

# Establishing Trustworthy Repositories of Scientific Data: Opportunities and Benefits

Robert R. Downs  
Center for International Earth Science  
Information Network (CIESIN),  
Columbia University  
61 Route 9W  
Palisades, NY 10964  
+1 (845) 365-8985  
rdowns@ciesin.columbia.edu

Ruth Duerr  
Ronin Institute for Independent  
Scholarship  
6637 W 95<sup>th</sup> PI  
Westminster, CO 80021  
+1 (303) 946-4842  
ruth.duerr@ronininstitute.org

Devan Ray Donaldson  
Department of Information & Library  
Science  
Indiana University  
1320 E. 10<sup>th</sup> St, Wells Library 019  
Bloomington, IN 47405  
+1 (812) 855-9723  
drdonald@indiana.edu

Sarah Ramdeen  
School of Information and Library  
Science  
University of North Carolina at Chapel  
Hill  
Chapel Hill, NC, 27599  
ramdeen@email.unc.edu

## ABSTRACT

Scientific progress often depends on the ability of the scientific community to build on the works of others. Such works include scientific data, published reports of findings, and other research-related information and artifacts that are produced as part of the scientific process. Providing capabilities for accessing and using such scientific works enables the reproducibility of published methods and results to identify opportunities for improvement. Access and use of science products also enables others to build on previous work. In an increasingly digital world, the science community accesses and uses relevant scientific resources that have been obtained from digital repositories, data centers, and archives, as well as from traditional sources such as publishers of journal articles. Digital repositories need to establish capabilities, which provide access to and enable the use of digital resources. These resources are needed by the science community to improve and build on the efforts of others. Digital repositories that provide free and unrestricted access to scientific data and research-related information can reduce the barriers to science. By reducing these barriers they will be able to offer opportunities for members of the scientific community to pursue research questions and challenges that were previously unapproachable. These may include opportunities for researchers to gather data from other domains and support interdisciplinary research. Opportunities to use the data products and services offered by digital repositories also can contribute to the development of the scientific community and to the emergence of new areas of study.

Being able to access scientific data and other research resources supports future science and is important to the research community. Given their role as stewards, digital repositories must be considered by this community to be trustworthy. With limited

resources available in many science domains, the scientific community as a whole cannot afford to lose science data and related resources. Digital resources may be particularly vulnerable to loss. Improving the infrastructure and practices for managing scientific data can reduce the potential for such losses. Trustworthy facilities are needed to curate, disseminate, and maintain these data and research-related materials. Furthermore, trustworthy repositories are needed to develop and improve data management services. They should also foster improvements in the capabilities and practices for scientific data stewardship.

Establishing digital repositories as trustworthy stewards of scientific data and related research products and services offers potential opportunities and benefits for science and society that can be leveraged to further research, educational, or decision making objectives. The direct beneficiaries of science data repositories include the individuals who serve as producers, stewards, and users of science data as well as the organizations that fund and host the digital repositories. Other potential beneficiaries include those who are not community stakeholders, currently, but may have an interest in these resources in the future. Furthermore, as described below, society at large also could benefit from digital repositories that have been recognized as trustworthy stewards of scientific data.

Data producers include scientists and other members of science project teams. Such data producers can compare digital repositories to determine those that are trustworthy, thereby enabling consumption of their data by interested researchers. If there is a variety of trustworthy repositories for their data, data producers can be selective and choose the repository that will best serve the community of potential users that the data producers are targeting for the use of their data products and services. Furthermore, selectivity among data producers for their choice repository could lead to competition among repositories that serve a particular science discipline, which may in turn lead to increased specialization of repositories to provide unique services.

Science data stewards include professionals in data management, information systems, and data services. These stewards can

iPres 2015 conference proceedings will be made available under a Creative Commons license.

With the exception of any logos, emblems, trademarks or other nominated third-party images/text, this work is available for re-use under a Creative Commons Attribution 3.0 unported license. Authorship of this work must be attributed. View a [copy of this licence](#).

compare choices among employment opportunities where they will apply their knowledge and skills, while also contributing to the curation, preservation, and dissemination of scientific data products and services. Tools should be available to enable data stewards to prepare, process, and preserve data for the future. These tools should also enable the dissemination of data products and services to diverse communities of users. Data stewards who accept positions at trustworthy science data repositories can enjoy the opportunities for professional development. These opportunities may be more abundant for individuals working within organizations that have been designated as trustworthy providers of scientific resources. Trustworthy digital repositories of science data and their staff should be able to demonstrate sustainable capabilities for managing data curation operations, for diligently preserving and disseminating science data, and for ensuring the integrity of their systems.

Users of science data products and services include researchers of all types, decision-makers, learners, and members of the general public. With improvements in the quality of products and services available from trustworthy digital repositories, these users can patronize trustworthy digital repositories that offer resources relevant to their goals and interests. Trustworthy repositories may recognize the potential for expanding their user base by supporting various levels of expertise, particularly in the science domain represented by the data that they archive and disseminate. As such, the availability of data products and services curated by trustworthy repositories for current and future users will enable use by increasingly diverse populations.

Benefactors of trustworthy digital repositories of science data will be able to observe and demonstrate how their support of trustworthy resources that have been prepared and disseminated, contributes to the future of science and the overall benefit of humanity. Funders that support the development and operation of trustworthy digital repositories could include government agencies, foundations, and philanthropists. The costs incurred enable the stewardship and dissemination of science data products and services. Perhaps, with such evidence of the value of their contributions, funders will be able to provide trustworthy digital repositories with the support needed to sustain science data stewardship operations and to help prepare science data repositories with the capabilities necessary to meet future challenges for the curation and dissemination of science data.

Organizations that host trustworthy digital repositories often include domain-specific data centers, archives, and institutional repositories. Upon attaining the status of trustworthy digital repositories, these organizations are likely to recruit the most qualified members of the science community based on their reputation for providing reliable data products and services. As more organizations become trustworthy digital repositories of science data, we also can expect the requirements of being trustworthy to increase. Just as other standards improve as technology evolves and new needs are recognized, the demands for increasing the requirements for trustworthy digital repositories, especially those that are responsible for science data, also should become more rigorous. This will ensure that science data and other research materials in digital form are being managed effectively for future use.

Members of society who are not traditional users of science data or other scientific resources also can benefit from the emergence of trustworthy science data repositories. Open science data that are accessible from trustworthy digital repositories will offer societal benefits as the data are used and translated into knowledge that contributes to the well-being of society at large. For example, educational institutions will be able to leverage the data and other research materials available to improve opportunities for educators and their students to learn from such resources. In addition, the benefits of trustworthy digital repositories for science can be realized by society as scientific breakthroughs, made possible by the continuing availability of science data products and services, thereby contributing to the lives of current and future generations as data are used to inform decision-making.

These are just a few of the opportunities and benefits that we can expect and hope for as digital repositories for science data attain the designation of being trustworthy. Taken together, the opportunities and benefits that can emerge from the availability of trustworthy digital repositories for science data can increase the maturity of the infrastructure and capabilities for managing, curating, disseminating, and preserving the digital data that exist today as well as those that will be produced in the future. Likewise, the availability of trustworthy science data repositories also has the potential to increase the professionalism of scientific data management practices, reducing the potential for the science data that have been created in digital form to be lost, through technological obsolescence, mismanagement, insufficient context for use, lapses in security, or other potential difficulties that could occur. Progress in the infrastructure for science can be achieved through the development and operation of trustworthy digital repositories for science data.

In summary, the availability of digital repositories that have been designated as trustworthy stewards for scientific data can contribute to the future availability of the science data that have been created during recent decades as well as the data that will be created in the future. Trustworthy digital repositories for science data can raise the expectations of stakeholder communities, increase the availability of choices for managing science data, improve scientific data stewardship practices, and contribute to the progress of science and the betterment of humanity.

## **General Terms**

Institutional opportunities and challenges; Infrastructure opportunities and challenges.

## **Keywords**

trustworthy digital repositories; scientific data centers; science data infrastructure; cyberinfrastructure; data archives; science data products; data services.

## **1. ACKNOWLEDGMENTS**

Support for Robert R. Downs has been received from the National Aeronautics and Space Administration (NASA) under Contract NNG13HQ04C for the continued operation of the Socioeconomic Data and Applications Center (SEDAC).