Purpose and Benefits of U.S. Geological Survey Trusted Digital Repositories

Federal mandates and U.S. Geological Survey (USGS, also known as the Bureau) Fundamental Science Practices (FSP) policies require that publicly funded scientific data, publications, and derivative works be openly accessible to researchers and the public. Open access helps to leverage the public investment by making the acquired data and published information products—collectively referred to as "data assets"—easier to locate, reproduce, and reuse. Open access also provides transparency to the processes used to acquire and analyze the data, thereby helping to ensure the scientific integrity of USGS data and products.

The data assets produced by USGS programs, science centers, and projects are preserved digitally in various USGS and non-USGS repositories (see, for example, fig. 1). To capitalize on the investment expended for data collection, analysis, and interpretation, these systems must remain useful and meaningful. For USGS repositories, the Bureau FSP Advisory Committee (FSPAC) has implemented an evaluation process to ensure that the systems being used to preserve these data assets are trustworthy, reliable, and secure. A system that is found to meet the reliability and suitability requirements is certified as a USGS Trusted Digital Repository (TDR).

Figure 1. Mass storage system for the USGS Earth Resources Observation and Science (EROS) Center data archive, a U.S. Geological Survey Trusted Digital Repository. Photograph by the U.S. Geological Survey.

Trusted Digital Repository status

Trusted Digital Repository (TDR) status implies that a digital repository has been internally certified in the USGS through a rigorous panel review of information provided by the repository manager. The evaluation criteria represent an internationally accepted measure for determining the suitability of the repository to acquire the status of "USGS Trusted Digital Repository." The evaluation process requires prospective repositories to demonstrate that USGS data are preserved, made accessible, documented, and backed up in a manner that reflects international standards for digital repositories.

What Is a USGS Trusted Digital Repository?

A USGS TDR is a digital system that has been certified by the Bureau as being one that meets international standards for digital repositories. Such repositories meet the Federal mandates for preserving and delivering scientific data to researchers and the public over a long period of time. A TDR adheres to community-established policies and practices that provide reliable methods to preserve, secure, and grant long-term, open access to preserved assets. The TDR designation indicates that the USGS repository was vetted according to a rigorous process used to evaluate the maturity and suitability of the data system to manage and preserve valuable USGS data assets appropriately.

In addition to preserving assets in the form of published products, TDRs preserve and provide open access to information and raw data obtained through such scientific activities as field observations and measurements, data analysis and modeling, and laboratory processes, and they help simplify data integration across suites of data products. Storing these assets in TDRs supports the trustworthiness of the data by improving data availability, enhancing scientific reproducibility, and providing evidence for data legitimacy in follow-on or derivative scientific applications.

USGS TDR standards recommend common practices for data management, archiving, and dissemination. For USGS digital assets (including systems, tools, workflows, and data), the TDR standards promote application of the Findable, Accessible, Interoperable, and Reusable (FAIR) principles (Wilkinson and others, 2016). The FAIR principles focus on the registration of persistent identifiers; the use of robust metadata; version management; and communication with users regarding changes to data or metadata, usage licenses, and links to related research products.

Basis for USGS Trusted Digital Repositories

In 2016, the USGS developed a plan to provide public access to its federally funded scientific research, titled "Public Access to Results of Federally Funded Research at the U.S. Geological Survey: Scholarly Publications and Digital Data" (fig. 2; U.S. Geological Survey, 2016). Known as the USGS Public Access Plan, this plan explains the requirements and processes to make scientific data and publications publicly usable, easily discoverable, and widely available. The concepts of open and accessible data embodied in the plan require repositories that (1) provide reliable and secure digital infrastructure to manage, preserve, and provide open access to stored assets; (2) are subject to sustainable data management practices to ensure that the data are properly curated beyond the duration of the projects in which they were acquired; and (3) serve the USGS, its researchers, and the public by making the data assets available in stable and reliable digital environments in the long term.
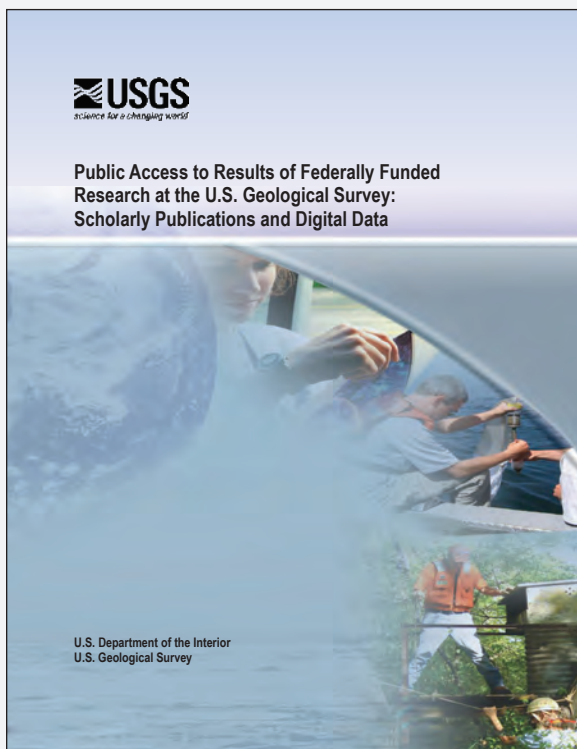
Figure 2: Cover of the U.S. Geological Survey plan for providing public access to its federally funded scientific research.

The USGS is also required to make integration of data stored by science theme experts in separate thematic databases possible. This capability is necessary to meet the USGS goal of providing integrated science, across various disciplines, to study the Earth's critical zone and Earth systems (Gallagher, 2012). By promoting common practices, USGS TDRs assist with this integration, especially by helping to overcome the challenges inherent in data discovery and reuse. Common practices make coordination of data among robust and reliable systems easier.

The USGS Public Access Plan was written in response to multiple Federal directives, including (1) the "Open, Public, Electronic, and Necessary Government Data Act," or the OPEN Government Data Act (44 USC 101, P.L.115–435 of January 14, 2019); (2) Office of Science and Technology Policy memorandum, "Increasing Access to the Results of Federally Funded Scientific Research" (Holdren, 2013); (3) Executive Order 13642, "Making Open and Machine Readable the New Default for Government Information" (Obama, 2013); and (4) Office of Management and Budget Memorandum M–13–13, "Open Data Policy—Managing Information as an Asset" (Burwell and others, 2013).

USGS Trusted Digital Repository owners and managers must—

- Have a mission to preserve USGS science data;
- Accept responsibility for long-term maintenance of digital resources on behalf of depositors and for the benefit of users;
- Support long-term viability of the repository and the digital information it contains;
- Demonstrate fiscal and managerial viability and sustainability;
- Design the repository in accordance with commonly accepted system conventions and standards to ensure ongoing management, access, and security of data assets;
- Provide system evaluation that meets community expectations of trustworthiness;
- Maintain communication regarding changes to data or metadata under their care; and
- Manage metadata for all datasets in compliance with the National Archives and Records Administration, the U.S. Department of the Interior, and USGS policies.

Examples of USGS Trusted Digital Repositories

USGS TDRs are used for the storage of science data produced primarily by the USGS and its cooperators and collaborators. The following are some of the repositories that have been certified as USGS TDRs. The data assets contained in these TDRs are publicly available.

ScienceBase

ScienceBase (https://www.sciencebase.gov) is a data management platform that provides centralized open access to preserved USGS science and products. ScienceBase was created to further the USGS goal of developing an integrated information environment that links digital assets and provides digital tools for researchers to use to establish connections among disparate sources of science information.

USGS Alaska Science Center (ASC) Data Repository

The USGS Alaska Science Center (ASC) Data Repository (https://www.usgs.gov/centers/asc/data-tools) provides public access to digital data generated by the USGS ASC. The repository includes materials related to scientific research of climate, ecosystems, environmental health, natural hazards, and natural resources (energy, minerals, and water).

USGS Earth Resources Observation and Science (EROS) Center Data Archive

The USGS Earth Resources Observation and Science (EROS) Center data archive (https://www.usgs.gov/centers/eros) contains one of the largest civilian collections of remotely sensed images of the Earth's land surface, focusing on land change, for use by researchers, resource managers, policymakers, and public users. The collection includes satellite imagery, aerial photography, elevation data, and land cover datasets (fig. 3).

Figure 3. This satellite image of Lake Erie during a harmful algal bloom event in September 2017 is preserved in the Earth Resources Observation and Science (EROS) Center data archive. Source: U.S. Geological Survey and the National Aeronautics and Space Administration.

National Geologic Map Database (NGMDB)

The National Geologic Map Database (NGMDB) (https://ngmdb.usgs.gov/) is the national digital archive of geologic maps and related geoscience reports and databases. The material in the NGMDB conforms to standardized formats to enable data assimilation, manipulation, and integration with comparative data for the purpose of land use management, assessment and use of natural resources, and environmental conservation and protection. Under the authority of the National Geologic Mapping Act of 1992, the USGS, in cooperation with the Association of American State Geologists, is responsible for establishing and maintaining the NGMDB.

Why Seek Certification?

The USGS FSP policies require USGS publications and data products approved for release to be preserved in digital repositories that meet USGS Public Access Plan requirements (U.S. Geological Survey, 2016). USGS digital repository owners or asset managers are advised to obtain TDR status to demonstrate the viability and capacity of the system or facility supporting the system infrastructure to meet long-term preservation requirements. USGS TDR status conveys that the data system is reliable and demonstrates the commitment of the USGS to meet Federal requirements to deliver scientific data and publication products to the public over the long term (U.S. Geological Survey, 2016).

Although USGS data may be preserved in accepted non-USGS repositories—for example, in non-USGS aggregated databases that contribute to the integration of worldwide research—authoritative digital copies of the data, including versioned data, must also be maintained in USGS systems. Some accepted non-USGS digital repositories include Incorporated Research Institutions for Seismology (IRIS), which is a database for seismic and other geophysical data; GenBank, which is a genetic sequence database; and AmeriFlux, which is a database for recorded field measurements of carbon, water, and energy fluxes used to study terrestrial ecosystems.

How does a USGS repository become a Trusted Digital Repository?

1. Repository staff alert the Trusted Digital Repository (TDR) working group of their interest in obtaining TDR certification.

2. Repository staff complete the TDR certification self-assessment available at https://www.usgs.gov/atom/84300.

3. Repository applicant and TDR working group staff work together to reconcile potential deficiencies to improve the application for certification.

4. Once TDR status is obtained, system capabilities continue to be developed to promote preservation and reuse of the scientific data assets.

5. Repository staff commit to recertifying the TDR through the USGS procedure or the international CoreTrustSeal process every 3 years.

Certification Process

To begin the process of submitting a repository for certification, USGS product owners or system managers (the "organization") alert the TDR working group of the organization's interest in applying for USGS TDR certification through the contact listed on the USGS TDR website (https://www.usgs.gov/about/organization/science-support/office-science-quality-and-integrity/trusted-digital-repository). Using the USGS Trustworthy Data Repository Application form that is downloadable from the USGS TDR website, the organization provides a self-assessment, graded on a scale of 0 (low) to 4 (high), for criteria associated with the organization's (1) infrastructure, including governance, financial sustainability, and staffing; (2) management of digital objects, including data quality and integrity, use of persistent identifiers and citations, and documented processes; and (3) technology, including technical infrastructure and security and risk management approaches. Requirements that are not applicable to the organization or for which the organization meets only the minimal qualifications must be documented.

Members of the USGS TDR working group then review the application, using an evaluation process that is both iterative and collaborative between the applicant and working group.

Deficiencies noted on the submission are discussed with the submitter to determine if the element was fully understood, and the working group may request additional information, if needed. Element scores of 2 or below assigned by the reviewers would not allow for USGS certification; however, the USGS TDR working group continues to work with the submitter to develop a pathway for the repository to attain USGS certification. Areas in which the applicant has lower graded compliance levels may be outweighed by highly graded criteria, especially during the implementation phase. During the implementation phase, the system processes are refined for continuous evolution to meet certification requirements in the long term.

USGS repositories are reassessed every 3 years to document improvements to and changes in systems and processes. The 3-year cycle promotes currency of information shared about USGS TDRs with users and stakeholders and allows the USGS to evolve its requirements for TDR evaluation, paralleling applied conventions and methods employed by internationally accepted communities of practice, particularly the international CoreTrustSeal certification process.

The international CoreTrustSeal certification approach (CoreTrustSeal, 2019a) embodies 16 criteria for evaluating the trustworthiness of data management and stewardship activities and infrastructure (https://www.coretrustseal.org/why-certification/ requirements/). Eventually, the intent is for USGS TDR applications to be submitted directly to the international CoreTrustSeal organization for CoreTrustSeal Data Repository certification (CoreTrustSeal, 2019b).
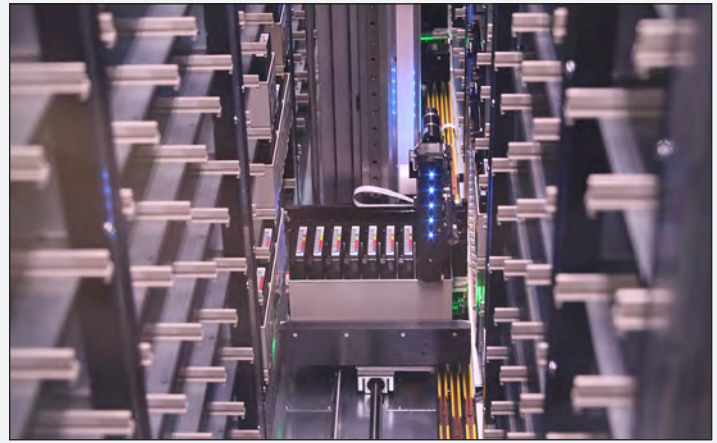
Summary

Those who manage USGS data repositories are held accountable for the care and preservation of valuable assets acquired using public funds (fig. 4). Some of these data, such as those collected for investigations of natural hazards, endangered species, aerial surveys, and other resource-intensive efforts, may never be recollected or duplicated because of changing conditions and high costs. TDR self-assessments and collegial reviews convey assurance that the data system is robust and capable of providing security and longevity to the assets it contains. The USGS has embraced Federal mandates to provide public access to its scientific research products by developing clear and effective FSP and policies that dictate reliable data management processes and robust systems capable of performing TDR functions now and into the future. When coupled with disciplined data management practices, USGS TDRs preserve and deliver scientific data to the public and promote the mission of the USGS to provide reliable scientific information to describe and understand natural systems to the benefit of society.

References Cited

Burwell, S.M., Mancini, D.J., Park, T., and VanRoekel, S., 2013, Open data policy—Managing information as an asset: Executive Office of the President, Office of Management and Budget Memorandum M–13–13, May 9, 2013, 12 p., accessed May 17, 2019, at https://obamawhitehouse.archives.gov/sites/default/files/omb/memoranda/2013/m-13-13.pdf.

CoreTrustSeal, 2019a, Home page: CoreTrustSeal, accessed May 17, 2019, at https://www.coretrustseal.org.

Figure 4. Interior of one of the U.S. Geological Survey's high-speed, high-capacity tape storage libraries. To retrieve information, a robotic transporter moves storage tape cartridges around within the library, obtaining the requested tape, inserting the requested tape into the tape drive to be read, and then reshelving the cartridge when the tape is no longer in use. Photograph by the U.S. Geological Survey.

CoreTrustSeal, 2019b, Webinars on CoreTrustSeal certification: CoreTrustSeal, accessed June 5, 2019, at https://www.coretrustseal.org/why-certification/requirements/.

Gallagher, K.T., 2012, Core Science Systems—Mission overview: U.S. Geological Survey Fact Sheet 2012–3009, 4 p., accessed May 17, 2019, at https://doi.org/10.3133/fs20123009.

Holdren, J.P., 2013, Increasing access to the results of federally funded scientific research: Executive Office of the President, Office of Science and Technology Policy Memorandum for the Heads of Executive Departments and Agencies, February 22, 2013, 6 p., accessed May 17, 2019, at https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf.

Obama, B., 2013, 2013, Making open and machine readable the new default for Government information—Executive Order No. 13642 of May 9, 2013: Federal Register, v. 48, no. 93, p. 28111–28113, accessed May 17, 2019, at https://www.federalregister.gov/documents/2013/05/14/2013-11533/making-open-and-machine-readable-the-new-default-for-government-information.

U.S. Geological Survey, 2016, Public access to results of federally funded research at the U.S. Geological Survey: Scholarly Publications and Digital Data, 18 p. plus 2 appendixes, accessed February 24, 2020, at https://www.usgs.gov/atom/72352.

Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., and others, 2016, The FAIR guiding principles for scientific data management and stewardship: Scientific Data, v. 3, no. 1, p. 160018, accessed May 17, 2019, at https://doi.org/10.1038/sdata.2016.18. (An addendum to this article was published on March 19, 2019, and is available at https://www.nature.com/articles/s41597-019-0009-6.)

By Natalie Latysh, Keith G. Kirk, and John Faundeen

For more information, contact:
Office of Science Quality and Integrity
U.S. Geological Survey, 12201 Sunrise Valley Drive, Reston, VA 20192
703–648–6774