



Improving Opportunities for New Value of Open Data: Assessing and Certifying Research Data Repositories

COLLECTION:
INTERNATIONAL
DATA WEEK

ESSAY

ROBERT R. DOWNS 

**Author affiliations can be found in the back matter of this article*

][ubiquity press

ABSTRACT

Investments in research that produce scientific and scholarly data can be leveraged by enabling the resulting research data products and services to be used by broader communities and for new purposes, extending reuse beyond the initial users and purposes for which the data were originally collected. Submitting research data to a data repository offers opportunities for the data to be used in the future, providing ways for new benefits to be realized from data reuse. Improvements to data repositories that facilitate new uses of data increase the potential for data reuse and for gains in the value of open data products and services that are associated with such reuse. Assessing and certifying the capabilities and services offered by data repositories provides opportunities for improving the repositories and for realizing the value to be attained from new uses of data. The evolution of data repository certification instruments is described and discussed in terms of the implications for the curation and continuing use of research data.

CORRESPONDING AUTHOR:

Robert R. Downs

Center for International Earth
Science Information Network
(CIESIN), Columbia University,
Palisades, NY 10964, USA

rdowns@ciesin.columbia.edu

KEYWORDS:

Data Repositories; Data Value;
Certification; Assessment; Data
Management; Data Reuse

TO CITE THIS ARTICLE:

Downs, RR. 2021. Improving
Opportunities for New Value
of Open Data: Assessing and
Certifying Research Data
Repositories. *Data Science
Journal*, 20: 1, pp. 1–11. DOI:
[https://doi.org/10.5334/dsj-
2021-001](https://doi.org/10.5334/dsj-2021-001)

Society invests in science and also benefits from the results of scientific research activities that have been conducted. As part of scientific and scholarly research, significant investments are made to collect data, which enables new activities to build on the efforts of earlier work. Studies are conceptualized, investigative teams are formed, needed resources are identified, support is acquired, procedures are planned, instruments are designed and produced, and experiments and observations are conducted to collect the data. Often, data collection and other research investments are supported by academic institutions, government agencies, or philanthropic organizations. After investing so much to obtain data, it is important for the data to be made publicly accessible as open data, if possible, so that researchers, planners, decision-makers, educators, students, and the general public have an opportunity to use the data without restrictions and realize the continuing value of such investments.

In addition to providing evidence of the methods and the results of a research study to facilitate verification of the conclusions, open data can be reused by subsequent investigations for new purposes. New hypotheses are investigated and new research questions are pursued if the data are freely available for other investigators to use, as revealed by data citations in various disciplines, despite inconsistent data citation practices (He & Nahar, 2016; Khan, Pink & Thelwall, 2020; Khan, Thelwall & Kousha, 2019; Robinson-García, Jiménez-Contreras & Torres-Salinas, 2016). Researchers also could use available data to conduct replication, reproduction, and comparison studies. Furthermore, new open data products and services could be created by integrating freely accessible open data with other open data that also are freely available for use. Leveraging such opportunities, the scientific community and non-scientists will continue to realize the benefits of data collection efforts long after the publication of the initial results of the studies that were conducted to collect the data. The producers of such data also can receive additional recognition for their research contributions as the use of their data enables new studies to be conducted and new data products and services to be created, fostering more research. By curating, preserving, and enabling new uses of data by additional users, the value of the initial investment of limited research funds also continues to be extended to propagate new studies, new results, and new benefits to society. In effect, benefits from open data sharing are engendered by both creators and users of data.

SOCIETAL VALUE OF OPEN DATA

Many datasets can be reused for purposes that were not necessarily foreseen by those who originally collected the data. Open data that are preserved, curated, and disseminated to foster diverse uses offer more opportunities to increase the value of the data through reuse. Such reuse can include non-scientific activities as well as those that contribute to science and scholarship. Researchers from the original data collection team or from the same discipline might want to extend the research that was conducted when the data were initially collected and analyzed. Scholars and researchers from different disciplines may see new opportunities for data analyses to investigate phenomenon that were not necessarily considered when collecting the data originally. New investigations may leverage the potential for integrating the data with data products collected within other disciplines to create entirely new data products and services. The potential for new value being derived from data increases with each act of reuse.

In addition to members of the research community deriving new value from data, other members of society also have the potential to derive new value from data. Journalists, educators and students, planners and policy-makers, representatives of commercial and non-profit entities, and members of the general public also generate value by reusing data. For example, planners and decision-makers are able to explore issues of interest that may be unrelated to the studies conducted by the original data producers. Educators have the ability to use the data to provide experiences for their students who may be studying techniques for analyzing data for various purposes. Members of the general public also may have interest in research data due to the location represented by the data or the time period when the data were collected, for example. Disseminating data products for use by broad audiences, including non-scientists, offers opportunities for shared understanding across society (Baker, Duerr & Parsons, 2015). Given such possibilities for new and diverse uses of data that leverage the initial investments in data collection, opportunities for enabling future use of the data should be

considered. Facilitating data reuse by diverse audiences, including the general public, requires efforts to reduce barriers that might prevent reuse, including such concerns as navigability, interpretability, accessibility, and analyzability (Grand et al., 2016; Levin et al., 2016). Data repositories that enable the use of data by diverse audiences, as well as enabling use by the designated community, improve the potential usefulness of data and increase opportunities for contributing to both the scientific value and the societal value of data.

If data are determined as having potential value for future use, creating opportunities for new uses enables future data users to explore ways to realize additional value from the previous investments that were made to collect the data. Documenting and submitting such datasets to a data repository that represents the domain of potential users offers current and future users with opportunities to realize such value. Enabling additional uses by scientists from other disciplines, as well as non-scientists, heightens that value. Archival activities should facilitate secondary value by enabling new uses (Borgerud & Borglund, 2020; Schellenberg, 1956, as cited in Jaillant, 2019). Infrastructure that enables new uses further increases the value of data as a result of such second-order benefits of reuse (Carrera & Hoyt, 2006). As new users make new discoveries, new value for society also will be realized from using the data that were previously collected. With appropriate recognition for data sharing, submitting datasets to a relevant data repository also enables the original data producers to realize new value from their initial investment in the data collection effort (Krzton, 2018). Even though attitudes vary among scientists for sharing data that they have collected (Tenopir et al., 2018), data producers and others who have contributed to the collection and sharing of open data should receive rewards for such contributions, including recognition for their data collection and sharing efforts from journals, sponsors, and promotion committees (Bierer, Crosas & Pierce, 2017). Recognizing contributions that enable the use of open data also can serve as an incentive and an inspiration for open science (Burgelman et al., 2019; Perrier, Blondal & MacDonald, 2020).

ADDED-VALUE OF DATA REPOSITORIES

Data repositories, such as research data centers and scientific archives, contribute to the potential for new value by facilitating reuse of data through the management and dissemination of open data products and services. By depositing data in a repository for dissemination as open data, data producers contribute to open science and to the reuse of their data for new purposes. Data repositories, particularly specialist or domain repositories, facilitate the reuse of data within and across disciplines by offering extensive data curation and stewardship services (Boté & Térmens, 2019). Quality control, documentation, and peer-review also are necessary data curation functions of repositories that facilitate reuse (Koltay, 2020). When open data are described and curated, potential users have an opportunity to explore data products to determine their potential usefulness for a particular purpose. Such data stewardship services add value by enabling data to be reused for new research, decision-making, and learning activities that otherwise might not be possible. Recognizing the importance of data stewardship to facilitate reuse, data repositories work with both data producers and data users to ensure that data are reusable. Meeting the need to foster open science by enabling the reuse of open data, data producers select and deposit their data in a repository that they can collaborate with on data curation to ensure that their data will be reusable by broad audiences and create new value for society.

IMPROVING RESEARCH DATA REPOSITORIES

Selection of the data repository, where future reuse of the data will be enabled, also contributes to the new value to be attained from the initial investments in data collection. Within this context, the characteristics of the data repository, such as disciplines and communities served, data policies, costs of deposit and preservation, discovery and access capabilities, user support services, sustainability, and reputation, as well as other selection criteria, are important. In addition, data depositors may be interested in the trustworthiness of the repository that will be responsible for providing stewardship and enabling continuing access to the data that have been submitted for future use. Characteristics of TRUST, including “transparency, responsibility, user focus, sustainability, and technology” (Lin et al., 2020, 4), contribute to the capabilities of a data repository that provide the stewardship to facilitate reuse of digital data

over time. Perceptions of reputation, societal commitment, and mission also may influence the trustworthiness of a data repository (Yoon, 2014). Knowing whether a candidate data repository has been independently assessed to be appropriately managing data for future use also can help data producers determine where to submit their data. Identifying a data repository that has been assessed and certified as trustworthy offers assurances to data depositors and other stakeholders that the repository has been independently evaluated in terms of its data stewardship capabilities and data curation practices.

Several instruments have emerged in recent decades to evaluate the trustworthiness of data repositories. Since the initial development of the Open Archival Information System (OAIS) Reference Model by the Consultative Committee for Space Data Systems in 2002 (CCSDS, 2002; 2012), and subsequent publication by the International Standards Organization (ISO) as ISO 14721:2003 and ISO 14721:2012, data repositories have strived to meet the requirements for becoming trustworthy (Garrett, et al., 2015). Following the initial ISO publication of the OAIS Reference Model, several instruments have been developed for measuring the trustworthiness of data repositories, since the OAIS Reference Model is a framework and not an instrument for measuring or assessing data repository trustworthiness. Early assessment instruments that were developed include the Data Seal of Approval (DSA) (Dillo & de Leeuw, 2015); the NESTOR Seal for Trustworthy Digital Archives that was published as the Deutsches Institut für Normung (DIN) 31644 standard (Harmsen, et al., 2013)); TRAC, the Trustworthy Repositories Audit & Certification: Criteria and Checklist (CRL, 2007); and DRAMBORA, the Digital Repository Audit Method Based on Risk Assessment (Innocenti, 2007), as well as others.

While the CCSDS and the ISO were conducting the reviews of the draft ISO 16363 standard, Audit and Certification of Trustworthy Digital Repositories (2011), which also emanated from the OAIS Reference Model, the European Framework for Audit and Certification of Digital Repositories was signed in 2010 by representatives of the DSA, DIN 31644, and ISO 16363 to offer guidance for data repositories seeking certification (Callaghan et al., 2014; Klump, 2011). Subsequently, the International Council of Science (ICSU) World Data System (WDS), which had been established from the redesign of the World Data Centers and has since become the International Science Council (ISC) WDS, published Certification of WDS Members to describe the evaluation criteria and procedures for WDS membership (2012). The ISO review of the CCSDS draft was completed and published as ISO 16363:2012 (2012). Also, the Data Management Maturity (DMM) Framework was established by the American Geophysical Union (Stall, 2015). After many repositories had been certified in compliance with the WDS requirements and with the DSA requirements, these two assessment instruments were merged to create the Core Trustworthy Data Repositories Requirements that are offered by the CoreTrustSeal (L'Hours, Kleemola & de Leeuw, 2019; Rickards et al., 2016). **Figure 1** depicts the evolution of several instruments that have been developed and used, internationally, for assessing trustworthy data repositories.

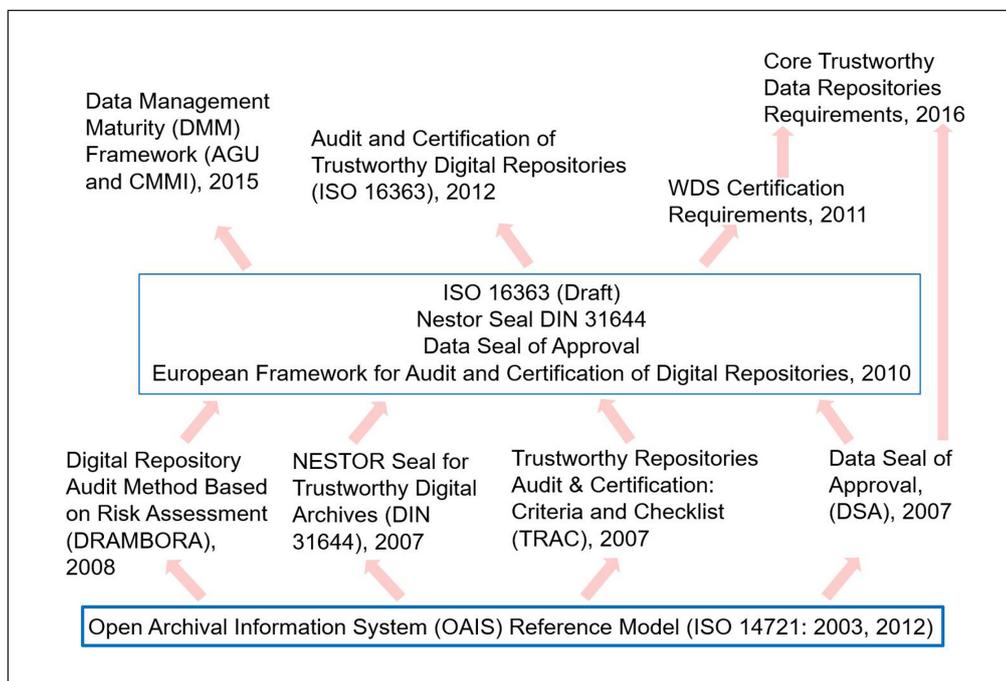


Figure 1 Evolution of Instruments for Assessing Trustworthy Data Repositories (Downs, 2019).

With many of these data repository assessment instruments, in addition to requesting an independent assessment or audit of a digital repository to determine compliance with the requirements of a particular instrument, repositories also can conduct a self-assessment. Using these instruments to conduct self-assessments enables repositories to determine whether they meet the requirements and enables the repositories to identify requirements for which additional capabilities need to be developed to attain compliance. These assessment instruments are being used by repositories to improve their management, infrastructure, policies, and practices and, in many cases, to also obtain certification. Receiving certification provides data repositories with a credential that attests to their capabilities and affords their recognition as trustworthy data repositories. In addition to repository staff members recognizing the value of attaining certification to be designated as trustworthy (Donaldson et al., 2017), several incentives and benefits, both extrinsic and intrinsic, for obtaining digital repository certification have been reported in the literature (Lindlar & Schwab, 2018). In addition, most repositories that have attained certification as trustworthy have publicly declared their achievement by posting information about the certification on their website (Donaldson, 2020).

Characteristics of data also contribute to the potential for new value that would be realized from subsequent use. Attainment of the FAIR Principles, for “findable, accessible, interoperable, and reusable” research objects, also has been recognized as necessary for enabling use (Wilkinson, et al., 2016, 3). From this perspective, the FAIR Principles also are relevant to the curation practices that enable future use and should be considered as integral when providing capabilities for data curation. Meeting the requirements associated with the FAIR Principles, in conjunction with certification of digital repositories, also has become a consideration for realizing the value of data (Koers et al., 2020; Mokrane & Recker, 2019). The evolution of digital repository certification requirements will need to consider the assessment of capabilities to meet aspects of both the TRUST Principles and the FAIR Principles. Other efforts, internationally, also contribute to the evolution of digital repository certification instruments, including, for example, the advent of maturity models for data. (Peng, et al., 2018). Similarly, other international efforts are influencing practices for improving data stewardship. The Global Earth Observations System of Systems (GEOSS) Data Management Principles (GEO, 2020) and the GEOSS Data Sharing Principles offer guidance to improve practices for the stewardship and sharing of data, particularly as open data (GEO, 2020).

CHALLENGES FOR THE EVOLUTION OF DATA REPOSITORY CERTIFICATION REQUIREMENTS

With the designation of the CoreTrustSeal data repositories requirements as the instrument that is being used for certification and re-certification of members of the WDS, several research data centers and other repositories around the globe are using these requirements to improve and certify their capabilities for managing and curating data (Core Certified Repositories, 2020). Even though the same CoreTrustSeal instrument is being used within many data repositories, the diversity of repositories and the differences among them often necessitates the application of individual approaches to attain compliance with the requirements for trustworthiness. For example, in serving its Designated Community, a repository may offer unique features to address particular disciplinary needs or even local context. As technology changes and community expectations increase, further innovation among data repositories also can be expected. By exploring and comparing the commonalities and differences among data repositories in terms of their efforts to attain capabilities for trustworthiness, recommended practices can be shared with more data repositories to meet the challenges of serving users of research data. Sharing of such practices offers opportunities for integration of technologies and for increasing efficiencies attained by the repositories that continually strive to improve their capabilities for enabling the use of data in a trustworthy manner. Data repositories that steadfastly improve their capabilities for using data will contribute to the evolving community of data repositories and to the new value that is gained from the use of the data that these repositories are sharing. The evolution of practices within the digital data stewardship community also will necessitate evolution of the instruments for assessing data repositories.

The CoreTrustSeal serves as an example of such evolution as it has since been revised in accordance with plans for reviewing the requirements every three years (CoreTrustSeal Standards and Certification Board, 2019). CoreTrustSeal also has begun exploring opportunities

to further meet the needs of the data repository community by expanding the types of certification that would be offered. Since emanating from the merging of the DSA and the WDS requirements for certification, CoreTrustSeal has been certifying many domain repositories that each serve a specified designated community and meet the certification requirements. However, recognizing that generalist repositories and related service providers also need to be certified in recognition of their capabilities, CoreTrustSeal has been conducting discussions with the broader community of domain repositories, generalist repositories, and data service providers to identify the requirements and the associated certifications that could be offered to meet the needs of the broader community (CoreTrustSeal, 2020). Like the CoreTrustSeal, other instrument developers also may need to consider the potential of broadening the scope of their instruments to reflect the evolving infrastructure for digital data and related research resources.

Generalist repositories, such as institutional repositories and large digital repositories that serve broad audiences, including non-scientists, recognize the importance of being entrusted with the responsibility of providing stewardship for the digital objects that they are being asked to manage. Likewise data service providers, such as those that offer storage, processing, access or other services to fulfill digital curation needs during the data lifecycle, also recognize the importance of being a part of the chain of trust that can be relied upon when providing services to support digital stewardship. Obtaining formal certification could offer evidence of attaining trustworthiness while assisting the entity that is seeking certification to improve its practices to be worthy of such trust. Likewise, pursuing periodic recertification would provide such entities with opportunities to continually earn and renew that trust while improving to meet the evolving requirements that are necessary for any standard to progress with changes in expectations and technology. Navigating the differences in requirements for attaining certification as trustworthy will be challenging as domain repositories, generalist repositories, and related service providers refine their roles within the digital data stewardship industry.

CONCLUSIONS

For research processes to continue effectively within the digital realm, it is necessary to establish, maintain, and verify sustainable infrastructure for managing and enabling access to data so that such capabilities can support current and future investigations. Current and future certification requirements will need to be developed by the communities that contribute to and are served by data repositories throughout the data lifecycle. In addition to reflecting the short and long-term objectives of data stewards, repository certification requirements also could reflect the evolving needs and workflows of data producers and users, as well as the perspectives of repository developers and service providers, tool developers, research institutions, sponsors, and publishers, for example. Improved channels of communication among such stakeholders, across repositories, disciplines, and international borders can help to identify common needs for improving the certification of data repositories, along with ways to address such needs by working together to improve the value of data. Similar to peer-review of research objects and institutions, community-review of data repositories offers opportunities to contribute to the capabilities needed to support scientific progress and the reuse of open data into the future.

The evolution of digital repository certification instruments appears to be contributing to the research data community's ability to navigate the abundance of issues associated with the development and provision of infrastructure for managing and enabling continuing use of data. Certification instruments are being improved over time and these, in turn, are being used to improve policies, procedures, skills, tools, services, and other capabilities for data repositories to enable data reuse on an ongoing basis. Research on the assessment and certification of data repositories also is needed to contribute to the improvement of approaches for managing and sharing data. Education, training, and the provision of learning resources that emphasize the importance of data stewardship policies and practices also is needed to improve understanding about the capabilities for enabling current and long-term use of data products and services by diverse communities, including those from disparate disciplines, as well as the general public.

The continuing evolution of instruments and processes for certifying data repositories is necessary to ensure that data are managed effectively and are available as open data for reuse and for attaining the value that can result from reuse. As certification instruments and processes evolve, it will be necessary for data repositories to leverage community resources

to improve while meeting the expectations of their user communities. Along with the new value that is being derived from the reuse of data, costs are incurred for enabling the reuse of data. Recognizing that data repositories are a critical part of the global research infrastructure and that diverse capabilities are needed to support the reuse of data, how might the research community and sponsors further nurture the development, operation, and certification of these facilities?

In consideration of the kinds of funding often obtained for operating research data repositories, it also will be necessary to maintain a balance between the costs of obtaining and maintaining certification and the costs of operating a repository so that the process is feasible, without creating excessive barriers for repositories to be certified. Such costs include improvements needed to prepare and apply for certification as well as the cost of any fees for procuring certification. Increases to current requirements could be identified through community participation so that infrastructure improvements for meeting new requirements can be planned for and conducted incrementally to fit within the budgets and resource constraints of data repositories, especially data repositories that rely on grants and contracts to sustain their operations. Similarly, the costs of certification fees cannot be so large as to prohibit small data repositories from obtaining and periodically renewing certification. If the costs of meeting new requirements and obtaining certification exceed the budgets of research data repositories, such repositories will be at risk of being eclipsed by large ventures that have the resources to incur the technological, administrative, and financial costs that are beyond the reach of smaller data repositories.

In addition, the evolution of users' expectations, technologies, standards, certification instruments, and requirements will place demands on repositories that will need to be met with improved efficiencies in terms of operations and for the development of improvements. Data curation, stewardship, and usage capabilities also will need to be enhanced. Data repositories also will need to reduce the potential resource demands that stem from continuous improvement and efforts to maintain certification by sharing resources within and across repository communities. Resource sharing offers an opportunity to reduce costs, improve efficiencies, and leverage economies of scale while improving capabilities to serve broader communities along with the designated community of users. Options for resource sharing include utilizing common infrastructure, services, and tools, as well as collaborating and sharing knowledge and competencies with other members of the international data repository community. In light of the potential benefits of resource sharing, how can resource sharing be facilitated among data repositories in an efficient and mutually-beneficial manner?

Cooperation among other community stakeholders also can contribute to the sustainability of repositories and to the value that can be attained from the reuse of data. Research sponsors could recognize the need to provide sustainable infrastructure for managing and facilitating the reuse of data, especially those data for which they already have invested to support their collection and initial use. Improving the capabilities and sustainability of repositories can facilitate the reuse of data by a variety of future users, beyond the disciplines represented by the data producers. Considering the additional gains that can be achieved by leveraging initial investments in data, how can sponsors ensure that repositories provide sustainable services to improve the value of data?

Furthermore, in addition to publishing reports of their work, data producers should have an opportunity to make informed decisions to facilitate the future use of their data. Early planning to begin preparing data for reuse and to select a repository can reduce the need for hasty decisions later, enabling data producers to choose a repository that they can collaborate with to foster broad reuse of their data by new users as open data. Recognizing that workflow changes are needed, how might journals, especially data journals, improve guidance to data producers on the selection of a data repository for the dissemination of data for reuse as open data?

A change in the research culture also is needed to facilitate the reuse of open data and foster open science. The value of open data is an outcome of a research ecosystem that requires participation by various global stakeholders that contribute to and benefit from scientific and scholarly information resources. In addition to contributing to the production, curation, preservation, dissemination and support of open data and related research resources, members of the ecosystem have an opportunity to contribute to the individual efforts that enable value to be derived from open data. Scientific and scholarly communities, as well as members of

society, can recognize the new value that is generated by producing, curating, and sharing open data when they are deciding on the allocation of resources and rewards. Recognition of the value of open data also includes proper data citation when data have been used to produce scholarly and scientific publications. Furthermore, review panels and promotion committees can contribute to the value of open data by acknowledging and recognizing data producers, curators, sponsors, repositories and hosting institutions for the contributions to science, scholarship, and society that are realized by the production, curation, and sharing of open data. Such acknowledgement and recognition could incentivize the contributions of stakeholders who contribute to the value of open data throughout the data lifecycle. Considering the need for cultural change across many institutions to further improve the value of data, we might ask ourselves the following question. How can we propagate needed changes throughout the research community to recognize and incentivize activities that contribute to the value of data to science, scholarship, and society?

ACKNOWLEDGEMENTS

The author is grateful for suggestions received from the anonymous 2018 SciDataCon reviewers to improve the extended abstract upon which this essay is based and also is grateful for the thoughtful recommendations offered by the anonymous reviewers of the Data Science Journal. The essay is partially based on the author's presentation, Improving Opportunities for New Data Value by Assessing and Certifying Scientific Data Repositories, ID 559, which was delivered to SciDataCon2018, as part of International Data Week (IDW) 2018 in Gaborone, Botswana, during the session: Improving Data Repository Trustworthiness, Part 1 (ID 149) that was held on Tuesday, 6 November 2018. This essay reflects the perspectives of the author and not necessarily those of any institution or agency.

FUNDING INFORMATION

The creation of this essay was supported by the National Aeronautics and Space Administration (NASA) under Contract 80GSFC18C0111 for the Socioeconomic Data and Applications Distributed Active Archive Center (DAAC).

COMPETING INTERESTS

The author has been involved in the development, review, adoption, and use of several of the data repository assessment instruments described in the essay and is currently a member of the CoreTrustSeal Standards and Certification Board. The author also is currently a member of the Editorial Board of the Data Science Journal.

AUTHOR CONTRIBUTIONS

Robert R. Downs designed, drafted, and submitted the original version of the manuscript to the Data Science Journal, based on the extended abstract that he had submitted to the SciDataCon 2018 Conference and on subsequent revisions completed in response to comments received from the anonymous reviewers. He subsequently revised the manuscript to reflect current developments and recommendations offered by the anonymous reviewers of the Data Science Journal.

AUTHOR INFORMATION

Dr. Robert R. Downs is the senior digital archivist and acting head of cyberinfrastructure and informatics research and development at CIESIN, the Center for International Earth Science Information Network, in the Earth Institute of Columbia University. He is an elected co-chair of the Group on Earth Observations (GEO) Data Working Group, co-chair of the Research Data Alliance (RDA) Interest Group on Repository Platforms for Research Data, Standards and Certification Board Member for the CoreTrustSeal, and Executive Committee Member for the EarthCube Council of Data Facilities. He is a co-chair of the Earth Science Information Partners (ESIP) Information Quality Cluster and serves on the ESIP Governance Committee and on the Editorial Board of the CODATA Data Science Journal.

Robert R. Downs  orcid.org/0000-0002-8595-5134

Center for International Earth Science Information Network (CIESIN), Columbia University, Palisades, NY 10964, USA

REFERENCES

- Baker, KS Duerr, RE and Parsons, MA.** 2015. Scientific knowledge mobilization: Co-evolution of data products and designated communities. *International Journal of Digital Curation*, 10(2): 110–135. DOI: <https://doi.org/10.2218/ijdc.v10i2.346>
- Bierer, BE, Crosas, M and Pierce, HH.** 2017. Data Authorship as an Incentive to Data Sharing. *The New England Journal of Medicine*, 376: 1684–1687. DOI: <https://doi.org/10.1056/NEJMs1616595>
- Borgerud, C and Borglund, E.** 2020. Open research data, an archival challenge? *Archival Science*, 20: 279–302. DOI: <https://doi.org/10.1007/s10502-020-09330-3>
- Boté, JJ and Térmens, M.** 2019. Reusing Data: Technical and Ethical Challenges. *DESIDOC Journal of Library & Information Technology*, 39(6): 329–337. DOI: <https://doi.org/10.14429/djlit.39.06.14807>
- Burgelman, JC, Pascu, C, Szkuta, K, Von Schomberg, R, Karalopoulos, A, Repanas, K and Schoupe, M.** 2019. Open science, open data and open scholarship: European policies to make science fit for the 21st century. *Frontiers in Big Data*, 2: 43. DOI: <https://doi.org/10.3389/fdata.2019.00043>
- Callaghan, S, Tedds, J, Kunze, J, Khodiyar, V, Lawrence, R, Mayernik, MS, Murphy, F, Roberts, T and Whyte, A.** 2014. Guidelines on Recommending Data Repositories as Partners in Publishing Research Data. *International Journal of Digital Curation*, 9(1): 152–163. DOI: <https://doi.org/10.2218/ijdc.v9i1.309>
- Carrera, F and Hoyt, L.** 2006. From plan-demanded data to plan-ready information: A rationale for comprehensive urban knowledge infrastructures. *Journal of Urban Technology*, 13(2): 3–23. DOI: <https://doi.org/10.1080/10630730600871999>
- Consultative Committee for Space Data Systems.** 2002. *Reference Model for an Open Archival Information System (OAIS)*. Also published as ISO 14721:2002. Superseded by CCSDS, 2012.
- Consultative Committee for Space Data Systems.** 2011. *Audit and Certification of Trustworthy Digital Repositories*. Also published as ISO 16363:2012. Available at <https://public.ccsds.org/Pubs/652x0m1.pdf> [Last Accessed 30 September 2020].
- Consultative Committee for Space Data Systems.** 2012. *Reference Model for an Open Archival Information System (OAIS)*. Also published as ISO 14721. Available at <https://public.ccsds.org/Pubs/650x0m2.pdf> [Last Accessed 30 September 2020].
- CoreTrustSeal.** 2020. *Meeting Community Needs: Exploring Opportunities for Expanding CoreTrustSeal Certification to Meet Community Needs*. Available at <https://www.coretrustseal.org/why-certification/meeting-community-needs/> [Last Accessed 30 September 2020].
- CoreTrustSeal Standards and Certification Board.** 2019. *CoreTrustSeal Trustworthy Data Repositories Requirements 2020–2022*. (Version v02.00-2020-2022). Zenodo. DOI: <http://doi.org/10.5281/zenodo.3638211>
- Dillo, I and de Leeuw, L.** 2015. Ten years back, five years forward: The Data Seal of Approval. *International Journal of Digital Curation*, 10(1): 230–239. DOI: <https://doi.org/10.2218/ijdc.v10i1.363>
- Donaldson, DR.** 2020. Certification information on trustworthy digital repository websites: A content analysis. *Plos One*, 15(12): e0242525. DOI: <https://doi.org/10.1371/journal.pone.0242525>
- Donaldson, DR, Dillo, I, Downs, RR and Ramdeen, S.** 2017. The Perceived Value of Acquiring Data Seals of Approval. *International Journal of Digital Curation*, 12(1): 130–151. DOI: <https://doi.org/10.2218/ijdc.v12i1.481>
- Downs, RR.** 2019. International Standards for Trustworthy Data Repositories. In: *National Institutes of Health (NIH) Trustworthy Data Repositories Workshop*. National Institute for Allergy and Infectious Diseases (NIAID). Rockville, MD, 8–9 Apr 2019. DOI: <https://doi.org/10.7916/d8-h7xf-ha70>
- European Framework for Audit and Certification of Digital Repositories.** Available at <http://trusteddigitalrepository.eu/> [Last Accessed 28 September 2020].
- Garrett, J, Giaretta, D, Downs, RR, Hughes, JS, Conrad, M, Ambacher, B, Lambert, S and Tibbo, H.** 2015. Certification of Digital Archives – A Brief History and Status Report. In: *Proceedings of the 2015 PV Conference, Ensuring Long-Term Data Preservation, and Adding Value to Scientific and Technical Data*, Darmstadt, Germany, Nov 3–5 2015. Available at http://www.eumetsat.int/website/home/News/ConferencesandEvents/DAT_2447480.html [Last Accessed 30 September 2020].
- Grand, A, Wilkinson, C, Bultitude, K and Winfield, AF.** 2016. Mapping the hinterland: Data issues in open science. *Public Understanding of Science*, 25(1): 88–103. DOI: <https://doi.org/10.1177/0963662514530374>
- Group on Earth Observations.** 2020. *Global Earth Observations System of Systems (GEOSS) Data Management Principles*. Available at http://earthobservations.org/open_eo_data.php# [Last Accessed 28 September 2020].

- Group on Earth Observations.** 2020. *Global Earth Observations System of Systems (GEOSS) Data Sharing Principles*. Available at http://earthobservations.org/open_eo_data.php# [Last Accessed 28 September 2020].
- Harmsen, H, Keitel, C, Schmidt, C, Schoger, A, Schrimpf, S, Stürzlinger, M and Wolf, S.** 2013. Explanatory notes on the Nestor seal for trustworthy digital archives. *Nestor Certification Working Group*, Available at <http://nbn-resolving.de/urn:nbn:de:0008-2013100901> [Last Accessed 28 September 2020].
- He, L and Nahar, V.** 2016 Reuse of scientific data in academic publications. *Aslib Journal of Information Management*, 68(4): 478–494. DOI: <https://doi.org/10.1108/AJIM-01-2016-0008>
- ICSU World Data System.** 2012. *Certification of WDS Members*. Available at <https://www.worlddatasystem.org/files/wds-certification-summary-11-june-2012.pdf> [Last Accessed 28 September 2020].
- Innocenti, P.** 2007. *Risk management foundations for digital libraries: DRAMBORA (Digital Repository Audit Method Based on Risk Assessment)*. <https://strathprints.strath.ac.uk/70481/> [Last Accessed 28 September 2020].
- International Organization for Standardization.** 2012 ISO 14721:2012 (CCSDS 650.0-M-2) *Space data and information transfer systems — Open archival information system (OAIS) — Reference model*. Available at <https://www.iso.org/standard/57284.html> [Accessed 30 September 2020].
- International Organization for Standardization.** 2012 ISO 16363:2012 (CCSDS 652.0-R-1) *Space data and information transfer systems — Audit and certification of trustworthy digital repositories*. Available at <https://www.iso.org/standard/56510.html> [Accessed 30 September 2020].
- Jaillant, L.** 2019. After the digital revolution: working with emails and born-digital records in literary and publishers' archives. *Archives and Manuscripts*, 47(3): 285–304. DOI: <https://doi.org/10.1080/01576895.2019.1640555>
- Khan, N, Pink, CJ and Thelwall, M.** 2020. Identifying Data Sharing and Reuse with Scholix: Potentials and Limitations. *Patterns*, 1(1): 100007. DOI: <https://doi.org/10.1016/j.patter.2020.100007>
- Khan, N, Thelwall, M and Kousha, K.** 2019. *Data Citation and Reuse Practice in Biodiversity – Challenges of Adopting a Standard Citation Model*. Available at <https://wlv.openrepository.com/handle/2436/623005> [Last Accessed 1 October 2020].
- Klump, J.** 2011. Criteria for the Trustworthiness of Data Centres. *D-Lib Magazine*, 17(1/2). DOI: <https://doi.org/10.1045/january2011-klump>
- Koers, H, Bangert, D, Hermans, E, van Horik, R, de Jong, M and Mokrane, M.** 2020. Recommendations for Services in a FAIR Data Ecosystem. *Patterns*, 1(5): 100058. DOI: <https://doi.org/10.1016/j.patter.2020.100058>
- Koltay, T.** 2020. Quality of Open Research Data: Values, Convergences and Governance. *Information*, 11(4): 175. DOI: <https://doi.org/10.3390/info11040175>
- Krzton, A.** 2018. Supporting the Proliferation of Data-Sharing Scholars in the Research Ecosystem. *Journal of eScience Librarianship*, 7(2): e1145. DOI: <https://doi.org/10.7191/jeslib.2018.1145>
- L'Hours, H, Kleemola, M and de Leeuw, L.** 2019. CoreTrustSeal: From academic collaboration to sustainable services. *IASSIST Quarterly*, 43(1): 1–17. DOI: <https://doi.org/10.29173/iq936>
- Levin, N, Leonelli, S, Weckowska, D, Castle, D and Dupré, J.** 2016. How do scientists define openness? Exploring the relationship between open science policies and research practice. *Bulletin of Science, Technology & Society*, 36(2): 128–41. DOI: <https://doi.org/10.1177/0270467616668760>
- Lin, D, Crabtree, J, Dillo, I, Downs, RR, Edmunds, R, Giarretta, D, De Giusiti, M, L'Hours, H, Hugo, W, Jenkyns, R, Khodiyar, V, Martone, M, Mokrane, M, Navale, V, Petters, J, Sierman, B, Sokolova, DV, Stockhause, M and Westbrook, J.** 2020. The TRUST Principles for Digital Repositories. *Scientific Data*, 7(144): 1–5. DOI: <https://doi.org/10.1038/s41597-020-0486-7>
- Lindlar, M and Schwab, F.** 2018. All that work... for what? Return on investment for trustworthy archive certification processes—a case study. In: *15th International Conference of Digital Preservation, iPRES*. DOI: <https://doi.org/10.17605/OSF.IO/8A3SC>
- Mokrane, M and Recker, J.** 2019. CoreTrustSeal-certified repositories: enabling findable, accessible, interoperable, and reusable (FAIR). In: *16th International Conference on Digital Preservation, iPRES*, DOI: <https://doi.org/10.17605/OSF.IO/9DA2X>
- Peng, G.** 2018. The State of Assessing Data Stewardship Maturity – An Overview. *Data Science Journal*, 17(7): 1–12. DOI: <https://doi.org/10.5334/dsj-2018-007>
- Perrier, L, Blondal, E and MacDonald, H.** 2020. The views, perspectives, and experiences of academic researchers with data sharing and reuse: A meta-synthesis. *PLoS One*, 15(2): e0229182. DOI: <https://doi.org/10.1371/journal.pone.0229182>
- Rickards, L, Vardigan, M, Dillo, I, Genova, F, L'Hours, H, Minster, JB, Edmunds, R and Mokrane, M.** 2016. DSA–WDS Partnership: Streamlining the landscape of data repository certification. In: *SciDataCon16*. <https://pure.knaw.nl/ws/portalfiles/portal/3998526/SciDataCon2016Dilloc.pdf> [Last Accessed 28 September 2020].
- Robinson-García, N, Jiménez-Contreras, E and Torres-Salinas, D.** 2016. Analyzing data citation practices using the data citation index. *Journal of the Association for Information Science and Technology*, 67(12): 2964–2975. DOI: <https://doi.org/10.1002/asi.23529>

- Schellenberg, TR.** 1956. *Modern Archives: Principles and Techniques*. Chicago: University of Chicago Press.
- Stall, S.** 2015. Good Data Can Be Better Data – How Data Management Maturity Can Help Repositories Improve Operations, Data Quality, And Usability, Helping Researchers. *American Geophysical Union, Fall Meeting 2015, IN14A-06*. <https://ui.adsabs.harvard.edu/abs/2015AGUFMIN14A..06S/abstract> [Last Accessed 28 September 2020].
- Tenopir, C, Christian, L, Allard, S and Borycz, J.** 2018. Research Data Sharing: Practices and Attitudes of Geophysicists. *Earth and Space Science*, 5(12): 891–902. DOI: <https://doi.org/10.1029/2018EA000461>
- The Center for Research Libraries and OCLC Online Computer Library Center.** 2007. *Trustworthy Repositories: Audit & Certification: Criteria and Checklist, Version 1.0*. Available at https://www.crl.edu/sites/default/files/d6/attachments/pages/trac_0.pdf [Last Accessed 28 September 2020].
- Wilkinson, MD, Dumontier, M, Aalbersberg, IJ, Appleton, G, Axton, M, Baak, A, Blomberg, N, Boiten, JW, da Silva Santos, LB, Bourne, PE and Bouwman, J.** 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*, 3(1): 1–9. DOI: <https://doi.org/10.1038/sdata.2016.18>
- Yoon, A.** 2014. End users' trust in data repositories: definition and influences on trust development. *Archival Science*, 14(1): 17–34. DOI: <https://doi.org/10.1007/s10502-013-9207-8>

TO CITE THIS ARTICLE:

Downs, RR. 2021. Improving Opportunities for New Value of Open Data: Assessing and Certifying Research Data Repositories. *Data Science Journal*, 20: 1, pp. 1–11. DOI: <https://doi.org/10.5334/dsj-2021-001>

Submitted: 31 March 2019
Accepted: 03 January 2021
Published: 21 January 2021

COPYRIGHT:

© 2021 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

Data Science Journal is a peer-reviewed open access journal published by Ubiquity Press.