USGS
science for a changing world

Opportunities To Improve Alignment With the FAIR Principles for U.S. Geological Survey Data

Open-File Report 2022–1043

U.S. Department of the Interior
U.S. Geological Survey

Opportunities To Improve Alignment With the FAIR Principles for U.S. Geological Survey Data

By Frances L. Lightsom, Vivian B. Hutchison, Bradley Bishop, Linda M. Debrewer, David L. Govoni, Natalie Latysh, and Shelley Stall

Open-File Report 2022–1043

U.S. Department of the Interior
U.S. Geological Survey

U.S. Geological Survey, Reston, Virginia: 2022

Suggested citation:
Lightsom, F.L., Hutchison, V.B., Bishop, B., Debrewer, L.M., Govoni, D.L., Latysh, N., and Stall, S., 2022, Opportunities to improve alignment with the FAIR Principles for U.S. Geological Survey data: U.S. Geological Survey Open-File Report 2022–1043, 23 p., https://doi.org/10.3133/ofr20221043.

ISSN 2331-1258 (online)

Acknowledgments

The authors are indebted to the generous discussions, insights, and experience offered by the 2019 U.S. Geological Survey FAIR Workshop participants, who are listed in the appendix of this report. The U.S. Geological Survey Community for Data Integration provided funding and encouragement for this roadmap project.

Contents

Figures

Tables

Abbreviations

BASIS+	Budget and Science Information System Plus
CDI	Community for Data Integration
EarthMAP	Earth Monitoring, Analyses, and Projections
FAIR	findable, accessible, interoperable, and reusable
FSP	Fundamental Science Practices
IPDS	Information Product Data System
ISO	International Organization for Standardization
OGC	Open Geospatial Consortium
ORCID	Open Researcher and Contributor Identifier
TDR	trusted digital repository
URL	uniform resource locator
USGS	U.S. Geological Survey

Opportunities To Improve Alignment With the FAIR Principles for U.S. Geological Survey Data

By Frances L. Lightsom,[1] Vivian B. Hutchison,[1] Bradley Bishop,[2] Linda M. Debrewer,[1] David L. Govoni,[1] Natalie Latysh,[1] and Shelley Stall[3]

Executive Summary

In 2016, an interdisciplinary, international group of 53 scientists introduced a framework named "the FAIR Principles" for addressing 21st century scientific data challenges (Wilkinson and others, 2016). The FAIR Principles are increasingly used as a guide for producing digital scientific products that are findable, accessible, interoperable, and reusable (FAIR), especially to enable use of such products in automated systems. Data aligned with the FAIR Principles can increase the efficiency of science integration capabilities such as those envisioned for the U.S. Geological Survey (USGS) Earth Monitoring, Analyses, and Projections (EarthMAP) initiative (Jenni and others, 2017).

The FAIR Principles clearly define the characteristics of reusable scientific products, but it is less clear how to facilitate consistency in achieving these characteristics across the Bureau. USGS data are produced by local research projects distributed over more than 100 centers in 7 regions. After data are approved for release, they could be managed in numerous repositories and online data systems. The diversity of USGS data is illustrated by the topical range of the USGS mission areas: Core Science Systems, Ecosystems, Energy and Minerals, Natural Hazards, and Water Resources. In the USGS context, realizing the EarthMAP vision for automated, predictive, integrated science that provides timely and actionable results involves providing knowledge and support services and developing the skills, infrastructure, and culture to enable Bureau-wide implementation of the FAIR Principles.

In 2019, the USGS Community for Data Integration funded a project to convene a broadly representative workshop and produce recommendations to enable consistency with the FAIR Principles across the USGS. The workshop, held in Fort Collins, Colorado, in September 2019, brought together 28 participants for 3 days to engage with the FAIR Principles, analyze USGS use cases, and discuss the roles of data producers and managers, data storage and catalogs, value-added services, and policy makers in implementing the FAIR Principles. Workshop participants agreed that scientific reproducibility requires the extension of the FAIR Principles beyond measured data to include physical samples, research methods, software, and tools at the USGS. Workshop discussions focused on how the USGS can implement the FAIR Principles by supporting research teams in creating data, metadata, and other scientific products and also by supporting enterprise systems that maintain and leverage the products' consistency with the FAIR Principles.

The resulting FAIR roadmap of recommendations describes nine proposed interdependent strategies that could be achieved by coordinated actions taken by different parts of the USGS. A proposed early action would be the creation of a coordinating council that includes representatives from the groups engaged in activities consistent with better alignment with the FAIR Principles. The nine proposed strategies, which are presented in more detail in this roadmap report, focus on enabling improvements to individual data products, providing infrastructure, and structuring administrative activities to support an organizational culture that values the FAIR Principles.

Data Strategies

- Machine-actionable metadata improve data discovery and reuse.
- Research methods generate interoperable and reusable data.
- Persistent identifiers link research products for automated discovery and access.

Infrastructure Strategies

- Repositories and data systems provide discovery and access through unified, compatible methods.
- Vocabulary services improve discovery and evaluation of fitness for use.
- Policy requirements improve consistency of products.

Administrative Strategies

- Continued culture shift, achieved through knowledge exchange, encourages activities and techniques that improve alignment with the FAIR Principles.
- Employee incentives improve the production of FAIR-aligned data.
- Research, development, and participation in professional communities improve capabilities to align with the FAIR Principles.

[1]U.S. Geological Survey.

[2]University of Tennessee.

[3]American Geophysical Union.

References Cited

Jenni, K.E., Goldhaber, M.B., Betancourt, J.L., Baron, J.S., Bristol, R.S., Cantrill, M., Exter, P.E., Focazio, M.J., Haines, J.W., Hay, L.E., Hsu, L., Labson, V.F., Lafferty, K.D., Ludwig, K.A., Milly, P.C., Morelli, T.L., Morman, S.A., Nassar, N.T., Newman, T.R., Ostroff, A.C., Read, J.S., Reed, S.C., Shapiro, C.D., Smith, R.A., Sanford, W.E., Sohl, T.L., Stets, E.G., Terando, A.J., Tillitt, D.E., Tischler, M.A., Toccalino, P.L., Wald, D.J., Waldrop, M.P., Wein, A., Weltzin, J.F., and Zimmerman, C.E., 2017, Grand challenges for integrated U.S. Geological Survey science—A workshop report: U.S. Geological Survey Open-File Report 2017–1076, 94 p., accessed June 29, 2021, at https://doi.org/10.3133/ofr20171076.

Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., Santos, L.B. da S., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A.J.G., Groth, P., Goble, C., Grethe, J.S., Heringa, J., Hoen, P.A.C. 't, Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M.A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., and Mons, B., 2016, The FAIR guiding principles for scientific data management and stewardship: Scientific Data, v. 3, no. 1, [article 160018], 9 p., accessed June 29, 2021, at https://doi.org/10.1038/sdata.2016.18.

Introduction

U.S. Geological Survey (USGS) data that align with the FAIR Principles are foundational to achieving the USGS mission in the 21st century. The USGS mission is to monitor, analyze, and predict the current and evolving dynamics of complex human and natural Earth system interactions and deliver actionable information at scales and timeframes relevant to decision makers. To do this in the 21st century, integration of multidisciplinary natural science data, research results, geospatial information systems, and predictive models will be necessary. This report introduces a roadmap composed of proposed actions to be taken across the USGS to establish this essential basis for success.

The FAIR Principles are international criteria that promote more effective data-intensive science by enabling automated processes to find, access, interoperate, and reuse data more easily. As such, the FAIR Principles provide the USGS with a framework that supports an "integrated and predictive capability that accounts for complex natural system interactions; anticipates the likelihood and consequences of evolving threats and hazards; and helps guide resilient adaptation and mitigation efforts" (James Reilly, former Director, U.S. Geological Survey, written commun., April 1, 2019). Better alignment with the FAIR Principles can benefit the USGS mission and stakeholders in the following ways:

- Ensuring that USGS scientific data and products are more easily findable by stakeholders through standard search engines, USGS-provided search tools, and broad U.S. Government discovery platforms utilizing keywords, tags, filter criteria, and facets that are meaningful to the searcher, relevant to the search criteria, and machine actionable.
- Providing scientific products that are more easily accessible to stakeholders who use convenient and open platforms for viewing, acquiring, and interacting with data and software.
- Producing more easily interoperable data to improve integration of different types of datasets, on various platforms, by using standard formats and providing robust metadata.
- Enabling data to be more reusable for new scientific analysis and decision making by including comprehensive metadata that support determination of whether the data are fit for the purpose intended and that provide clear information on licenses.

The USGS has made substantial strides to publicly release the scientific data that are used in interpretive scholarly publications; however, continued efforts can help ensure that digital assets across the Bureau have maximum influence on the scientific research community and maximum utility in addressing complex societal problems. Adoption of the FAIR Principles by the USGS as a whole can provide efficiencies that benefit researchers who share and integrate data to advance scientific interpretations, software developers and modelers who build tools and implement workflows for synthesis of large datasets, Federal agencies charged with making sound, science-based decisions to manage and conserve natural resources, and the American public who wish to access and use USGS scientific information.

The roadmap introduced in this report resulted from the 2019 USGS FAIR Workshop, which gathered 28 participants from across the USGS and from other organizations, including USGS data stakeholders, data professionals, and managers of USGS data systems (see participant list in appendix 1). The 3-day workshop was designed to produce recommendations for actions the USGS can take to increase alignment of Bureau data to the FAIR Principles. To this end, several invited workshop participants, including both scientists and data experts, gave short presentations describing use cases. Each use case demonstrated strengths in USGS data generation and management procedures, as well as deficiencies that inhibit those data from full alignment with the FAIR Principles. Breakout sessions allowed small groups of workshop participants to

discuss each use case and develop observations and recommendations for a USGS roadmap to enable better alignment with the FAIR Principles for USGS data. The resulting recommendations offer many opportunities to increase consistency in policies and practices guiding the creation and management of USGS data.

This roadmap report examines the extent to which current USGS policies and practices align with the FAIR Principles and identifies actions to improve the efficiency and effectiveness of research, as well as usefulness of USGS data, through more comprehensive adoption of the FAIR Principles. This report provides background and makes connections among 88 discrete proposed actions, which could be taken by groups that include Bureau and science center management, the Core Science Systems Mission Area, the Fundamental Science Practices Advisory Council, the Community for Data Integration, the Human Capital Office and Office of Employee Development, and a proposed new coordination and communication council to oversee the process of enabling better alignment with the FAIR Principles in the USGS. Fully realized, this roadmap would provide USGS personnel delivering research data, data systems, and repositories with the understanding, information, tools, and support they need to produce and maintain FAIR-aligned digital resources.

Background: The FAIR Principles

The acronym "FAIR" was introduced in 2016 to the international science community in "The FAIR Guiding Principles for Scientific Data Management and Stewardship" (Wilkinson and others, 2016). The goal of the FAIR Principles is to communicate criteria for well-managed and machine-actionable data and metadata, in a time when the volume of data continues to increase, the ability to discover and access data is inconsistent, and the frequency of data reuse is poor. For each of the FAIR foundational elements (findable, accessible, interoperable, and reusable), Wilkinson and his coauthors provide specific guiding principles for data and metadata (see box 1), many of which have been previously highlighted in efforts to encourage organizations and researchers to value datasets and software as important contributions to the scientific record (see summary in Starr and others, 2015). In the Federal Government, such efforts included the Office of Science and Technology Policy memorandum "Increasing Access to the Results of Federally Funded Scientific Research" (Holdren, 2013) and the Foundations for Evidence-Based Policymaking Act of 2018 (Public Law 115–435, 132 Stat. 5529). In the USGS, a need for improved data management to support new research efforts was expressed by a 2017 workshop entitled "Grand Challenges for Integrated U.S. Geological Survey Science" (Jenni and others, 2017).

Acceptance of the FAIR Principles within the scientific community is reflected in the 2019 revision of the data position statement of the American Geophysical Union (American Geophysical Union, undated). The revised statement affirms that

> Earth and space science data are a world heritage, and an essential part of the science ecosystem. All players in the science ecosystem—researchers, repositories, publishers, funders, institutions—should work to ensure that relevant scientific evidence is processed, shared, and used ethically, and is available, preserved, documented, and fairly credited.

The FAIR Principles provide a clear goal for institutions, funders, societies, publishers, and researchers who see the need to improve their practices and policies for better data management, including meaningful data-management plans, tools and platforms that support researchers, and repositories that support data preservation and reuse.

The international scientific community has embraced the tenets of FAIR-aligned data as means to focus efforts across domains and geography. The FAIR Principles help the community address the challenge of enabling scientists to test and learn from published scientific results by reproducing research methods (National Academies of Sciences, Engineering, and Medicine, 2019). The FAIR Principles help governments address the need for transparency of scientific findings that are used in making policy decisions (for example, the Foundations for Evidence-Based Policymaking Act of 2018; Public Law 115–435, 132 Stat. 5529). The FAIR Principles are also bringing awareness to the substantial amount of time researchers spend acquiring, preparing, and integrating data prior to, as well as after, the initial data collection and analysis (Directorate-General for Research and Innovation [European Commission], 2019). The FAIR-aligned data-management practices may initially appear to some as unrecognized and unrewarded efforts added to the research workload. A continued USGS culture shift to embrace the FAIR Principles could enable researchers to receive recognition for this work through data citations and also demonstrate that time dedicated to data management is paid forward when that same scientist, or another researcher, finds it easier and faster to use those well-prepared data for new research (Tenopir and others, 2018).

Though the FAIR Principles do not require data to be freely available and open to all, the international research community encourages broad data sharing as part of the research culture (for example, see Landi and others, 2020). Data and software that are both open and FAIR-aligned can optimize data utility and research efficiency. But if data must be protected (not freely available), those data can still be consistent with the FAIR Principles.

Innovations to enable FAIR research products are being implemented internationally and can provide models and tools for use in the USGS.

Box 1. The FAIR Guiding Principles

(Modified from Wilkinson and others, 2016.)

To be findable:

F1. Data and metadata are assigned a globally unique and persistent identifier.
F2. Data are described with rich metadata (defined by R1 below).
F3. Metadata clearly and explicitly include the identifier of the data they describe.
F4. Data and metadata are registered or indexed in a searchable resource.

To be accessible:

A1. Data and metadata are retrievable by their identifier using a standardized communications protocol.
  A1.1. The protocol is open, free, and universally implementable.
  A1.2. The protocol allows for an authentication and authorization procedure, where necessary.
A2. Metadata are accessible, even when the data are no longer available.

To be interoperable:

I1. Data and metadata use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2. Data and metadata use vocabularies that follow the FAIR Principles.
I3. Data and metadata include qualified references to other data and metadata.

To be reusable:

R1. Data and metadata are richly described with a plurality of accurate and relevant attributes.
  R1.1. Data and metadata are released with a clear and accessible data usage license.
  R1.2. Data and metadata are associated with detailed provenance.
  R1.3.Data and metadata meet domain-relevant community standards.

- The Coalition for Publishing Data in the Earth and Space Sciences (COPDESS; https://copdess.org) has a project for enabling FAIR data that includes guidelines, goals, resources, and a statement of commitment. The USGS is one of more than 100 organizations that are signatories of the coalition's commitment (Coalition for Publishing Data in the Earth and Space Sciences, 2018).
- As part of its Horizon2020 effort, the European Commission published an expert report: "Turning FAIR into Reality: Final Report and Action Plan from the European Commission Expert Group on FAIR Data" (European Commission, 2018). In that report, many of the world's leading experts on the FAIR Principles provide a guide for use by all countries and scientific domains to initiate the changes necessary to incorporate the FAIR tenets and change research culture.
- The USGS is a member of the Open Geospatial Consortium (OGC), an open standards body for geospatial data services. The OGC mission statement is "Make location information Findable, Accessible, Interoperable, and Reusable (FAIR)" (Open Geospatial Consortium, undated).
- The Australian Research Data Commons requires application of the FAIR Principles and provides guidelines to assist researchers in producing FAIR-aligned data (Australian Research Data Commons, 2021).
- Some of the authors of the initial FAIR guiding principles paper (Wilkinson and others, 2016) received funding from European governments and organizations to create GO FAIR (https://www.go-fair.org), a community of practice that provides training, cross-domain sharing, and progress metrics.

"Since its establishment in 1879, the USGS has had a firm commitment to providing public access to scientific results through timely, technically sound, peer-reviewed, and professionally presented scholarly publications and data generated from unclassified research funded wholly or in part by the USGS. The USGS adheres to rigorous policies, standards and processes for the development, review, approval and release of USGS scholarly publications, whether they are published internally in one of several USGS series publications or externally in scientific journals or other outlets." (U.S. Geological Survey, 2016, p. 1).

Background: Data Management at the U.S. Geological Survey

The USGS has historical commitments to maintaining scientific integrity and providing high-quality scientific products that serve the Nation and the world. The USGS conducts research across a wide range of scientific disciplines, which historically resulted in organizational units producing and maintaining their data products in ways consistent with the requirements and standards of their respective scientific discipline and stakeholder communities. Presently, USGS data are produced in the course of scientific activities by multiple projects within each of five mission areas: Core Science Systems, Ecosystems, Energy and Minerals, Natural Hazards, and Water Resources (see https://www.usgs.gov/science/mission-areas). Scientists and technicians within individual projects are responsible for the observations and measurements that result in data; data management, processing, analysis, and quality control; and release of the project's data to the public. The USGS does not have a central data-management facility, but 21st century trends and Federal mandates toward open science and interdisciplinary research have moved the USGS to introduce Bureau-wide processes for data management and release, data infrastructure, and research administration.

Throughout its history, the USGS has delivered its science primarily through published research papers and has managed and released data since 1889, with the first streamgage measurement on the Rio Grande River in New Mexico (Gunn and others, 2014), which eventually led to establishment of the USGS National Water Information System (NWIS).

In 2007, the USGS released a strategic science plan that included a new vision: "The USGS will use its information resources to create a more integrated and accessible environment for its vast resources of past and future data. It will invest in cyberinfrastructure, nurture and cultivate programs in natural-science informatics, and participate in efforts to build a global integrated science and computing platform" (U.S. Geological Survey, 2007, p. 56).

To help realize this vision, in 2010, USGS employees self-organized to establish the USGS Community for Data Integration (CDI; https://www.usgs.gov/centers/cdi), a community of practice that includes members from all the USGS mission areas and receives administrative and leadership support from the USGS Executive Leadership Team. The CDI is a forum for discussing new ideas, challenges, and solutions that support data integration and also for encouraging development of strategies, methods, tools, data, and infrastructure that support integrated science capability in the USGS. The community uses a rigorous proposal process to offer funding for projects focused on data-integration methods, tools and services. Through these activities, the CDI has encouraged broadly representative teams that have developed enterprise-level products such as the USGS data-management website (U.S. Geological Survey, 2021) and the USGS Science Data Lifecycle Model (Faundeen and others, 2013). The CDI also funded the workshop that inspired this report.

The USGS-wide discussions that produced the USGS Science Data Lifecycle Model, the resulting publication (Faundeen and others, 2013), and use of the model in CDI activities were important steps toward the development of a Bureau-wide culture of data management. The primary Science Data Lifecycle Model elements, which parallel the stages of data use in a scientific research project, are to plan, acquire, process, analyze, preserve, and publish and share (fig. 1). Additionally, the Science Data Lifecycle Model includes three elements that describe activities that are performed throughout the lifecycle: describe (metadata, documentation), manage quality, and backup and secure. The Science Data Lifecycle Model provides a scheme for organizing the USGS tools, workflows, policy, and guidance needed to implement scientific data management; it has been incorporated into multiple USGS tools, information products, and policies, including Survey Manual Chapter SM 502.6 (U.S. Geological Survey, undated d), the USGS Data Management website (U.S. Geological Survey, 2021), and the CDI Science Support Framework (Chang and others, 2015). The Science Data Lifecycle Model identifies the roles of people who create and manage data during, and beyond, the research lifecycle. To enable the USGS to consistently produce and maintain FAIR-aligned data, this roadmap addresses the whole science data lifecycle. In addition, Bureau-wide use of the USGS Science Data Lifecycle Model is a step toward the consistency that is the foundation for enabling better alignment with the FAIR Principles in the USGS.

In response to the Federal Data Quality Act of 2001 (Public Law 106–554, app. C, §515, 114 Stat. 2763A–153), the USGS began updating its policies and procedures for ensuring the quality and integrity of its research and monitoring activities and the resulting products. In 2006, this resulted in the introduction of the USGS Fundamental Science Practices (FSP; see U.S. Geological Survey, undated a). USGS Survey Manual chapters establish longstanding Bureau-wide policies, standards, instructions, and general procedures. Chapters of the Survey Manual continue to be

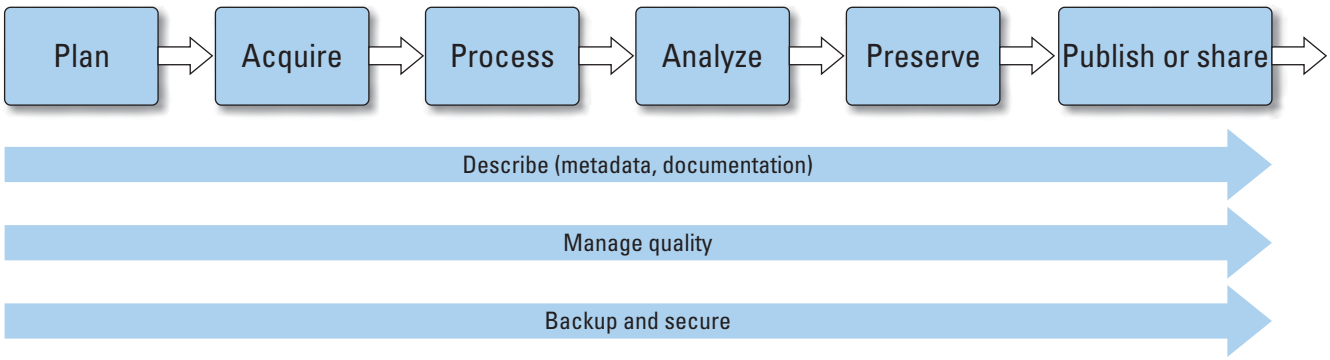

Figure 1. A diagram of the U.S. Geological Survey Science Data Lifecycle Model. Some data-management elements are parallel to stages in the research process, and other elements are addressed throughout research projects. Arrows at the right end of the diagram are a reminder that data maintenance and reuse continue after the end of research activity that creates the data. Diagram from Faundeen and others (2013).

revised to implement FSP (see U.S. Geological Survey, undated d). In 2016, the USGS published four FSP chapters focused on data management. This was coordinated with the completion of the USGS public access plan (U.S. Geological Survey, 2016), which was developed in response to the 2013 Federal Government open data and machine-readable data policies (Burwell and others, 2013; Holdren, 2013; Obama, 2013) that later culminated in the Foundations for Evidence-Based Policymaking Act of 2018 (Public Law 115–435, 132 Stat. 5529).

The public access plan accelerated USGS progress toward widely available, open scientific data. Implementation of the plan produced a substantial cultural shift in the management of scientific data across the Bureau. The new data policies introduced Bureau-wide adoption of modern practices of scientific data management: project data-management plans; public release of data in machine-readable, nonproprietary formats; provision of robust, granular metadata records in standard, machine-readable formats; use of USGS trusted digital repositories to maintain and share data; use of digital object identifiers to ensure long-term discoverability; and establishment of a comprehensive public data catalog. The USGS FAIR roadmap proposes extensions of these efforts to establish a foundation for data-intensive integrated science.

Current U.S. Geological Survey Practices Relative to the FAIR Principles

Development of a USGS FAIR roadmap required an assessment of the extent to which current policy and practice aligns with the FAIR Principles. At the 2019 FAIR Workshop, participants discussed a variety of use cases, spanning mission areas and data types, to identify current successes in producing FAIR-aligned scientific products, common challenges that prevent USGS products from being fully consistent with the FAIR Principles, and potential improvements in USGS infrastructure and administration that could enable the production of more fully FAIR-aligned products. This evaluation of the current status inspired 88 recommendations for activities that form this roadmap.

The USGS already has many successes in aligning USGS data with the FAIR Principles, reflecting current USGS policies, procedures, and guidelines for open scientific data and data quality. Some examples of ways the USGS provides data and information products that already align with the FAIR Principles are the following:

- Assignment of globally unique and persistent identifiers to publications and publicly available scientific data.
- Requirement that released data be described by detailed metadata records following the content standards endorsed by the Federal Geographic Data Committee.
- Mandates that metadata records contain the digital object identifier (DOI) for the data and are registered and indexed in the USGS Science Data Catalog (U.S. Geological Survey, undated b) for further exposure in downstream discovery platforms.
- Issuance of policies and tools for documenting and maintaining access to data and metadata versions.
- Development and implementation of data delivery systems such as ScienceBase (U.S. Geological Survey, undated c) to improve the accessibility and reusability of USGS data and reduce the effort required of research projects to release their data.

Success in moving toward an organizational culture supportive of the FAIR Principles is shown by the Bureau's communities, such as the CDI, that identify evolving data-management needs, drive updates to policy, and foster

knowledge sharing. Communication, tutorials, and outreach have demonstrated the value of the Bureau's requirement to publish datasets and metadata along with scholarly publications. The shift in data culture has required scientists to address research objectives while handling increased and more visible data-management responsibilities. As a result, researchers have produced more metadata documentation, included in the USGS Science Data Catalog, which improves data findability. Production of better documentation has contributed to improvements in data discovery and in evaluating and understanding the suitability of data for reuse (Lynch, 2008).

Although the Bureau has come a long way in supporting increasingly data-intensive integrated science research, infrastructure and support services can be improved to make USGS data more discoverable, easier to access, usable across a variety of platforms, and understandable. The FAIR workshop participants identified the following areas for improvement:

- The USGS could improve coordination across standards communities, especially by increasing consistency in data standards among similar disciplines (such as seismology and volcanology), as well as for similar types of data (for example, biological data) produced by different projects throughout the Bureau. This could be done by building an infrastructure that uses standards and supports management of the types of data and associated metadata produced throughout the USGS, which would demonstrate and encourage commonality of expression, while remaining flexible to address emerging data and policy requirements.
- The Bureau lacks sufficient guidance on techniques for documenting and sharing information about data quality and other information that is needed to determine the availability and fitness of data for reuse.
- Improved practices could help identify, describe, and discover physical samples and other scientific collections to improve accessibility and research transparency.
- For reproducibility and transparency, researchers, when using data from dynamic repositories that are continually updated, need the ability to easily share a snapshot of the data they analyzed.
- Although the USGS has improved the methods provided for known-item searching, such as when a specific author or product is known, more robust capabilities could help both humans and machines reliably discover and access the full range of data and information the USGS has to offer.
- Investment and enhancements in education and training, designed for staff in various roles across the data lifecycle, could help continue to foster a culture that supports and promotes FAIR-aligned data-management practices and tools.

USGS FAIR workshop participants also identified challenges and solutions for better alignment with the FAIR Principles. These include the following:

Putting policy into action.—For alignment with the FAIR Principles, adherence to policy requirements would be a priority at all levels of the organization that have roles in data creation, management, and distribution. Providing resources to implement the policy requirements fully is a challenge. Enhancements to communication and training could help to ensure science center staff understand the policies and their respective roles.

Working with dynamic systems.—The nature of the USGS systems that provide continuously updated data can make it difficult to access versions of data and metadata that were available at previous times of access and use, which can lead to barriers for reproducibility and transparency of scientific findings.

Dealing with emerging data types.—Data-related guidance will need to evolve to support scientific innovation and provide appropriate flexibility to enable various complex projects and data types.

Managing and using large data.—Current USGS information technology infrastructure provides insufficient capacity to support computational capabilities for efficient integration and application of large data.

Supporting enterprise systems.—Currently, there is inadequate organizational acknowledgement that some locally managed systems have been elevated to act at an enterprise level. These systems experience increased usage, throughout the Bureau, because they enable compliance with policy. This inadequate acknowledgement can lead to unpredictable long-term funding and staffing instability.

Goals of the Roadmap for Enabling the FAIR Principles

After discussing the basic characteristics of data and metadata that are aligned with the FAIR Principles and the current status of USGS data and data-management practices, workshop participants turned their attention to the roles of data producers, data storage and catalogs, value-added services, and policy makers in producing USGS scientific products. Implicit in these discussions was the recognition that open science requires the extension of the FAIR Principles to samples, methods, software, and tools (see, for example, Coalition for Publishing Data in the Earth and Space Sciences, 2018). Workshop participants focused on organizational structures to support USGS research teams in creating FAIR-aligned data, metadata, and other scientific products, as well as USGS enterprise systems to maintain the products' FAIR alignment in the future. The following goals emerged from synthesis of the workshop recommendations:

Leadership.—USGS leaders at all levels communicate the value of FAIR-aligned scientific products to achieving the USGS mission and provide resources for the necessary activities.

Policies.—The USGS Survey Manual and related policy guidance provide consistent, granular directions for carrying out the work of producing and maintaining FAIR-aligned scientific products.

Organizational culture.—Research and support staff throughout the USGS continue to increase their understanding of the value of FAIR-aligned scientific products to achieving the USGS mission and take pride in their excellent performance of the necessary activities.

Information and training.—USGS staff have access to current guidelines and best practices needed for excellence in creating and maintaining scientific products that align with the FAIR Principles.

Tools.—The USGS provides research teams, support personnel, and enterprise systems with tools and services that expedite FAIR alignment.

Research and development.—The USGS engages with relevant scientific communities to improve tools, services, and standards that improve alignment with the FAIR Principles.

Repositories and value-added data systems.—USGS enterprise systems ingest, expose, and maintain scientific products for discoverability, interoperability, and reuse.

Strategies for Enabling Better Alignment With the FAIR Principles

At present, USGS data are in many cases discoverable, accessible, and potentially interoperable for reuse, but not fully FAIR-aligned. The FAIR Principles and this roadmap's proposed strategies reach further than current USGS data practices and push toward providing a consistent, cross-disciplinary, integrated data foundation for data reuse. Implementing the strategies could allow for additional automation through machine learning, artificial intelligence, and other techniques that advance scientific research. This roadmap's data, infrastructure, and administrative strategies could make it possible for previously collected data to lead to new and unanticipated scientific discoveries. The 88 unique yet interconnected activities that compose the strategies provide clear steps for the USGS to expand its capability for integrated science through more efficient FAIR-aligned workflows and practices. The roadmap activities in aggregate present a route to FAIR-aligned USGS data. Whereas each activity, by itself, expands the potential value of USGS data, the coordinated implementation of the complete roadmap would maximize that potential and ensure that USGS data are ready for analysis and fit for reuse. The following sections provide a summary of the proposed activities, which constitute nine strategies that fall into three action areas: data, infrastructure, and administrative (fig. 2).

Table 1 lists the nine proposed strategies and the number of activities in each. Each strategy is described in more detail in the later tables, which list activities along with a suggested prioritization. The priority categories are these:

- Essential activities are first steps that are foundational to success.
- Important activities are necessary for success, but in many cases depend on prior completion of one or more essential activities.
- Useful activities are optional and could be done later to improve the alignment of USGS products with the FAIR Principles.

For convenience, each activity is assigned an identifier, which consists of the table number and the position of the activity in the table. For example, identifier 2–6 is assigned to the sixth activity listed in table 2. These identifiers are used to refer to activities in the discussions below.

Data Strategies

The FAIR Principles describe the characteristics of data and other digital products that will be ready for integration into systems that provide automated predictive capabilities. The USGS FAIR roadmap strategies in the data action area address organizational support for USGS data producers to enable creation of data that are better aligned with The FAIR Principles. Within this action area are 3 strategies that include 34 activities.

Metadata Strategy

Twenty activities that contribute to the strategy "Machine-actionable metadata improve data discovery and reuse" are listed in table 2. Automated systems will require machine-actionable metadata for findability and reusability. An essential activity for data reuse is to clarify policy requirements for all data to specify a clear and machine-accessible data usage license (see principle R1.1 in box 1) (2–1). Other essential activities can improve the consistency and machine accessibility of metadata fields that are necessary for alignment with the FAIR Principles. These activities could develop standardized approaches for providing information about data access (2–2), persistent identifiers of products (2–3), qualified references to related products (2–4), and the geospatial location of the research being presented (2–5). The resulting standardized approach for geospatial fields in USGS metadata records could enable consistent and accurate geospatial searches in catalogs and repositories. For this reason, a proposed roadmap activity is to improve interfaces and application programming interfaces for geospatial searching (2–6). Using the International Organization for Standardization (ISO) metadata content standards could facilitate standardized and

Figure 2. A diagram of the nine strategies to enable the U.S. Geological Survey to create and maintain scientific products that are aligned with the FAIR Principles, which are in three action areas: data, infrastructure, and administrative.

Table 1. Nine proposed strategies for enabling better alignment with the FAIR Principles at the U.S. Geological Survey, the number of activities in each strategy, and the table in this report with more details.

[FAIR, findable, accessible, interoperable, and reusable]

Strategies	Number of activities	Table
Data		
Machine-actionable metadata improve data discovery and reuse.	20	Table 2
Research methods generate interoperable and reusable data.	8	Table 3
Persistent identifiers link research products for automated discovery and access.	6	Table 4
Infrastructure		
Repositories and data systems provide discovery and access through unified, compatible methods.	22	Table 5
Vocabulary services improve discovery and evaluation of fitness for use.	7	Table 6
Policy requirements improve consistency of products.	6	Table 7
Administrative		
Continued culture shift, achieved through knowledge exchange, encourages activities and techniques that improve alignment with the FAIR Principles.	13	Table 8
Employee incentives improve the production of FAIR-aligned data.	3	Table 9
Research, development, and participation in professional communities improve capabilities to align with the FAIR Principles.	3	Table 10

Table 2. Proposed activities for the strategy "Machine-actionable metadata improve data discovery and reuse" and suggested priority levels.

[USGS, U.S. Geological Survey; BASIS+, Budget and Science Information System Plus; IPDS, Information Product Data System]

Identifier	Activity	Priority
2–1	Develop policy that clarifies requirements for specifying a clear and machine-accessible data usage license.	Essential
2–2	Develop a standardized approach for USGS metadata to ensure each record fully defines the procedures, online locations, and requirements for accessing data.	Essential
2–3	Determine a standardized approach for inclusion of the appropriate digital object identifiers (DOIs) in metadata records.	Essential
2–4	Develop a standardized approach to enable USGS metadata to identify relations with other metadata and data.	Important
2–5	Develop a standardized approach for geospatial fields in USGS metadata records that enables consistent and accurate geospatial searches in catalogs and repositories.	Important
2–6	Improve data catalog search application programming interfaces (APIs) and human search capabilities to offer searches of specific metadata fields, such as location keywords, author identifiers, related publications or websites, funding sources, or projects.	Important
2–7	Create a team to develop approaches for using International Organization for Standardization (ISO) geospatial metadata standards to improve interoperability and reusability while maintaining the value of the current USGS metadata collection.	Important
2–8	Create a research and development effort to develop services that efficiently and accurately translate metadata from other standards to the standards required by the USGS.	Important
2–9	Maintain processes to match USGS data with the requirements for discovery through standard dataset search services, such as data.gov and Google Dataset Search.	Important
2–10	Develop and implement processes to make metadata remain accessible even when data are no longer available.	Important
2–11	Develop policy requiring that metadata remain accessible even when data are no longer available.	Important
2–12	Create a team to develop improvements in dynamic creation of metadata elements that are needed for reusability, including the following: analysis steps; provenance including methods used, reprocessing, and new data versions; and methods to reuse the data.	Important
2–13	Link USGS information systems (such as BASIS+, IPDS, Publications Warehouse, ScienceBase, and USGS Digital Object Identifier [DOI] Tool for Data Release) to enable extraction and sharing of common information elements to reduce redundancy in documentation workflow.	Important
2–14	Develop standard templates for language in metadata records, including USGS access and use constraints.	Useful
2–15	Develop standard methods for providing detailed provenance information with data and metadata.	Useful
2–16	Work with the Metadata Reviewers Community of Practice to provide authors with a simplified checklist of key informational components needed for interoperability and where to populate them.	Useful
2–17	Create a team to develop tools to prepare metadata for integration.	Useful
2–18	Create a team to develop procedures, similar to the National Aeronautics and Space Administration's processing levels approach,[1] to identify the most appropriate version of data for a particular use, in order to enable documentation and automated access.	Useful
2–19	Create tools and procedures to categorize and assign terms for digital object type (such as project, data model, publication, service, or software) or processes of data creation (such as model, observation, or analysis).	Useful
2–20	Work with Data.gov and other Federal systems to enable advanced query capability, such as using geospatial and temporal metadata.	Useful

[1]See National Aeronautics and Space Administration (2019).

machine-actionable metadata, and so a proposed roadmap activity is to plan for the USGS transition to those standards (2–7).

Several important proposed activities could result in process and service enhancements. Services that efficiently and accurately translate metadata among content standards will be essential when the USGS transitions to the ISO standards. In the meantime, these services would ensure that USGS data are findable in most Federal systems (2–8). Processes that assist in creating metadata tags (for example, schema.org terms that are used by popular online data discovery services) could make USGS data more findable by a broader community of potential users (2–9). A process is needed to ensure that USGS metadata remain findable and accessible after the data are no longer available (see principle A.2 in box 1) (2–10); a companion activity is to develop policy governing this situation (2–11). Metadata quality and metadata production efficiency could both benefit by the creation of project teams to develop tools for automated production of metadata elements (2–12), including reuse of information that is already present in USGS information systems such as Budget and Science Information System Plus (BASIS+) and the Information Product Data System (IPDS; 2–13).

Other proposed useful FAIR roadmap activities for metadata improvement could help improve the FAIR alignment of USGS data or assist USGS metadata authors. These activities could assist authors by providing guidance on standard language for access and use constraints (2–14), on methods for providing provenance information (2–15), and on other metadata fields that are important for interoperability (2–16). A project team could be formed to develop tools to prepare metadata for combination when automated processes perform data integration (2–17). To improve the reusability of USGS data, other teams could develop defined data processing levels to assist in recognizing fitness for use (2–18), create tools and procedures to categorize digital object type (such as project, data model, publication, service, and software) (2–19), or develop tools to create and assign standard terms for processes or methods of data creation (such as model, observation, and analysis) (2–19). Another potential roadmap activity is to work with Federal partners to add geospatial search to Federal data systems like Data.gov, building on the increased standardization of geospatial information in metadata records (2–20).

Research Methods Strategy

The strategy "Research methods generate interoperable and reusable data" includes eight proposed activities that are listed in table 3. The most essential of these is to ensure that fieldwork, data processing, and data analysis staff have tools to capture rich metadata throughout the data lifecycle, so that more useful and reliable metadata records are produced by the end of each project. In many cases, such tools will require development by science programs or science centers (3–1).

During events like floods, volcanic eruptions, and oil spills, the USGS responds rapidly to collect data that can help protect life and property. When USGS teams are responding to emergencies, their data products may be most useful if they are in a state that enables immediate accessibility and interoperability so that they can be used rapidly to produce actionable information. This is challenging because in emergencies, in addition to a need for quick data processing, analysis, and dissemination, there is often loss of expected communication

Table 3. Proposed activities for the strategy "Research methods generate interoperable and reusable data" and suggested priority levels.

[USGS, U.S. Geological Survey]

Identifier	Activity	Priority
3–1	Science mission leadership ensure that appropriate staff (field crew, principal investigators) have access to and use tools that help capture rich metadata during data acquisition, processing, and analysis. This can ensure more useful and reliable metadata records by the end of the project.	Essential
3–2	Create a team to develop and implement standard operating procedures and skill-building activities to support data management for emergency response situations.	Important
3–3	Examine the provisional data release policy to enable more rapid data release in support of first responders and emergency management decision makers.	Useful
3–4	Create a team to develop registries of standard research methods to encourage their use and enable their citation in metadata.	Useful
3–5	Create web-based registries for recommended tools and standards to help USGS staff produce interoperable data.	Useful
3–6	Create a team to develop or discover standard data dictionaries and provide them online to encourage their use and enable citation in metadata.	Useful
3–7	Provide training to scientists on modern ways to create interoperable data throughout the research lifecycle.	Useful
3–8	Keep the USGS aware of domain-relevant community standards for data and metadata, especially those that enable USGS data to be combined with data from desired data-interoperability partners.	Useful

technology (such as internet access), and the work may be done by quickly assembled project teams. However, the participants in the FAIR Workshop did not have the required expertise to address the challenges of producing data during these events that are ready for immediate integration with other data to provide actionable information for responders. A proposed roadmap activity is to create a team to develop and implement standard operating procedures and skill-building activities to support data management for emergency response situations (3–2). A second proposed activity is to examine the USGS provisional data release policy to ensure its responsiveness to the need for data production and distribution during emergency events (3–3).

The remaining proposed useful activities in this strategy focus on how USGS research methods may be more precisely described to encourage data reusability: the creation of information systems that facilitate use of standard methods and enable consistent machine-actionable citations in metadata (3–5), especially a registry of standard research methods (3–4) and another registry of standard data dictionaries (3–6). The training activity 3–7 underscores the importance of understanding and adopting standards and techniques for creation and maintenance of interoperable data throughout the research lifecycle. Beyond any specific training, the USGS could also make efforts to keep staff aware of domain-relevant community standards for data and metadata, especially those that enable USGS data to be compatible with data from relevant scientific communities and stakeholders (3–8).

Persistent Identifiers Strategy

Six proposed activities, listed in table 4, compose the strategy "Persistent identifiers link research products for automated discovery and access." Explicitly identified as essential in the FAIR Principles (see box 1, principles F1, F3, and A1), persistent identifiers ensure that research products are uniquely identifiable and that the objects to which they are applied are always findable and accessible. Assigned to digital data products and publications and managed appropriately, persistent identifiers can ensure that links to these objects always resolve correctly. Applied to terms and vocabularies, persistent identifiers enable automated recognition and matching that subsequently enables interoperability and reusability. USGS policies currently require several types of persistent identifiers, including Open Researcher and Contributor Identifier (ORCID) identifiers for authors (ORCID is the authority control for data creators), digital object identifiers (DOIs) for publications and data releases, and USGS metadata identifiers for metadata files. The proposed FAIR roadmap activities would expand USGS policy and provide tools for persistently identifying additional related scientific data, objects (such as physical samples), collections, and ancillary information (such as field notebooks). Three essential activities are proposed to develop standards and policies that address the scope of objects which must be assigned persistent identifiers (4–1), the types of identifiers that the USGS will use for particular classes of objects (4–2), and the need for persistent uniform resource locators (URLs) that assure access to the identified objects (4–3). Other proposed essential activities are to develop and support additional tools to assist scientists in the generation and assignment of new types of persistent identifiers and persistent URLs (4–4), and also to update the USGS collections policy to include assignment of persistent identifiers to physical specimens to ensure they are digitally findable and to preserve provenance relations among samples, subsamples, derived data, and publications (4–5). ORCID identifiers require management by individual scientists, so the final important activity in this action area is additional training and outreach to help scientists create and properly manage their ORCID identifiers (4–6).

Table 4. Proposed activities for the strategy "Persistent identifiers link research products for automated discovery and access" and suggested priority levels.

[URL, uniform resource locator; USGS, U.S. Geological Survey]

Identifier	Activity	Priority
4–1	Develop standards and policies to clarify which additional scientific data, objects, collections, and ancillary information need to have persistent identifiers.	Essential
4–2	Develop policy addressing assignment of appropriate persistent identifiers, in addition to ORCID and digital object identifiers (DOIs), which allow retrieval through standardized communications protocols.	Essential
4–3	Develop USGS standards and policies for persistent URLs to allow uninterrupted online access to data, metadata, associated documentation, and other digital products.	Essential
4–4	Develop and support tools that apply suitable, standardized, persistent identifiers to enable findability for research objects.	Essential
4–5	Update the USGS collections policy to include assignment of persistent identifiers to physical specimens to ensure findability of samples and preserve provenance of relations among samples, subsamples, and publications.	Essential
4–6	Develop additional training and outreach materials to instruct scientists on the importance of creating and properly managing their ORCID identifiers.	Important

Infrastructure Strategies

Thirty-five activities in three proposed strategies relate to infrastructure. In contrast to the data strategies outlined above, the infrastructure strategies relate more broadly to the information systems and Bureau policies that can enable better alignment with the FAIR Principles.

Repositories and Data Systems Strategy

The infrastructure strategy with the most proposed activities (22, listed in table 5) is "Repositories and data systems provide discovery and access through unified, compatible methods." An essential first step is definition of policy and guidance to move toward more unified, compatible methods for Bureau repositories (5–1). Overall, the effect of this activity is to elevate the curation methods used in repositories for data and metadata, for collections as well as individual items, to methods described in USGS policy and followed consistently throughout the Bureau. Another proposed essential activity is provision of funding and staffing for USGS facilities that preserve physical specimens (5–2). To provide a machine-actionable path for finding and accessing USGS data and to promote interoperability, each public USGS data system or repository needs publicly available machine-operable technical documentation that identifies and links to the operational system parameters for data models and application programing interfaces (APIs) (5–3). A cooperative effort could provide this documentation. Another important proposed activity is to develop a system that provides a comprehensive, cloud-based data store connecting existing USGS data catalogs, repositories, and associated services (5–4). The data store could directly enhance findability and accessibility, while participation in the data store could also provide repository managers with guidance on standards and services that support effective data integration and reuse.

Many of the activities in this proposed strategy implicitly leverage existing repositories and data systems; however, this roadmap explicitly identifies the need to provide consistent funding and staffing for repositories and data systems (5–5). Management of data for USGS research activities requires sustained funding of systems and staff to enable USGS repositories, data warehouses, and dark archives to be properly certified by the USGS as trusted digital repositories (5–6, 5–9). This designation affirms that those systems are reliable, curated, efficient, and machine-to-machine accessible, in compliance with international standards for managing data for longevity (5–7). A specific activity is to provide funding and staffing to projects and science centers for preservation of physical specimens and for archival of physical and digital objects while they are in active use by research projects (5–8).

Three proposed activities are collaborations among USGS repositories to develop capabilities to help find and access data. First, users of the USGS websites could benefit from a capability for interfaces that link data catalogs and repositories with website tools that enable them to find data geospatially, in addition to thematically and temporally (5–10). Second, the USGS could provide federated search capabilities and access services that would improve machine and human access to data within USGS online products, including continuously updated data, such as measurements of earthquakes or streamflow (5–11). A third useful activity could improve the reusability of data in repositories by providing notice if data reuse would be affected because sensitive information was not included in the public data release (5–12). Ideally, these notifications should use both human-readable indicators on landing pages and also machine-readable flags in metadata, to allow users to quickly discover that sensitive information is not included in the data release, which might affect reuse.

This roadmap also proposes two teams to design Bureau-wide improvements in data discovery. One team could explore the development of a unified search interface that enables deeper, introspective discovery of USGS data that are contained in all USGS public data systems and repositories (5–13). A second team could investigate the design of services that offer interactive queries and visual examination of data, which can be used by multiple repositories to assist in finding and evaluating data (5–14). Another activity is to provide an enterprise method to preserve digital data through format and media conversions necessary for long-term data integrity (5–15). The roadmap proposes developing training and outreach to USGS repository operators about options to enable appropriate reuse of restricted data (5–16), providing information to data producers as well as repositories on data formats and delivery protocols currently in use by data reusers and integrators (5–17) and the advantages of making data available in multiple file formats (5–21).

The final proposed useful activities in table 5 concern the reusability of legacy data in the repositories. One activity addresses the possibility that some legacy data may be essential for multiple future projects. A cooperative effort including the trusted digital repositories could design and implement a process to identify and prioritize investments in improvements to legacy data, identifying those older data and accompanying metadata that could be of highest value to the research community if they were more FAIR-aligned (5–18). Two useful other activities propose project teams to enable automated interoperability and reusability of data that are not yet FAIR-aligned. These project teams could explore development of techniques, using artificial intelligence or machine learning, to enable machines to determine the format and organization of data so that interoperability and reusability will be machine actionable (5–19), and investigate the design of services that assist users by transforming data packages to make them accessible in multiple formats (5–20). The final proposed activity in table 5 calls for developing partnerships between data managers and communications staff to provide data access directly from social media and press releases, facilitating additional pathways to discover and access USGS data (5–22).

Table 5. Proposed activities for the strategy "Repositories and data systems provide discovery and access through unified, compatible methods" and suggested priority levels.

[USGS, U.S. Geological Survey; TDR, trusted digital repository; FAIR, findable, accessible, interoperable, and reusable; ASCII, American Standard Code for Information Interchange]

Identifier	Activity	Priority
5–1	Define USGS policy and guidance on standardized curation methods in repositories, including data and metadata, for both collections and individual items.	Essential
5–2	Provide funding and staffing for facilities across the USGS that preserve physical specimens that are used to produce scientific conclusions.	Essential
5–3	Make publicly available, for each public USGS data system or repository, a set of machine-operable technical documentation that identifies and links to operational system parameters for data models, application programming interfaces, and other details to promote interoperability.	Important
5–4	Develop a comprehensive, cloud-based data store as a value-added activity to build on and connect to existing USGS data catalogs, repositories, and associated services, and provide guidance on standards and services to support their effective integration and use.	Important
5–5	Provide funding and staffing to ensure that USGS repositories, data warehouses, and dark archives are trusted, reliable, curated, and efficient.	Important
5–6	Provide funding and staffing to USGS TDRs to ensure data persistence (and findability), including updating deprecated file formats to ensure continued data integrity and utility.	Important
5–7	Support USGS data systems to become USGS TDRs to ensure that standard practices are applied and that the systems comply with international methods to manage data for longevity.	Important
5–8	Provide consistent funding and staffing to research projects and science centers for archiving physical and digital objects they are using for research.	Important
5–9	Provide funding and staffing to USGS data systems to enable them to provide machine-to-machine accessibility to data contained within them.	Useful
5–10	Implement capability for interfaces between data catalogs and repositories with tools on USGS websites to enable data to be found geospatially, thematically, temporally.	Useful
5–11	Provide continuously updated federated search and access services for USGS digital online products that reach into individual databases and repositories and are accessible to machines as well as people.	Useful
5–12	Improve reusability of data in repositories by developing human-readable indicators on landing pages, along with machine-readable flags in metadata, for quick identification if elements are not included in the data release, such as sensitive location information.	Useful
5–13	Create a team to design a unified search interface that enables discovery of USGS data that are contained in all USGS public data systems and repositories.	Useful
5–14	Create a team to investigate and recommend the design of services that offer interactive queries and visual examination of data, which can be used by multiple repositories to assist in finding and evaluating data.	Useful
5–15	Provide repositories with an enterprise method to preserve digital data through format and media conversions.	Useful
5–16	Develop training and outreach on options for enabling appropriate reuse of restricted data.	Useful
5–17	Provide information to data producers and repositories on data formats and delivery protocols currently in use by data reusers and integrators.	Useful
5–18	Design and implement a process to identify and prioritize investments in improving the FAIR alignment of older data and its accompanying metadata.	Useful
5–19	Create a team to develop techniques, using artificial intelligence or machine learning, to enable machines to determine the format and organization of data (for example, tabular, ASCII) so that interoperability and reusability will be machine actionable.	Useful
5–20	Create a team to design services that assist users by transforming data packages so that data can be accessed in multiple formats to respond to user needs.	Useful
5–21	Increase individual data releases made available in multiple file formats and protocols.	Useful
5–22	Develop partnerships between data managers and communication staff to provide findability from social media and press releases.	Useful

Vocabulary Services Strategy

Seven proposed activities related to the second-largest infrastructure strategy, "Vocabulary services improve discovery and evaluation of fitness for use," are listed in table 6. The FAIR Principles specify that, to enable interoperability, metadata should use vocabularies that are themselves FAIR-aligned (see box 1, principles I1 and I2). In addition, the current decentralization of USGS data systems requires development of a capacity to conduct efficient federated searching and data access across platforms, which is included in the previous strategy. The foundation for this capacity will be a stabilized and extended USGS vocabulary infrastructure to support discovery across existing data systems. The first proposed essential activity to enable USGS use of standard vocabularies that align with FAIR Principles is to establish a machine-actionable catalog of USGS-approved controlled vocabularies and ontologies for use in metadata keyword fields (6–1), an activity that requires coordinated action from many groups. A second essential activity proposes to establish guidance for USGS metadata to use keywords from approved vocabularies and also to assign responsibility for identifying and endorsing those vocabularies (6–2). This strategy has two other proposed interrelated activities that are important. First, develop a standardized approach so that USGS metadata consistently include terms and keywords from the controlled vocabularies used by USGS catalogs and repositories, especially for named places (6–3). Second, create the online system that provides machine-actionable access to the approved vocabularies and ontologies (6–4).

If implemented, these activities could enable improvement of data catalog and repository search capabilities through the consistent use of USGS-approved controlled vocabularies and ontologies. An opportunity for future development is automatic linking of synonyms, especially alternate names for a single place, such as place names used historically or names used in different operational contexts (6–5 and 6–6). A future proposed activity to develop policy for use of terms from USGS approved vocabularies or ontologies in data dictionaries, would, if implemented, support machine understanding of the data values in data releases (6–7).

Policies Strategy

The six proposed activities that fall under the strategy "Policy requirements improve consistency of products" (table 7) all propose creation of a guidance infrastructure in the form of policy that produces consistent alignment with the FAIR Principles. An essential proposed first step is to review the software release policy (instructional memorandum OSQI 2019–01, see U.S. Geological Survey, 2019) to ensure alignment with FAIR (7–1). The following two proposed activities are also important: adding plans to support findability to the set of issues addressed in research project data-management plans (7–2) and ensuring that USGS policy for data versioning enables accessibility of both previous and new versions (7–3). The USGS could work with the National Archives and Records Administration (NARA) to review and update USGS scientific record schedules to align retention requirements with the FAIR Principles (7–4). The USGS could also develop guidance that coordinates the preservation and documentation requirements of records management with the FAIR Principles (7–5). Finally, this strategy proposes updated policy and guidance to clarify requirements for USGS projects to follow USGS FSP for data management and documentation, regardless of their funding source (7–6).

Table 6. Proposed activities for the strategy "Vocabulary services improve discovery and evaluation of fitness for use" and suggested priority levels.

[USGS, U.S. Geological Survey; FAIR, findable, accessible, interoperable, and reusable]

Identifier	Activity	Priority
6–1	Establish a machine-actionable catalog of USGS-approved, application programming interface (API)-enabled, controlled vocabularies and ontologies for use in metadata keyword fields.	Essential
6–2	Establish guidance for inclusion of keywords derived from USGS-approved controlled vocabularies in metadata. Establish responsibility to approve vocabularies.	Essential
6–3	Develop a standardized approach so that USGS metadata consistently include terms and keywords from the controlled vocabularies and ontologies used by USGS catalogs and repositories, especially location names, including guidance for consistent citing of controlled vocabularies.	Important
6–4	Provide access to USGS-approved FAIR vocabularies and ontologies so that all necessary information can be provided in a way that is understood by a computer without human intervention.	Important
6–5	Improve data catalog search capabilities to use USGS approved controlled vocabularies and ontologies to enable successful searches (for example, by automatically searching for similar terms and synonyms).	Important
6–6	Create a team to develop additional vocabulary services to enable discovery through use of alternate search terms (for example, historical place names and related observation sites).	Useful
6–7	Develop policy on the use of terms from USGS-approved vocabularies or ontologies (which are themselves findable, accessible, interoperable, and reusable) in data dictionaries that define the data values in data releases.	Useful

Table 7. Proposed activities for the strategy "Policy requirements improve consistency of products" and suggested priority levels.

[USGS, U.S. Geological Survey; FAIR, findable, accessible, interoperable, and reusable; FSP, Fundamental Science Practices]

Identifier	Activity	Priority
7–1	Review and update the USGS software policy (instructional memorandum OSQI 2019–01; U.S. Geological Survey, 2019) for alignment with the FAIR Principles.	Essential
7–2	Update policy on data-management plans to specify, at the outset of the project, how the project will enable data findability.	Important
7–3	Develop policy regarding accessibility to versions of data.	Important
7–4	Work with the National Archives and Records Administration (NARA) to review and update USGS scientific records schedules for alignment of retention requirements with the FAIR Principles.	Important
7–5	Work with the USGS Records Management Program to develop guidance that aligns preservation and documentation requirements of records management with the FAIR Principles (such as persistent identifiers assigned as a means of ensuring data discovery and provenance).	Important
7–6	Update policy and guidance to clarify the requirement for USGS projects to follow USGS FSP for data management and documentation, regardless of funding source.	Useful

Administrative Strategies

USGS staff create and manage the data and infrastructure, and these final 3 proposed strategies, which together total 19 activities, address human capacity and the administrative capabilities necessary to support and encourage alignment with the FAIR Principles. At present, the CDI, the John Wesley Powell Center for Analysis and Synthesis, and some programs and projects work across the USGS to foster a culture that supports practices that promote data reuse. These proposed administrative strategies reflect opportunities for further support and for providing USGS employees with information about the benefits and methods of alignment with the FAIR Principles.

Continued Culture Shift Strategy

The greatest share of these proposed activities (13, listed in table 8) contribute to the strategy, "Continued culture shift, achieved through knowledge exchange, encourages activities and techniques that improve alignment with the FAIR Principles" and address the need for continued evolution in the USGS culture surrounding FAIR Principles. The first proposed essential activity is to create a FAIR Coordinating Council to promote communication and collaboration about implementing and maintaining activities across the USGS that enable better alignment of USGS scientific products with the FAIR Principles (8–1). An essential early proposed activity is for the FAIR Coordinating Council to engage the Executive Leadership Team in guiding a continued culture shift and emphasize its importance to individual scientists and to the organization (8–2). Vital to the success of this roadmap is management support of, and messaging of alignment with, the FAIR Principles as key to the future of the USGS scientific reputation and the USGS's ability to effectively provide complex information that addresses important societal issues. Another essential proposed activity is to assess and identify gaps in the comprehensiveness and quality of existing training modules that address alignment with the FAIR Principles and to collaborate to update or create new training materials (8–3). The USGS has a new policy on Scientific Working Collections Management (U.S. Geological Survey, 2019), and an essential activity is to provide guidance and training for physical specimens and their associated data, metadata, and other artifacts (8–4). The public USGS Data Management website (https://www.usgs.gov/products/data-and-tools/data-management) is an established source of guidance and training on data-management details. The roadmap proposes to expand the information and resources available on the website to include additional guidance for alignment with the FAIR Principles (8–5).

The USGS could develop outreach materials to increase scientists' awareness and understanding of data manager roles and the value of these activities to the research project deliverable (8–6). The roadmap proposes improved policy clarity (via FSP frequently asked questions [FAQ]) on the necessity of providing the relevant provenance information to enable reproducibility and determination of fitness for use (see box 1, principle R1.2) (8–7).

This roadmap includes two proposed evaluative activities. At the Bureau level, the continued culture shift necessary to align USGS scientific products with the FAIR Principles will require establishment of procedures to evaluate improvement in USGS-wide production of FAIR-aligned data, based on repeatable metrics (8–8). For evaluation at the research project level, checklists and other rubrics could assist programs, science centers, and research projects in designing research methods and workflows that are aligned with the FAIR Principles (8–9).

This roadmap proposes formation of a team to explore and raise awareness of the multiple uses for metadata and their importance to the FAIR Principles and open data, and to provide a conceptual model for types and purposes of metadata in the USGS (8–10). Additionally, clearly established responsibility could help maintain up-to-date FAIR-aligned training modules that are posted on the public USGS Data Management website (8–11). As time is available, it could

Table 8. Proposed activities for the strategy "Continued culture shift, achieved through knowledge exchange, encourages activities and techniques that improve alignment with the FAIR Principles" and suggested priority levels.

[FAIR, findable, accessible, interoperable, and reusable; USGS, U.S. Geological Survey; FSP, Fundamental Science Practices; FAQ, frequently asked questions]

Identifier	Activity	Priority
8–1	Create a FAIR Coordinating Council to promote communication and collaboration across the USGS in order to enable better alignment with the FAIR Principles and advocate for needed support for both implementation and maintenance.	Essential
8–2	Engage the USGS Executive Leadership Team in leading a continued culture shift to emphasize the importance of the FAIR Principles as a key component of Fundamental Science Practices to the individual scientist and the organization.	Essential
8–3	Assess, identify gaps in, and work with the Office of Employee Development to update or create new training modules that address topics such as FSP, the FAIR Principles, and other data-related responsibilities.	Essential
8–4	Develop guidance and training to support USGS Survey Manual instructional memorandum CSS 2019–01 (U.S. Geological Survey, 2019) to assure proper care of physical specimens and their associated data, metadata, and other artifacts.	Essential
8–5	Expand the existing USGS Data Management public website to incorporate additional information and resources related to the FAIR Principles.	Important
8–6	Develop outreach materials to increase scientists' awareness of the role and value of data managers throughout their research projects.	Important
8–7	Improve policy clarity through FSP FAQ web pages on the necessity of providing the relevant provenance information to enable findability, reproducibility, and determination of fitness for use.	Important
8–8	Establish and implement a procedure to monitor improvement in production of FAIR-aligned data based on repeatable metrics.	Important
8–9	Create checklists to evaluate FAIR alignment of USGS workflows to aid in developing research techniques that align with the FAIR Principles.	Useful
8–10	Create a team to explore and raise awareness of the multiple purposes of metadata and their importance to the FAIR Principles and open data, which could include a conceptual model for types and purposes of metadata in the USGS.	Useful
8–11	Establish responsibility and provide resources to ensure that FAIR training modules are kept up to date and made available on the public USGS Data Management website.	Useful
8–12	Engage data-management practitioner groups to continue to leverage, expand, and make broadly available data-management guidance, training, best practices, and standards to further implement the FAIR Principles.	Useful
8–13	Work with the Office of Communications and Publishing to highlight or showcase scientific activities that build on FAIR-aligned data, including recognition of data providers. The target audience includes center directors, researchers, and field and lab technicians.	Useful

also be helpful to engage with data-management practitioner groups to continue to leverage, expand, and make broadly available data-management guidance, training, best practices, and standards for better alignment with the FAIR Principles (8–12), and to work with the Office of Communications and Publishing to highlight or showcase scientific activities that build on FAIR-aligned data, including recognition of data providers (8–13).

Employee Incentives Strategy

Three proposed activities, listed in table 9, incentivize scientists and others to adopt and advance data production and management activities aligned with the FAIR Principles and contribute to the strategy "Employee incentives improve the production of FAIR-aligned data." Although the many tasks performed by data managers are well understood within research communities, official recognition of data management as a profession is a recent phenomenon (Zuccala and others, 2008). An activity to offer professional career advancement at USGS for data managers could strengthen the implementation of FAIR-aligned data-management practices (9–1). Another important step could be for a team to consider Bureau-wide research and data-management cultures and explore possible incentives for good data management, tailored to different roles (9–2). Additionally, elements and standards related to FAIR-aligned tasks could be added to employee performance appraisal plans as appropriate (9–3).

Research, Development, and Community Participation Strategy

The "Research, development, and participation in professional communities improve capabilities to align with the FAIR Principles" strategy includes three proposed activities to keep abreast of new data-management capabilities that support the USGS mission (table 10). Support for USGS participation in national and international conferences, forums, and training

Table 9. Proposed activities for the strategy "Employee incentives improve the production of FAIR-aligned data" and suggested priority levels.

[USGS, U.S. Geological Survey; FAIR, findable, accessible, interoperable, and reusable]

Identifier	Activity	Priority
9–1	Offer professional career advancement for data managers to incentivize employees, continue to build a professional and institutional culture, and ensure the longevity of the FAIR Principles in USGS data management.	Important
9–2	Create a team to identify ways to incentivize good data management for various roles (for example, research grade and nonresearch grade USGS scientists).	Important
9–3	Incorporate FAIR-aligned elements into employee performance appraisal plans, as appropriate, to advance the FAIR Principles.	Useful

Table 10. Proposed activities for the strategy "Research, development, and participation in professional communities improve capabilities to align with the FAIR Principles" and suggested priority levels.

[USGS, U.S. Geological Survey]

Identifier	Activity	Priority
10–1	Stay abreast of emerging procedures necessary for making data findable by discovery services, through participation in conferences, national and international forums, and training.	Important
10–2	Create a pilot project to determine the benefits and efficiency of using cloud-based resources to improve accessibility, interoperability, and reusability of data across multiple domains or organizations.	Important
10–3	Advocate for, support, and coordinate USGS engagement with the Open Geospatial Consortium and other open-source initiatives that are critical to the USGS mission.	Useful

devoted to emerging techniques, infrastructures, and technologies that support the FAIR Principles will be vital for enabling the Bureau's data to be open, accessible, and ready for analysis (10–1). Creation of a pilot project is a proposed activity to determine the benefits and efficiency of using cloud-based resources to improve accessibility, interoperability, and reusability of data across multiple domains or organizations for data-intensive science (10–2). The final proposed useful roadmap activity is to engage with and support open-source communities, especially the Open Geospatial Consortium, as a way of learning about and implementing new capabilities that support the FAIR Principles (10–3).

First Steps Toward Better U.S. Geological Survey Alignment With the FAIR Principles

The proposed strategies of the FAIR roadmap described above include high-priority early steps to be taken by several different groups within USGS. This section of the report emphasizes these essential first steps for each group.

An important early proposed step in implementing and following this roadmap is forming a FAIR Coordinating Council (8–1) to include representatives from each of the USGS groups that are engaging in specific activities within the nine proposed strategies of the roadmap. As the individual groups consider the activities they are engaging, the FAIR Coordinating Council can assist in evaluating priorities, methods, and potential partners. The council can enable coordination of contingent activities such as training and information activities that depend on policy, or tool development that depends on guidance and policy and then informs training.

Upon its formation, the FAIR Coordinating Council is encouraged to prioritize the following three essential activities:

- Coordinate actions to establish a machine-actionable catalog of controlled vocabularies and ontologies that can be used in metadata tools to produce consistent keywords (6–1).
- Engage with the Executive Leadership Team to provide leadership for a continued culture shift (8–2).
- Plan for additions and improvements to training on the purpose and value of the changes USGS employees will be asked to make (8–3).

Developing policy and guidance could be important to enable better alignment with the FAIR Principles by establishing consistency in USGS products. The following are proposed essential activities to address:

- Develop policy that clarifies requirements for data usage licenses in metadata (2–1).

- Develop policy to clarify the use and requirement of persistent identifiers and URLs for selected scientific data, objects, collections, metadata, and other digital products (4–1, 4–2, 4–3).
- Develop policy on curation methods in USGS digital repositories (5–1).
- Develop policy and procedures to create and maintain a machine-actionable catalog of approved vocabularies and ontologies, and on using keywords from these sources in metadata (6–2).
- Review the USGS software policy for alignment with the FAIR Principles (7–1).

When these essential activities are complete, it would be important to develop USGS policy for activity 8–8, establishment of procedures to evaluate improvement of USGS-wide production of FAIR-aligned data, based on repeatable metrics. Full implementation of this roadmap would require development of processes for continuous qualitative and quantitative evaluations of USGS progress toward consistently providing FAIR-aligned digital products.

The proposed essential role for metadata specialists is to guide USGS staff to produce consistent and informative metadata by using the metadata content standards required by Survey Manual chapter SM 502.7 (see U.S. Geological Survey, undated d). The following two activities are of high priority:

- Develop consistent approaches for providing information about how to access data (2–2).
- Determine placement of digital object identifiers in metadata (2–3).

The proposed essential role for developers of data-management tools is development of tools that apply persistent identifiers consistently and appropriately (4–4).

Managers of the USGS collections policy have an essential role in aligning the USGS with the FAIR Principles. The following essential activities are of high priority:

- Update policy to include persistent identifiers in USGS collections (4–5).
- Offer guidance and training to assure proper care of physical specimens and associated data (8–4).

Finally, USGS management has an essential role in enabling improved USGS alignment with the FAIR Principles by providing leadership for a continued culture shift, which is most clearly communicated by authorizing resources. The proposed essential first activity is to ensure that facilities for physical specimens are adequately funded and staffed (5–2). In addition, science mission leadership has an essential role to equip their staff with tools to capture data documentation throughout their projects (3–1).

Conclusion

As a leading Federal science organization, the U.S. Geological Survey (USGS) recognizes the importance of addressing scientific data challenges by adopting scientific standards that provide efficiencies in scientific research. Among these standards, the FAIR Principles are international criteria that promote more effective data-intensive science by enabling automated processes to find, access, interoperate, and reuse data more easily. A workshop funded by the Community for Data Integration explored establishment of a FAIR roadmap for the USGS and engaged scientists and technologists from across the Bureau to identify nine proposed strategies for enabling better USGS alignment with the FAIR Principles.

Implementing this FAIR roadmap would result in creation of interrelated data policies, repositories, services, methods, and tools to enable the USGS to better serve the Nation by more efficiently producing and disseminating advanced scientific products and insights. Implementing this FAIR roadmap would require coordinating the contributions of many parts of the USGS. Scientists and data producers, data repositories and catalogs, applications, and value-added services would need the support of USGS management, human resources, and information technologists. Implementing this roadmap could help to strengthen a USGS culture that values efficient reuse of scientific products to produce new insights in combination with new information from our changing world.

Data that are aligned with the FAIR Principles are democratic data, because they are readily discoverable, available, and useful to a broad cross section of users. FAIR-aligned data are consistent with the open data movement and promote transparency of Federal Government-sponsored science and research. Importantly, USGS data that are better aligned with the FAIR Principles can be more readily usable by the scientific community, cooperators, and stakeholders, reinforcing the reputation of the USGS as a trusted data source.

References Cited

American Geophysical Union, [undated], Data at AGU: American Geophysical Union web page, accessed June 25, 2021, at https://www.agu.org/Learn-About-AGU/About-AGU/Data-Leadership.

Australian Research Data Commons, 2021, Working with the ARDC: Australian Research Data Commons web page, accessed September 1, 2021, at https://ardc.edu.au/about_us/policies-and-guidelines/.

Burwell, S.M., VanRoekel, S., Park, T., and Mancini, D.J., 2013, Open data policy—Managing information as an asset: Executive Office of the President memorandum M–13–13, 12 p., accessed February 12, 2021, at https://www.whitehouse.gov/sites/whitehouse.gov/files/omb/memoranda/2013/m-13-13.pdf.

Chang, M.Y., Carlino, J., Barnes, C., Blodgett, D.L., Bock, A., Everette, A.L., Fernette, G.L., Flint, L.E., Gordon, J., Govoni, D.L., Hay, L.E., Henkel, H.S., Hines, M.K., Holl, S.L., Homer, C., Hutchison, V.B., Ignizio, D.A., Kern, T., Lightsom, F.L., Markstrom, S.L., O'Donnell, M., Schei, J.L., Schmid, L.A., Schoephoester, K.M., Schweitzer, P.N., Skagen, S.K., Sullivan, D.J., Talbert, C., and Warren, M.P., 2015, Community for data integration 2013 annual report: U.S. Geological Survey Open-File Report 2015–1005, 36 p., accessed March 25, 2022, at https://doi.org/10.3133/ofr20151005.

Coalition for Publishing Data in the Earth and Space Sciences, 2018, Commitment statement in the Earth, space, and environmental sciences: Coalition for Publishing Data in the Earth and Space Sciences web page, accessed July 9, 2021, at https://copdess.org/enabling-fair-data-project/commitment-statement-in-the-earth-space-and-environmental-sciences/.

Directorate-General for Research and Innovation (European Commission), 2019, Cost-benefit analysis for FAIR research data—Cost of not having FAIR research data: European Union, 37 p., accessed November 19, 2021, at https://op.europa.eu/en/publication-detail/-/publication/d375368c-1a0a-11e9-8d04-01aa75ed71a1/language-en/format-PDF/source-search.

European Commission, 2018, Turning FAIR into reality—Final report and action plan from the European Commission expert group on FAIR data: European Union, prepared by Collins, S., Genova, F., Harrower, N., Hodson, S., Laaksonen, L., Mietchen, D., Petrauskaité, R., and Wittenburg, P., 78 p., accessed June 29, 2021, at https://op.europa.eu/en/publication-detail/-/publication/7769a148-f1f6-11e8-9982-01aa75ed71a1.

Faundeen, J.L., Burley, T.E., Carlino, J.A., Govoni, D.L., Henkel, H.S., Holl, S.L., Hutchison, V.B., Martín, E., Montgomery, E.T., Ladino, C.C., Tessler, S., and Zolly, L.S., 2013, The United States Geological Survey science data lifecycle model: U.S. Geological Survey Open-File Report 2013–1265, 4 p., accessed June 29, 2021, at https://doi.org/10.3133/ofr20131265.

Gunn, M.A., Matherne, A.M., and Mason, R.R., Jr., 2014, The USGS at Embudo, New Mexico—125 years of systematic streamgaging in the United States: U.S. Geological Survey Fact Sheet 2014–3034, 4 p., accessed March 25, 2022, at https://doi.org/10.3133/fs20143034.

Holdren, J.P., 2013, Increasing access to the results of federally funded scientific research: Executive Office of the President memorandum, 6 p., accessed January 18, 2021, at https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf.

Jenni, K.E., Goldhaber, M.B., Betancourt, J.L., Baron, J.S., Bristol, R.S., Cantrill, M., Exter, P.E., Focazio, M.J., Haines, J.W., Hay, L.E., Hsu, L., Labson, V.F., Lafferty, K.D., Ludwig, K.A., Milly, P.C., Morelli, T.L., Morman, S.A., Nassar, N.T., Newman, T.R., Ostroff, A.C., Read, J.S., Reed, S.C., Shapiro, C.D., Smith, R.A., Sanford, W.E., Sohl, T.L., Stets, E.G., Terando, A.J., Tillitt, D.E., Tischler, M.A., Toccalino, P.L., Wald, D.J., Waldrop, M.P., Wein, A., Weltzin, J.F., and Zimmerman, C.E., 2017, Grand challenges for integrated U.S. Geological Survey science—A workshop report: U.S. Geological Survey Open-File Report 2017–1076, 94 p., accessed June 29, 2021, at https://doi.org/10.3133/ofr20171076.

Landi, A., Thompson, M., Giannuzzi, V., Bonifazi, F., Labistida, I., Bonino da Silva Santos, L.O., and Roos, M., 2020, The "A" of FAIR—As open as possible, as closed as necessary: Data Intelligence, v. 2, no. 1–2, p. 47–55, accessed June 29, 2021, at https://doi.org/10.1162/dint_a_00027.

Lynch, C., 2008, How do your data grow?: Nature, v. 455, p. 28–29, accessed November 19, 2021, at https://doi.org/10.1038/455028a.

National Academies of Sciences, Engineering, and Medicine, 2019, Reproducibility and replicability in science: Washington, D.C., The National Academies Press, 256 p. [Also available at https://doi.org/10.17226/25303.]

National Aeronautics and Space Administration, 2019, Data processing levels: National Aeronautics and Space Administration web page, accessed June 29, 2021, at https://earthdata.nasa.gov/collaborate/open-data-services-and-software/data-information-policy/data-levels.

Obama, B.H., 2013, Executive order 13642—Making open and machine readable the new default for government information: Federal Register, v. 48, no. 93, p. 28111–28113, accessed February 8, 2021, at https://www.govinfo.gov/content/pkg/FR-2013-05-14/pdf/2013-11533.pdf.

Open Geospatial Consortium, [undated], About OGC: Open Geospatial Consortium web page, accessed September 1, 2021, at https://www.ogc.org/about.

Starr, J., Castro, E., Crosas, M., Dumontier, M., Downs, R., Duerr, R., Haak, L., Haendel, M., Herman, I., Hodson, S., Hourclé, J., Kratz, J., Lin, J., Nielsen, L., Nurnberger, A., Proell, S., Rauber, A., Sacchi, S., Smith, A., Taylor, M., and Clark, T., 2015, Achieving human and machine accessibility of cited data in scholarly publications: PeerJ Computer Science, v. 1, no. 1, accessed September 1, 2021, at https://doi.org/10.7717/peerj-cs.1.

Tenopir, C., Christian, L., Allard, S., and Borycz, J., 2018, Research data sharing—Practices and attitudes of geophysicists: Earth and Space Science, v. 5, no. 12, p. 891–902, accessed June 29, 2021, at https://doi.org/10.1029/2018EA000461.

U.S. Geological Survey, 2007, Facing tomorrow's challenges—U.S. Geological Survey science in the decade 2007–2017: U.S. Geological Survey Circular 1309, 67 p., accessed June 29, 2021, at https://doi.org/10.3133/cir1309.

U.S. Geological Survey, 2016, Public access to results of federally funded research at the U.S. Geological Survey—Scholarly publications and digital data: U.S. Geological Survey, 22 p., accessed December 17, 2020, at https://prd-wret.s3.us-west-2.amazonaws.com/assets/palladium/production/atoms/files/USGS-PublicAccessPlan-APPROVED-v1-03.pdf.

U.S. Geological Survey, 2019, IM CSS 2019–01: U.S. Geological Survey web page, accessed July 8, 2021, at https://www.usgs.gov/about/organization/science-support/survey-manual/im-css-2019-01.

U.S. Geological Survey, 2021, Data management: U.S. Geological Survey web page, accessed June 29, 2021, at https://www.usgs.gov/products/data-and-tools/data-management.

U.S. Geological Survey, [undated] a, Fundamental science practices (FSP) policy directives: U.S. Geological Survey web page, accessed June 29, 2021, at https://www.usgs.gov/about/organization/science-support/office-science-quality-and-integrity/policy-directives.

U.S. Geological Survey, [undated] b, Science data catalog (SDC) (ver. 3.0.1): U.S. Geological Survey database, accessed June 30, 2021, at https://data.usgs.gov/datacatalog/.

U.S. Geological Survey, [undated] c, ScienceBase (ver. 2.169.0-489–g73a767f): U.S. Geological Survey database, accessed June 30, 2021, at https://www.sciencebase.gov/catalog/.

U.S. Geological Survey, [undated] d, Survey manual: U.S. Geological Survey web page, accessed June 29, 2021, at https://www.usgs.gov/about/organization/science-support/survey-manual.

Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., Santos, L.B. da S., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A.J.G., Groth, P., Goble, C., Grethe, J.S., Heringa, J., Hoen, P.A.C. 't, Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M.A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., and Mons, B., 2016, The FAIR guiding principles for scientific data management and stewardship: Scientific Data, v. 3, no. 1, [article 160018], 9 p., accessed June 29, 2021, at https://doi.org/10.1038/sdata.2016.18.

Zuccala, A., Oppenheim, C., and Dhiensa, R., 2008, Managing and evaluating digital repositories: Information Research, v. 13, no. 1, [paper 333], accessed November 19, 2021, at http://informationr.net/ir/13-1/paper333.html.

Glossary

dark archive An archive that does not grant public access and preserves the information it contains. The purpose of a dark archive is to be a repository for information that can be used as a failsafe during disaster recovery.

data management Activities that enable the location, sharing, and reuse of data.

data warehouse A facility for organized storage and access to data.

Earth Monitoring, Analyses, and Projections (EarthMAP) A U.S. Geological Survey initiative to integrate its natural science activities by using data integration, high performance computing, artificial intelligence, modeling, analytics, laboratory facilities, visualization, and decision-support tools.

enterprise systems An organization's centralized computing systems and applications, often serving as data repositories or catalogs, or as services that provide standard data-management methods.

Fundamental Science Practices A set of policies, practices, philosophical premises, and operational principles used by the U.S. Geological Survey that safeguard the integrity of scientific research and products.

Fundamental Science Practices Advisory Council The advisory and working committee that provides support for full implementation of the U.S. Geological Survey Fundamental Science Practices.

instructional memoranda Within the Survey Manual of the U.S. Geological Survey, temporary policies and general procedures that will be incorporated into the Survey Manual after the policy of the memorandum has been implemented and revised, if needed.

John Wesley Powell Center for Analysis and Synthesis This U.S. Geological Survey center fosters innovative thinking in Earth system science through collaborative synthesis activities.

metadata Descriptive information about data, including who, what, where, when, why, and how, so that they can be understood, reused, and integrated with other data.

ORCID An organization that provides a persistent digital identifier that distinguishes an individual from other researchers and a record that supports automatic links among a researcher's professional activities.

snapshot In information management, a snapshot is a read-only copy of a database that preserves a copy of the content of the database at a certain time.

use case In information system development, a use case is a description of the steps taken by a hypothetical system user in order to achieve that user's goal. A collection of use cases is analyzed to determine the functionality that the information system must provide.

Appendix 1. FAIR Workshop Participants

Neil Baertlein, U.S. Geological Survey
Tara Bell, U.S. Geological Survey
Wade Bishop, University of Tennessee
Tom Burley, U.S. Geological Survey
Linda Debrewer, U.S. Geological Survey
Jon Dewitz, U.S. Geological Survey
Angie Diefenbach, U.S. Geological Survey
Jason Ferrante, U.S. Geological Survey
David Govoni, U.S. Geological Survey
Michelle Guy, U.S. Geological Survey
Viv Hutchison, U.S. Geological Survey
Drew Ignizio, U.S. Geological Survey
Mikki Johnson, U.S. Geological Survey
Keith Kirk, U.S. Geological Survey
Natalie Latysh, U.S. Geological Survey
Fran Lightsom, U.S. Geological Survey
Tom Murray, U.S. Geological Survey
Daniel Pearson, U.S. Geological Survey
Ray Plante, National Institute of Standards and Technology
Emily Read, U.S. Geological Survey
Carma San Juan, U.S. Geological Survey
Jason Sherba, U.S. Geological Survey
Rich Signell, U.S. Geological Survey
Chris Skinner, U.S. Geological Survey
Shelley Stall, American Geophysical Union
Nancy Sternberg, U.S. Geological Survey
Roland Viger, U.S. Geological Survey
Lisa Zolly, U.S. Geological Survey

For more information about this report, contact:
Director, Woods Hole Coastal and Marine Science Center
U.S. Geological Survey
384 Woods Hole Road
Quissett Campus
Woods Hole, MA 02543–1598
508–548–8700 or 508–457–2200
whsc_science_director@usgs.gov
or visit our website at
https://www.usgs.gov/centers/whcmsc

Publishing support provided by the
Pembroke Publishing Service Center

ISSN 2331-1258 (online)
https://doi.org/10.3133/ofr20221043