-driven organisations offering basic certification in order to comply with the schedule and practical deliverables expected of RDA-endorsed WGs. We hope to gain knowledge in this first phase of work that can then be applied more broadly.

Engagement with existing work in the area

A review of the literature shows that data sharing and trusted repositories go hand in hand. Kowalczyk and Shankar⁶ point out that for data sharing to be effective, the data (and their location) must be: (1) findable over time via consistent pointers, (2) preserved and accessible for the long term, and (3) of sufficient quality to be useable. Data repositories are a key component of scientific

⁶ Kowalczyk, S. & Shankar, K. (2013). Data Sharing in the Sciences. *Annual Review of Information Science and Technology*, 45(1), 247-294. doi:10.1002/aris.2011.1440450113

infrastructure, providing a needed mechanism for data access and sharing. By archiving and safeguarding data, repositories stimulate additional research. Indeed, Pienta et al.⁷ found that when social science data were shared formally through a repository, at least twice as many publications based on the original data were produced. Similar findings have also been shown for other disciplines.

As part of the Opportunities for Data Exchange project, Dallmeier-Tiessen et al.⁸ modelled a set of drivers, barriers, and enablers to describe factors that motivate, inhibit, and enable the sharing of research data. They suggest that barriers—such as the perceived trustworthiness of data, data usability, and pre-archive activities—can be overcome by a combination of actions, one of which is certification of data centres for data quality and usability by a trustworthy body. Kowalczyk and Shankar also suggest that data repositories require systematic evaluation for multiple reasons: the value of building these systems, their usefulness and usability, their importance to researchers, and other metrics. Repository assessment and certification can provide such evaluations, leading to greater transparency around the operation of repositories.

There are currently several options and levels for repository audit and certification, which has resulted in a confusing picture for repositories seeking to benchmark their efforts against a standard. As the grid in Appendix A illustrates, options range from basic certification involving relatively few requirements to full certification against the ISO 16363 standard. For the latter, a repository must demonstrate to external auditors its compliance with over 100 detailed requirements.

In Europe, an integrated framework for auditing and certifying of digital repositories has been generated. This framework consists of a three-level sequence of increasing trustworthiness:

1. *Basic Certification*, granted to repositories that obtain DSA certification.
2. *Extended Certification*, granted to Basic Certification repositories that additionally perform a structured, externally reviewed, and publicly available self-audit based on ISO 16363 or DIN 31644.
3. *Formal Certification*, granted to repositories that additionally to Basic Certification obtain a full external audit and certification based on ISO 16363 or equivalent DIN 31644.

We seek to coordinate certification at the basic level; streamlining certification for repositories lacking the resources to undertake full certification. Our ultimate goal is to include those initiatives listed above and in Appendix A to establish a global framework for certification. We believe that the DSA–WDS collaboration will stimulate additional assessment and certification efforts worldwide, thus strengthening the infrastructure related to data sharing and leading to a global system of federated and trusted repositories.

Work Plan

The main goals are to improve—and possibly combine—the current ICSU-WDS and DSA certification catalogues and procedures, thereby identifying synergies between these organisations and ultimately other organisations and initiatives. Three steps are planned:

⁷ Pienta, Amy M.; Alter, George C.; Lyle, Jared A. (2010). The Enduring Value of Social Science Research: The Use and Reuse of Primary Research Data. <http://hdl.handle.net/2027.42/78307>

⁸ Dallmeier-Tiessen, S., Darby, R., Gitmans, K., Lambert, S., Suhonen, J., Wilson, M., Science, H., et al. (2012). Compilation of results on drivers and barriers and new opportunities. <http://goo.gl/7y3NYT>

1. Mapping and comparing the guidelines, processes, and review procedures of DSA and ICSU-WDS in order to explore further collaboration and understand similarities and differences. Of particular interest will be examination of the review procedures within scientific networks, such as the ICSU Scientific Unions. Other important aspects to explore include the expertise of reviewers, as well as that of applicants.
2. Compiling a common catalogue of certification criteria. Both organisations have developed criteria for a repository function. The WG will first attempt to merge the generic parts of the WDS and DSA catalogues related to data repositories. It will then look to appropriately describe and group all the specialized functions that data services may possess, before finally establishing a catalogue of criteria to handle these.
3. Developing a common testbed, and its surrounding organisational framework, for peer review and certification based on the current WDS and DSA catalogues and certification procedures. The testbed will provide practical insight into the proposed common WDS–DSA catalogue and review process, thus enabling iterative improvements to those procedures. The testbed will be driven by the DSA Board and the WDS Scientific Committee. A pool of reviewers will be set up to test the common procedures developed.

Adoption Plan

The form and description of final deliverables of the WG

If consensus is reached, the certification procedures will be adopted by ICSU-WDS and DSA, and maintained as a common certification mechanism with a common pool of reviewers experienced in evaluating data repositories and services. It is envisioned that such a mechanism could expand beyond ICSU-WDS and DSA; involving more organisations and standards. This might possibly be the next stage in the process, and should be addressed at the level of the RDA/WDS IG on Certification of Digital Repositories.

The form and description of milestones and intermediate documents, code or other deliverables that will be developed during the course of the WG's work

Deliverables

- Comparative description and mapping of existing DSA and WDS certification catalogues (Month 3)
- Comparative description and mapping of existing DSA and WDS certification procedures (Month 6)
- Certification catalogue comprising DSA and WDS criteria (Month 9)
- Certification procedure as a common procedure of ICSU-WDS and DSA (Month 9)
- Testbed for peer review and certification driven by ICSU-WDS and DSA (Month 9–18)

A description of the WG's mode and frequency of operation

The WG will hold open teleconferences approximately every six weeks to discuss both assignments and progress towards the deliverables.

A description of how the WG plans to develop consensus, address conflicts, stay on track and within scope, and move forward during operation

DSA and ICSU-WDS are two examples of organizations that have attempted to set up mechanisms for basic certification; there may be more in the future.

Because DSA and WDS have existing identities, infrastructures, and web presences, it is necessary for them to reach agreement on outcomes that could affect how they operate. Any final

decisions that impact on the regulations or structures of ICSU-WDS or DSA remain entirely matters for the relevant boards, and not for the RDA-endorsed WG.

Decisions within the WG will be made by consensus. In the event that the WG cannot reach agreement, each member will have a vote, and decisions will be taken on a simple majority. It is expected that members will contribute and vote as individuals. Therefore, members will make it clear if their comments represent the views of DSA, ICSU-WDS, or another body or organisation. Project management will be supplied by the WDS International Programme Office (WDS-IPO).

The above paragraph should in no way be interpreted to mean that the WG is closed to participants not involved in the DSA or ICSU-WDS. Rather, we seek to conduct the work in an open atmosphere, involving all interested participants. One of our first jobs will be to contact individuals from other certification standards to apprise them of the work and to invite them to join the effort. Other RDA participants are welcomed also.

A description of the WG's approach to broader community engagement and participation

Webinars, conferences, and intermediate reports will be used to communicate findings and step-by-step deliverables to members of the IG on Certification of Digital Repositories, WDS Members, the DSA community, the wider RDA membership, and additional stakeholders and initiatives addressing the same topic. Of particular interest will be discussion of the results with ICSU Scientific Unions; especially, the International Union of Geology and Geophysics and its Associations. Generally, organisations such as the ICSU Scientific Unions have their own internal organisation and management, which includes the monitoring of their services.

Membership

- Lesley Rickards (UK, PSMSL, WDS-SC) **[Co-chair]**
- Mary Vardigan (USA, ICPSR, DSA Board) **[Co-chair]**
- Kevin Ashley (UK, Digital Curation Centre)
- Michael Diepenbroek (Germany, Pangaea, WDS-SC)
- Ingrid Dillo (The Netherlands, DANS, DSA Board)
- Françoise Genova (France, CDS, WDS-SC)
- Hervé L'Hours (UK, UK Data Archive, DSA Board)
- Guoqing Li (China, CEODE, WDS-SC)
- Jean-Bernard Minster (UCSD, and Chair of WDS Scientific Committee)
- Paul Trilsbeek (The Netherlands, MPI for Psycholinguistics, DSA Board)

Ex officio: Rorie Edmunds and Mustapha Mokrane (WDS-IPO)

Appendix A: [List of Certification Standards](#)

Name	Website	Year Started	# Certifications/ Members	#Requirements	Procedures	External Audit Required?	Domains Covered	Schedule for Recertification	Effort Required	Geographic Distribution
Trusted Digital Repositories (European Framework)	http://www.trusteddigitalrepository.eu/Site/Trusted%20Digital%20Repository.html	2011								
Data Seal of Approval	www.datasealofapproval.org/	2008	20	16	Repository applies online; adds evidence and URLs for 16 guidelines and assigns ratings; DSA Board member reviews application; DSA awarded and displayed with a logo and linkable application	No	Humanities, social science	2 years	16 hours	Europe, USA, Australia
World Data System	http://www.icsu-wds.org	2011	53 regular members; 7 network members; 2 partner members; 13 associate members	17	Organization completes expression of interest; completes application; two reviewers review application; accreditation received with the right to display logo	No, but is possible	Physical and life sciences but branching out to social sciences	3-5 years	24 hours	Global -- # regular members in different parts of the world (Australia 2, Asia 15, Africa 1, Europe 20, North America 18)
Trustworthy Repositories Audit and Certification (TRAC)	http://www.crl.edu/archiving-preservation/digital-archives/metrics-assessing-and-certifying-0	2005	4	84		Yes	Various		Weeks	U.S., Canada,
Trustworthy Digital Repositories (TDR) Checklist, ISO 16363	http://public.ccsds.org/publications/archive/	2011		Over 100		Yes	Various		Weeks	
DRAMBORA	http://www.repositoryaudit.eu/	2007	Over 18	10	Organization completes self-audit using Drambora toolkit	No	Various		24-40 hours	Europe, Japan, U.S. S.
Nestor Seal for Trustworthy Digital Archives	http://www.langzeitarchivierung.de/Subsites/nestor/EN/n	2004	15 nestor partners, no seals granted yet	34	Organization expresses interest; completes application; two reviewers review application; accreditation received with the right to display logo	No	Various, with a focus on public memory institutions	No recertification required, although recommended (seal displays date)	Maximum three months	Germany, Europe
DIN 31644 -- Criteria for Trustworthy Digital Archives	http://www.nabd.din.de/cmd?level=tpl-art-detailansicht&committee	Published 2012		34			Various		Maximum three months when used for nestor accreditation	Germany, Europe
World Meteorological Organization (WMO) Information System	http://goo.gl/BzpG7N	2009		42	(1) Candidate organizations submit questionnaire plus evidence of their capabilities and results of demonstration tests; (2) the ET-GDDP assesses the information; (3) ET-GDDP endorses the candidate organization	Yes, but restricted to the technical service supplied	Atmospheric data		24 hours	Global