



Public Access to Results of Federally Funded Research at the U.S. Geological Survey: Scholarly Publications and Digital Data

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



Contents

1.0 Background and Purpose	1
2.0 Scope	3
2.1 Definitions.....	3
3.0 Requirements	4
4.0 Applicability.....	5
5.0 Authority.....	5
6.0 Roles and Responsibilities.....	6
7.0 Implementation	7
8.0 Objectives for Public Access to Scholarly Publications.....	9
8.1 Timely digital access to publications.....	9
8.1.1 Public search, discovery, and access.....	10
8.1.2 Metadata associated with Publications and Data	11
8.1.3 Public-private collaboration	11
8.1.4 Attribution to authors, journals, and publishers.....	11
8.1.5 Unauthorized Distribution	11
8.1.6 Archival.....	12
9.0 Objectives for Maximizing Public Access to Digital Scientific Data.....	12
9.1 Data management plans	13
9.2 Data management costs.....	13
9.3 Evaluation of data management plans	13
9.4 Data management compliance.....	14
9.5 Data repositories and databases.....	14
9.6 Public-private collaboration	14
9.7 Attribution in scientific datasets	15
9.8 Training, education, and workforce development.....	15
9.9 Data preservation	15
10.0 Metrics, Compliance and Evaluation.....	16
11.0 Public Consultation Experience	16
12.0 Interagency Coordination.....	16
13.0 Public Notice.....	17
14.0 Update and Re-evaluation of the Plan.....	17
15.0 Timeline for Implementation	17
16.0 Resources	17
17.0 Risks.....	18
18.0 References Cited.....	18
Appendix A -- Summary of USGS Policies Relating to Public Access and Open Data.....	19
Appendix B -- List of Acronyms	22

USGS Public Access Plan Revision History

The original version of this document was dated January 8, 2016, to reflect the date it was cleared for dissemination. As processes evolve or other updates to this document are needed, they will be listed as changes in this revision history. Subsequent versions of the document will not be dated.

The table below shows the date when substantial changes (excludes minor typographic corrections) have been made and provides a brief description of the changes. Any questions about this revision history should be directed to kkirk@usgs.gov.

DATE	CHANGE
3/9/2016	Revised text related to the timing for public access to agency funded data. The revised text clarifies that funded data upon which scholarly conclusions are based must be made publically accessible at the same time or prior to release of the publication of the scholarly conclusions and that conversely, agency funded data not directly associated with a scholarly publication is made publically accessible as specified in a data management plan.

1.0 Background and Purpose

The U.S. Geological Survey (USGS), a Bureau of the U.S. Department of the Interior (DOI), conducts objective scientific research in ecosystems, climate and land use change, energy and mineral assessments, environmental health, natural hazards, and water resources to inform effective decision making and planning; produces information to increase understanding of natural hazards such as earthquakes, volcanoes, and landslides; conducts research on oil, gas, and alternative energy potential, production, consumption, and environmental effects; and leads the effort on climate change science research for the DOI. The results of USGS research, generally released in the form of publications, maps, data, and models, are used by policymakers at all levels of government and by the private sector to support appropriate decisions about how to respond to natural risks and manage natural resources.

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 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 (refer to appendix A).

USGS data are currently being used to spur economic growth, competitiveness and innovation in the private sector. For example, USGS streamgage data are widely used in many industries from tourism and recreation to construction planning, and USGS Landsat data are used globally to optimize crop production, identify drought and inform decision makers on sustainable land use practices.

USGS scholarly publications and associated data are discoverable online at these USGS Websites: publications at the USGS [Publications Warehouse](#) and data from the USGS [Science Data Catalog](#) and [Maps, Imagery, and Publications](#) portal. Currently, citations for the more than 50,000 USGS series publications are available, and 10,000 of these are also available free to the public as downloadable digital files. Additionally, more than 41,000 scholarly publications authored by the USGS but published externally are cataloged in the Publications Warehouse, and links to original published sources are provided.

On February 22, 2013, the Office of Science and Technology Policy (OSTP) issued a memorandum, *Increasing Access to the Results of Federally Funded Scientific Research* (Holdren, 2013), which called on all Federal agencies with annual research and development (R&D) expenditures of more than \$100 million to develop a plan to increase public access to the direct results of federally funded scientific research, including specifically peer-reviewed publications and digital data. The DOI's total annual Research and Development (R&D) budget in FY 2015 was approximately \$925 million. 74 percent (\$686 million) of that funding was allocated to the USGS, with lesser amounts, all less than \$100 million, distributed among other DOI Bureaus. This Plan focuses specifically upon the USGS's 'public access' activities, policies, and plans, as they affect both intramural and extramural R&D. Of the \$686 million USGS allocation in FY 2015, \$70 million, just over 10 percent, was allocated for extramural R&D.

On May 9, 2013, the Administration issued Executive Order 13642, *Making Open and Machine Readable the New Default for Government Information* (Obama, 2013), and the Office of Management and Budget (OMB) also released Memorandum M-13-13, *Open Data Policy-- Managing Information as an Asset* (Burwell et al., 2013). Individually and collectively these directives established the mandates for the Federal Government to transform data and information into useable and accessible digital artifacts and promote and accelerate their release, subject to certain limitations imposed by privacy, confidentiality, and national security considerations.

This document elucidates how the USGS currently provides and upon the effective date of this plan will further enable public access to USGS scholarly publications and digital data subject to: law; agency mission; resource constraints; U.S. national, homeland, and economic security; and the objectives listed in the OSTP'S public access memorandum. It also describes how existing USGS processes and procedures are being modified to meet the requirements of the OSTP "public access" and OMB "open data" memoranda. Note that at the first use of terms commonly referred to throughout this document the term acronym directly follows, shown in parenthesis and a complete list of the terms and their related acronyms is found in Appendix B.

2.0 Scope

This plan applies to scholarly publications and unclassified or otherwise unrestricted digital research data (i.e. digital data required to validate research findings) produced in whole or in part by the USGS, contract employees, financial assistant awardees, other grantees, and other contractor entities where the publication and data are produced with complete or partial USGS funding, unless otherwise prohibited by law, regulation or policy.

This plan builds on existing USGS policy, summarized in appendix A, which requires public access be provided for any scholarly publications and associated data that:

- Arise from research conducted directly by USGS or by others using USGS funding,
- Published by the USGS, with no embargo
- Published externally by USGS scientists or extramural USGS funded scientists, with embargo.

This existing policy provides no embargo for digital data required to validate research findings, as those data must be made available at the time of publication to support scholarly conclusions.

2.1 Definitions

For the purposes of this Plan, the following terms are defined:

Dark Archive – is an archive that does not grant public access and preserves the information it contains. The purpose of a dark archive is to function as a repository for information that can be used as a failsafe during disaster recovery. Making a document “bright” refers to retrieving it from the dark archive and placing it in a publicly accessible repository.

External publications – Are any USGS information published by any non-USGS entity, including but not limited to scientific journals, professional society volumes, cooperating agency series, and university or commercial publishers.

Fundamental Science Practices (FSP) – The set of mandated requirements, codified in USGS policy, that describe the peer review process for publishing research results, releasing data, and other aspects of the scientific process conducted by USGS scientists.

Information Product – An information product presents scientific knowledge (observations, facts, or interpretations thereof) communicated by speech, text, graphics, or maps, in any medium (e.g., print, digital, Web, audiovisual), to a defined audience or customer, scientific or nonscientific, internal or external. It includes both publications and data.

Research data – Research data is defined as the recorded factual material commonly accepted in the scientific community as necessary to validate research findings, but not any of the following: Preliminary analyses, drafts of scientific papers, plans for future research, peer reviews, or communications with colleagues. This “recorded” material excludes physical objects (e.g., laboratory samples).

Scholarly publications and associated research data – Any publication that presents the results of USGS-funded research and any data associated with that research and which are included within the publication or accompany the published work as a separate digital data set referenced back to the scholarly publication.

USGS Information Product Data System (IPDS) – The internal USGS IPDS is a Web based application and dark archive that documents and tracks the review, approval, and dissemination of information products that are subject to Fundamental Science Practices (FSP). The IPDS also provides bibliographic metadata to the USGS Publications Warehouse for creating citations. IPDS acts as a dark archive for information products not yet available for public release. IPDS houses the accepted manuscripts (AM) from journals for publications funded by the USGS, both intramural and extramural. If a publisher fails to provide public access after the embargo period expires a copy of the AM from IPDS will be provided to the USGS publications warehouse to meet public access requirements.

USGS ScienceBase – is a digital repository and collaborative data management platform providing public access to machine readable data and metadata. ScienceBase also serves as a cross-program institutional repository, ensuring long-term stewardship of the results of federally-funded research, when a more specific Program-level data repository is not available. ScienceBase allows scientists to contribute new and original data content in any file format, providing advanced access and integration capabilities for certain types of formats (e.g., shapefiles and GeoTIFF served via Open Geospatial Consortium Web services). ScienceBase is designed to integrate into the *Science Data Lifecycle* (Faundeen et al, 2013), offering features for project teams to develop and manage data securely and to facilitate final review and approval for public release.

USGS Science Data Catalog – is a public search and discovery tool for fully documented USGS science data, for example by the FGDC Content Standard for Digital Geospatial Metadata (CSDGM). Metadata are available through the USGS Science Data Catalog from across the USGS Mission Areas and Programs. USGS Data Stewards contribute metadata to the Science Data Catalog via the Science Data Catalog Dashboard application, which allows for data contributions via multiple methods (i.e. Web-accessible folders, existing USGS systems) and provides various summary reports.

USGS Publications Warehouse – is an online citation index for USGS-authored publications managed by the USGS Library that serves as the authoritative source for information and access to USGS publications. Each publication has its own descriptive citation page that is dynamically generated based on information stored in a database. The Publications Warehouse cataloging team builds and maintains records based on data derived from a variety of sources, including the USGS Information Product Data System (IPDS), USGS Science Publishing Network pages and announcements and other bibliographic databases. The Publications Warehouse site is built in such a way to allow easy indexing by web search crawlers, and provides both basic and advanced search capabilities. The Publications Warehouse also provides a number of different Web services, including a customizable RSS feed and a MODs XML service. By the effective date of this plan, the publications warehouse will also contain citations for and provide access to publications produced by USGS funded extramural researchers after any embargo period has expired. It will provide XML versions of all publications published after the effective date.

USGS series publications – A set of numbered information products, each series is established to meet a specific audience need and produced in-house by the USGS and publicly released gratis on the USGS Website.

3.0 Requirements

Much of this plan refers to requirements or activities that already exist or are in the process of being implemented. This plan also identifies tasks that will impose additional requirements on both USGS intramural and extramural entities funded by the USGS upon its effective date. Specifically, this plan requires that an electronic copy of either the final accepted manuscript (AM) or the final publication (i.e., publication of record (PoR)) shall be available free-of-charge for public access not more than twelve months after the official publication date. Supporting digital data shall be released and made available free to the public prior to the publication of the PoR or at the same time. Further, final data associated with USGS funded research will be made available free to the public. These data are released and made available free to the public unless the agency agrees that a demonstrated special circumstance prevents the digital research data from being made publicly available, for example because it provides location data for endangered species. This plan also requires provision of formal Data Management Plans (DMPs) for all new research proposals. The DMP requirement became effective February 2015 for USGS intramural scientists and will apply to all USGS funded extramural scientists beginning January 2016.

4.0 Applicability

Upon the effective date of this plan these requirements will apply to all publications of research results and digital data arising from complete or partial USGS funding, unless otherwise prohibited by law, regulation, or policy. These information products and data must be provided free-of-charge.

Currently the USGS requires public access for all approved peer-reviewed science information products,¹ including associated data,² regardless of media, whether published and disseminated by the USGS³ or by an outside entity,⁴ if the author has full time, part time, volunteer, or emeritus USGS affiliation or shared affiliation (for example, between the USGS and a university).⁵

5.0 Authority

The mandate to publish data and findings from USGS science activities dates to the Bureau's creation by the signing of the Sundry Civil Bill (U.S. Statutes at Large, v. 20, p. 394-395) on

¹ SM Chapter 502.4 effective 12/16/11: "Review, Approval, and Release of Information Products" and SM Chapter 500.24 effective 4/9/93: "Policy for Release of Computer Databases and Computer Programs."

² IM OSQI 2015-001 effective 2/19/2015: "Fundamental Science Practices: Scientific Data Management Foundation Policy", IM OSQI 2015-002 effective 2/19/2015: Fundamental Science Practices: Metadata for Scientific Data, Software, and Other Information Products, IM OSQI 2015-003 effective 2/19/2015 Fundamental Science Practices: Release of Scientific Data and IM OSQI 2015-004 effective 2/19/2015: Fundamental Science Practices: Preservation Requirements for Scientific Data.

³ SM Chapter 1100.3 effective 12/29/2003: "U.S. Geological Survey Publication Series."

⁴ SM Chapter 1100.4 effective 07/02/04: "Use of Outside Publications, Including Abstracts."

⁵ Upon the effective date of this plan these USGS policies will apply to both intramural and extramural funded research results and associated data arising from USGS funding.

March 3, 1879, establishing the USGS. This bill also defined the requirement to report the results of investigations by the USGS to the public.

In addition, 2 CFR § 200.315 Intangible property provides the Federal awarding agency the right to reserve a royalty-free, nonexclusive and irrevocable right to reproduce, publish, or otherwise use any work that is subject to copyright and was developed, or for which ownership was acquired, under a Federal award for Federal purposes, and to authorize others to do so; and pursuant to 2 CFR 215.36 Intangible Property, the Federal government has the right to: (1) Obtain, reproduce, publish, or otherwise use the data produced under a Federal award; and (2) Authorize others to receive, reproduce, publish, or otherwise use such data for Federal purposes.

6.0 Roles and Responsibilities

USGS entities responsible for implementing the actions outlined in this Plan include:

- **The Office of Science Quality and Integrity (OSQI)** – which is responsible for maintaining and enforcing the USGS Fundamental Science Practices which provide the primary policy basis for the processes elucidated in this plan.
- **The Office of Communications and Publishing** – which through the USGS Science Publishing Network is responsible for editing, production, preparation, section 508 compliance, and release of USGS Series publications.
- **The Core Science Systems Mission Area** – which is responsible for the USGS Publications Warehouse (part of the USGS Library system), USGS Science Data Catalog and USGS ScienceBase data repository. The Community for Data Integration which created and manages the USGS Data Management Website is facilitated by this mission area.
- **The Office of Enterprise Information** – which is responsible for the administration of certain repositories and archives.
- **USGS Science Center Directors** – which are responsible for ensuring the scientists reporting to those science centers follow the USGS Fundamental Science Practices.
- **Office of Acquisition and Grants** – which is responsible for providing funding through financial assistance to scientists extramural to the USGS and enforcing requirements on those extramural scientists, including creation and adherence to data management plans and delivery to the USGS of digital versions of their published manuscripts and associated data.

7.0 Implementation

Many of the organizational and technological resources needed to fully implement public access to USGS scholarly publications and digital research data already exist within the USGS. The mechanisms for ensuring public access and measuring progress are established through the USGS FSP.

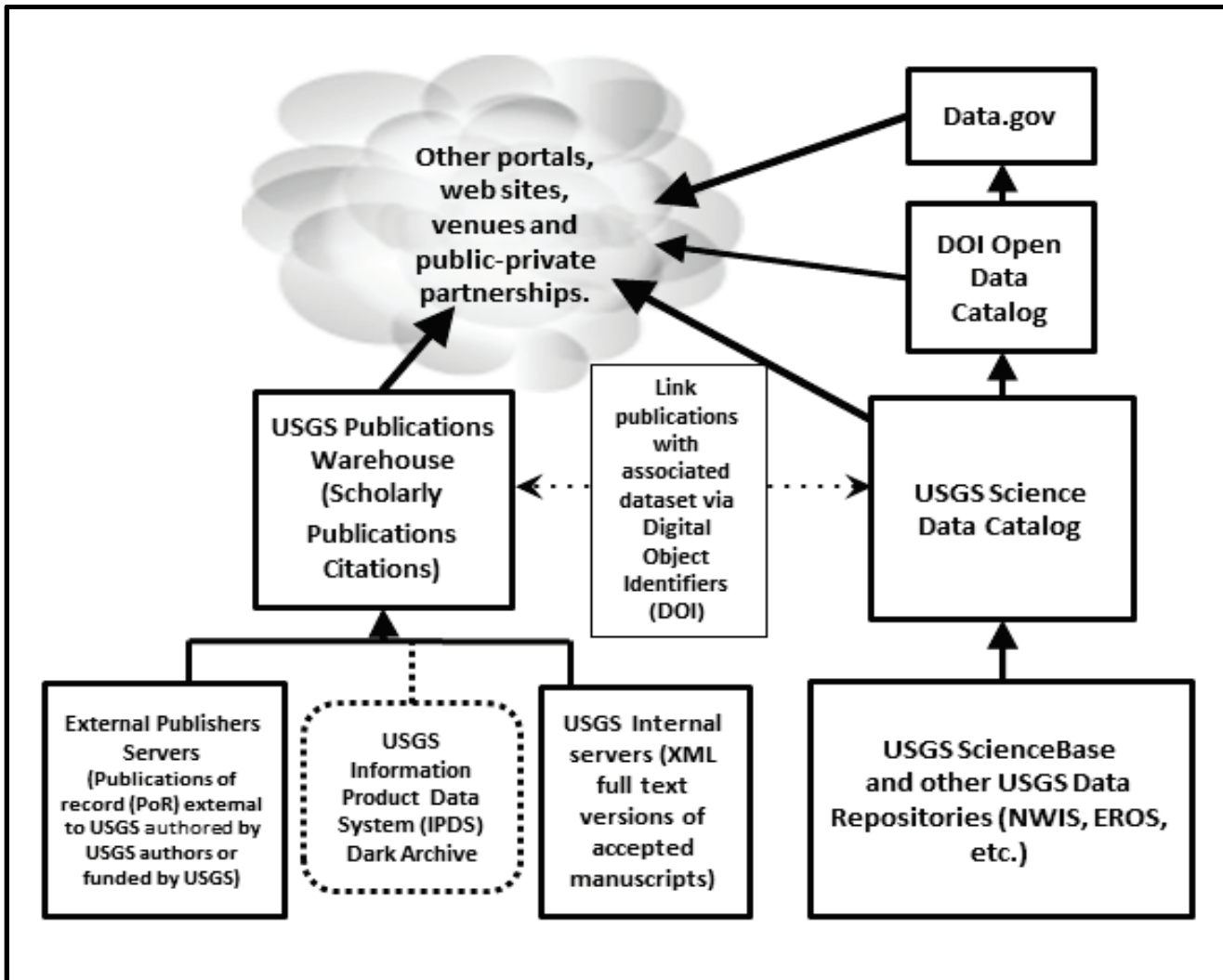


Figure 1. A generalized diagram illustrating the public release process for USGS scholarly publications and digital research data. *Note: Metadata from the USGS Publications Warehouse meet Library of Congress metadata standards while USGS Science Data Catalog metadata meet Federal Geographic Data Committee metadata standards. Details of how publications and data are linked via use of DOIs are described in text.*

The USGS public access process for scholarly publications and digital research data that will be in place after the effective date of this plan is illustrated in figure 1. Key elements include:

For scholarly publications

- Upon completion of peer review, Bureau approval, and production, USGS series publications will be submitted to the USGS Publications Warehouse for cataloging and public release in both PDF and XML formats.
- As part of the USGS Bureau approval process, the full-text of all USGS final manuscripts will reside in the Information Product Data System (IPDS), a [National Archives and Records Administration](#) (NARA)-certified repository that functions in part as a dark archive for all USGS-funded information products. This includes electronic copies of all USGS-funded publications by scientists extramural to the USGS. Publications by intramural scientists will be entered in the IPDS dark archive by those scientists, and publications by extramural scientists funded by the USGS will be provided to the USGS project scientist responsible for the USGS funded extramural research, who will place the publication into the IPDS dark archive.
- Upon release, all USGS series publications and scholarly publications published externally (resulting from both intramural and extramural research) will be cataloged in the USGS Publications Warehouse. Associated data will be cataloged in USGS Science Data Catalog prior to, or concurrent with the release of either the AM or PoR (that is, irrespective of any embargo period). Links will be provided to the full text of all information products regardless of publisher (USGS or non-USGS) and type of access (publicly accessible or accessible only to subscribers).
- As part of this plan the USGS has contracted with [Clearinghouse for the Open Research of the United States](#) (CHORUS), to provide notification to the USGS of the release of USGS-funded extramural publications on non-USGS publishers' websites for those publishers participating in the CHORUS service. These extramural publications (i.e., not published as USGS series reports using the USGS in-house publishing process of the USGS Science Publishing Network) and their metadata will then be cataloged in the USGS Publications warehouse, and a link provided to the full-text of the article on the publisher's website, where free public access will be provided after the embargo period has expired.
- For those publications for which the full text cannot be accessed through CHORUS (publisher not part of CHORUS) or because the publishers embargo period is longer than 12 months, the accepted manuscripts stored in the IPDS dark archive will be made bright via USGS internal servers and accessible through the USGS Publications Warehouse.

For digital data

- Data associated with USGS scholarly publications currently are either contained within the publication or may be released as a separate data entity (e.g., as digital datasets) after passing through quality assurance, quality control, formal description, review, and approval for release as established in Survey Manual (SM).⁶
- Data associated with USGS scholarly publications and other data approved for release are currently submitted to the internal USGS data repository most appropriate to their content, level of data management needed, and dissemination method required. A trusted third-party distribution repository may be used to increase access, but the authoritative version of the data is maintained by the USGS. Regardless of the repository, metadata describing the data are indexed in the USGS Science Data Catalog, enabling consistent public discoverability.
- USGS data are subject to Bureau approval under FSP; once approved for release, data are made available to the public in the appropriate forms and formats free-of-charge with no embargo period.

The USGS (for USGS series publications and data sets in ScienceBase and other USGS data repositories) and journal publishers (for USGS scholarly publications published externally) assign and register Digital Object Identifiers (DOIs) to aid in unique identification, citation, recordkeeping, and cross-referencing between related resources.

In keeping with the OSTP public access memorandum, the following sections describe how the USGS is either meeting required objectives or the actions that will be taken by FY 2017 to ensure the objectives are met.

8.0 Objectives for Public Access to Scholarly Publications

The following objectives describe the USGS plan to provide public access to scholarly publications and align with Section 3 of the OSTP memorandum:

8.1 Timely digital access to publications

Timely search, discovery and access to all publications subject to this plan will be provided by the USGS Publications Warehouse. USGS series publications (including historical publications) are made publicly available free-of-charge in digital form as PDF for download through the Warehouse immediately upon public release. All other scholarly publications by USGS authors are cataloged in the USGS Publications Warehouse with appropriate bibliographic metadata shortly after their publication in the respective journals, and links are provided to the full text of those scholarly publications. The full text will be made available free-of-charge to the public to read, download, and analyze in digital form not later than 12 months after the official date of publication from either the publisher's website or from USGS. In either case, USGS will maintain a full-text archive of all peer reviewed publications resulting from its research.

⁶ See SM Chapters 502.2 ([Planning and Conducting Data Collection and Research](#)), and 502.4 ([Review, Approval, and Release of Information Products](#)) and SM Instructional Memoranda (IM) IM OSQI 2015-001 ([Foundation Data Policy](#)), IM OSQI 2015-002 ([Metadata Policy](#)), IM OSQI 2015-003 ([Data Release Policy](#)), and IM OSQI 2015-004 ([Data Preservation Policy](#)).

8.1.1 Public search, discovery, and access

The USGS will enable public search, discovery, and access to full text Section 508-compliant⁷ digital versions of all USGS funded publications resulting from its research via its existing USGS Publications Warehouse. The USGS is retooling its scientific publishing process to provide all publications resulting from its research in machine-readable XML format in addition to other human-readable formats to facilitate discovery of and access to these publications by automated systems. The anticipated implementation date of the XML format delivery for all new publications is end of FY2016. The files and associated bibliographic metadata will continue to be released online in these formats through the USGS Publications Warehouse free-of-charge. Providing USGS information products and their associated metadata in XML will promote indexing by public search engines, further enhancing discoverability.

All USGS funded, peer-reviewed and disseminated publications released outside of the USGS (e.g., as journal articles) will be cataloged in the Publications Warehouse with appropriate metadata to ensure discovery by the public. The full text of the final, accepted, peer-reviewed manuscript will be stored in the IPDS ready for release at the end of any embargo period and linked to the record in the Publications Warehouse. Under existing terms of award for contracts and financial assistance, the full text of publications produced by USGS funded scientists extramural to the USGS must be stored electronically in the IPDS.

To supplement this capability, provide a mechanism for determining the official date of publication and end of the embargo period, and provide access to the final published version of the article, USGS has entered into an agreement with CHORUS to provide free public access to the journal articles, if available, on a partnering publisher's site. For journals not partnered with CHORUS or partnered but not complying with the 12-month (maximum) embargo, the version of the manuscript held in the IPDS will be made bright (i.e., publicly accessible at no charge).

These mechanisms, taken together, will facilitate timely access to peer-reviewed scholarly publications and associated metadata directly arising from research funded by the USGS. They will allow the public to read, download and analyze the publications by machine and will support downstream information processing and dissemination activities by using machine-readable and open formats, data standards, and common core and extensible metadata.

The USGS will provide a mechanism as part of USGS Publications Warehouse for accepting petitions for changes to the embargo period. It will be available commensurate with the effective date of this Plan. Requesters will fill out a web form with details about their embargo change request. The petition will be received and coordinated by the USGS Office of Science Quality and Integrity. If the petition is granted and would result in immediate release for publication, the publication will be made public within two weeks of the decision to allow for processing time needed by USGS to move from the IPDS dark archive to the USGS Publications Warehouse.

⁷ Meaning compliant with section 508 of the Rehabilitation Act of 1973 as amended in 1998 (29 U.S.C. § 794d)

8.1.2 Metadata associated with Publications and Data

After the effective date of this plan, metadata for USGS series publications and all journal articles from publishers outside the USGS that result from USGS-funded research will be included in the USGS Publications Warehouse and publically available free-of-charge at the time of publication. This catalog will be used to facilitate discovery of USGS intramural and extramural publications as well as aggregation of publications resources into Data.gov and any other specialized indexes and discovery mechanisms, as illustrated in figure 1.

Metadata associated with USGS data products are released free to the public with no embargo through the USGS Science Data Catalog coincident with or before the publication dependent on that data. These two systems represent the authoritative source for discovery of published research and digital data produced or funded by the USGS. Both resources provide metadata in standard formats and use methods that conform to established industry standards. These methods (such as embedding structured metadata in Web pages) facilitate the use of common commercial and open-source tools such as bibliographic management software and promote innovative access to research results in industry and the global scientific community.

8.1.3 Public-private collaboration

The USGS encourages public-private collaboration by allowing third parties to easily harvest metadata records and links from both the USGS Publications Warehouse and USGS ScienceBase that may then be used in external discovery and repository systems. The USGS participates in CrossRef and DataCite, using DOIs designed to promote access to research publications and associated data. USGS metadata is being harvested from and provided to a variety of third party search engines and vendors of library systems. This metadata exchange provides discovery of USGS-funded publications. USGS provides multiple options for accessing and extracting metadata records including via machine APIs and in downloadable formats including RIS, CSV, TSV, Excel, RSS, and JSON.

8.1.4 Attribution to authors, journals, and publishers

Using existing processes, metadata for both USGS series publications and all journal publications made accessible through the USGS Publications Warehouse complies with existing Library of Congress standards for metadata encoding and release so that publications are fully attributed to the appropriate authors, journals, and publishers. The USGS will continue to use these existing processes to meet this requirement and implement improved mechanisms as they evolve.

8.1.5 Unauthorized Distribution

Neither the USGS Publications Warehouse or USGS Science Base are capable of providing mass downloading of multiple publications or multiple data sets which assists in limiting unauthorized distribution. The USGS monitors online retailers of publications to determine if USGS funded publications are being mass-produced and sold through these outlets. When instances of illegal mass duplication and distribution are discovered they are reported to the USGS solicitor's office, who in turn contact the web seller, and the illegal publications are removed.

8.1.6 Archiving

USGS series publications are currently stored as 508 compliant and downloadable files from a system that is both replicated and accessible. This system, the USGS IPDS, is part of the USGS Records Disposition Schedule and therefore meets NARA requirements for digital preservation, that is, publications and metadata are stored in an archival solution that provides for long-term preservation and access to the content without charge to the public. Final versions of both intramural and extramural publications retained in the USGS IPDS are identified as permanent records in the USGS NARA Records schedule. The text and associated data for USGS series publications are currently stored in nonproprietary formats requiring no specialized software for access.

By the beginning of FY 2017, the USGS will provide all intramural and extramural publications in nonproprietary, open, 508 compliant formats, including archival JATS XML. Bibliographic metadata and the digital contents of USGS series publications have been 508-compliant since 2001. The USGS uses a suite of tools and techniques to achieve compliance. Access to bibliographic metadata and the digital content of USGS series publications is provided using formats and digital services that are standard within the library and online access community. The USGS will continue to improve integration and interoperability across the Federal community through consultation with the scientific community and with other Federal agencies and facilitate the establishment of best practices and standards for identification, description and citation of resources, terms of use, information exchange, data integration, and trusted repositories.

9.0 Objectives for Maximizing Public Access to Digital Scientific Data

The USGS has, since its creation in 1879, made the data it collects publically accessible. The USGS provides online public access to the vast majority of data produced by its research and monitoring/observation programs through digital downloads and more robust digital service capabilities. Some data, such as records collected by the national streamgage network, are provided in near real-time. All USGS-created or USGS-funded data is publically accessible and free except for cases where restrictions must be applied because of security, privacy, confidentiality, and other legal constraints

USGS will use its Science Data Catalog to facilitate discovery of and provide access to digital data resulting from its funded research. With the creation of the USGS Science Data Catalog in 2013, USGS data are made available free in compliance with Federal Open Data standards, policies and practices (as described in Obama, 2013 and Burwell et al., 2013). As a result, the ability of the public to locate and access USGS data has been improved greatly. USGS data collected under confidentiality or proprietary agreements are managed separately. Classified data are not released to the public. All released data are compliant with appropriate safeguards under the Federal Information Security Management Act, the Privacy Act, and other authorities and thereby protect confidentiality and personal privacy, recognize proprietary interests, business confidential information, and intellectual property rights and thus avoid significant negative impact on intellectual property rights, innovation, and U.S. competitiveness. In addition, every reasonable attempt is made to assure that released data are stored in a way that supports

downstream information processing and dissemination activities by using machine-readable and open formats, data standards, and complete metadata.

The USGS will continue to work extensively with the private sector to develop and employ innovative ways of distributing data such as Landsat imagery or the GeoPDF output from the [National Map](#) via cloud services such as Google Earth Engine. The USGS remains responsible as a primary point of interface with its data, but works with many partners to ensure broad distribution.

9.1 Data management plans

USGS Fundamental Science Practice requirements for data acquisition and management include a requirement for data management plans as part of all research conducted or funded by the USGS. Prior to initiating research, intra- or extramural, approved plans must identify appropriate methods for digital data management, data release, and appropriate preservation in accordance with the USGS Records Disposition Schedules. The plans must also address making data available in appropriate long-term repositories (refer to section 8.1.6) and stress the importance of non-proprietary, open formats for improved accessibility. Currently, USGS-funded data approved for release require, in all cases without exception, preservation and access. There are no exceptions.

To further clarify the obligations of extramural researchers, proposed language changes to USGS extramural funding agreements will be introduced by the effective date of this plan to require data management plans for all USGS funded research. Both intramural and extramural data management plans will describe how to maximize access while protecting privacy, confidential, and proprietary and intellectual property rights, and will balance value and cost of long-term preservation and access. This proposed language in extramural funding agreements related to required data management plans will be similar to that used by NSF in its grants process. Existing funding agreement language is elucidated below in section 10 Compliance, which requires a copy of any publication and data be provided to the USGS.

9.2 Data management costs

USGS Fundamental Science Practice policies on data management also require estimation and inclusion of appropriate data management and release costs in the plans and clarification of management functions and responsibilities at various levels in the organization across the variety of data types and repositories. The USGS will continue these practices after the effective date of this plan.

9.3 Evaluation of data management plans

USGS FSP policy also describes the processes for evaluating data management plans in the overall research review process. These processes are elucidated at the [USGS Data Management Website](#) where detailed guidance is provided in the form of explanatory text and checklists to ensure appropriate evaluation of the merits of submitted data management plans by research proposal reviewers.

9.4 Data management compliance

Current USGS FSP policies ensure compliance by USGS intramural, staff scientists with their approved data management plans. Awardees and other federally funded researchers will be informed of their obligations to meet similar data management requirements during the planning process and as part of the award language (refer to 9.1.2 above). Compliance for awardees and other federal funding researchers is and will continue to be ensured through progress reporting as required in the funding agreement and the Financial Assistance Monitoring Protocol used by USGS pursuant to 2 CFR 200.205(c)(3). Starting in January 2016, USGS will be required to report such recipients to Federal Awardee Performance and Integrity Information System (FAPIIS) as required by 2 CFR 200.212. Funds are withheld if an awardee is in noncompliance.

9.5 Data repositories and databases

The USGS has a long history of making data publicly accessible and usable by building and releasing integrated databases and robust data systems that promote the deposit of data in publicly accessible databases. Examples range from the longstanding access to robust data on the water resources of the Nation via the [National Water Information System](#) (NWIS) to integrated databases that distribute value-added data products such as BISON ([Biodiversity Information Serving Our Nation](#)) that integrates millions of species occurrence records from many discrete datasets. The developers of these types of systems have or are establishing clear policies and workflows for continuous contributions of appropriate new data. Work is ongoing to provide robust digital data repository services at different levels of the organization, from ScienceBase as a Bureau-wide capability to Program-specific repositories, with accompanying policies and procedures to ensure long-term public availability. All intramural data acquired with USGS funds is stored and managed in one of the USGS trusted repositories or dark archives. All data acquired by USGS funded extramural scientist will be stored and managed in accordance with the data management plan that is part of the research proposal. In instances where the USGS collects digital data resulting from the supported research, additional requirements for data management may be necessary to ensure compliance with the requirements of the Open Data Policy.

9.6 Public-private collaboration

The USGS actively participates and will continue to participate in collaborative forums with the private sector focusing on improving discovery, access, and use of USGS and other Federal data assets. Some specific partnerships include the [Federation of Earth Science Information Partners \(ESIP\)](#), the National Science Foundation [EarthCube](#) initiative, and the USGS [Community for Data Integration \(CDI\)](#). These opportunities have resulted in creative problem solving to make USGS data available in more relevant and viable ways. As a DOI assigning authority, the USGS collaborates with both CrossRef.org and DataCite.org for issuing its own DOIs for publications and data.

The USGS is actively examining shared cost models, budget initiatives, new partnerships, and other mechanisms to ensure adequate funding for data infrastructure. The USGS is working to identify and implement technical and management efficiencies in its cyberinfrastructure across underlying systems – Publications Warehouse, Science Data Catalog, ScienceBase, and other Program-level data facilities – to help in applying more resources to data and information management and curation. However, efficiency gains alone will not be sufficient to

accommodate the many new data resources that have not been previously managed and distributed to the level and in ways needed to support public access, nor will they be sufficient to adequately support new methods for managing and releasing USGS Series Reports. New approaches will be needed.

The USGS will continue, as it has always done, to identify additional approaches involving both public and private sector entities to expand and improve public access to its science information and data. Using existing systems and relationships with other Departments and Agencies as models, the USGS will explore the development of a research data commons, a federated system of research databases for storage, discoverability, and reuse of data, with a particular focus on making the data underlying the conclusions of peer-reviewed scientific publications resulting from federally funded scientific research available for free at the time of publication

9.7 Attribution in scientific datasets

The USGS continues to invest in new data documentation methodologies and formats that enhance the usability of metadata. Activities range from detailed implementation, planning and guidance for the latest international standard for geographic metadata under the work of the [Federal Geographic Data Committee](#) to participation in the formation of embedded metadata within the Hierarchical Data Format—a method whereby detailed attribution is embedded directly with data.

To provide appropriate attribution for scientific datasets, data are linked to a corresponding publication by reference to the publication's CrossRef.org DOI in the FGDC-complaint metadata for the dataset. In return, publications are linked to the data by their DataCite.org DOI, which is displayed on the Publications Warehouse landing page for the publication, and in the publication's citations. This linkage is illustrated in figure 1.

9.8 Training, education, and workforce development

Currently the publicly available [USGS Data Management](#) Website provides extensive training and educational materials related to data management, analysis, storage, preservation, and stewardship to USGS scientists and outside funded investigators, as well as other interested parties. The Fundamental Science Practices internal Website provides further guidance and training specific to USGS scientists. In addition, the USGS relies heavily on informal networks throughout the organization where scientific best practices are generally tested and developed such as the USGS CDI. Through community of practice activities including the USGS CDI and the Federation for Earth Science Information Partners, along with several formalized training programs, the USGS is committed to developing a workforce knowledgeable in the application of data management policies and practices.

9.9 Data preservation

Current [USGS policy](#) addresses digital data preservation. The USGS also conducts a [National Geological and Geophysical Data Preservation Program](#) in collaboration with State Geological Surveys and other DOI Bureaus for physical geoscience data and materials (e.g., rock cores, samples, paper records, etc.) as well as other data preservation activities for specific collections and data assets. Given the long history of the USGS, there is a wide array of legacy data assets, such as the records of field expeditions that contributed much of what is known of the Nation's geologic resources, which could be made more accessible in digital forms. Programs that

conduct data rescue and preservation activities constantly weigh the relative costs and benefits of preserving these resources on the basis of current research needs and public demand for them. Metadata record development resulting from inventories of physical (not digitally accessible) artifacts will make more assets discoverable in the USGS Science Data Catalog and will help set priorities for digitization and other preservation actions.

10.0 Metrics, Compliance and Evaluation

Increasing access to scholarly publications and data is a responsibility shared by all USGS employees and as stated in the current USGS FSP policy. It is the responsibility of authors and data producers under these policies to ensure that the products they develop are valid and robust. It is the responsibility of the Bureau as a whole to establish and enforce procedures that ensure scientific quality and integrity of the Bureau's products as indemnified in the policies established in the USGS SM. Requirements for these actions are embodied in the SM chapters and instructional memoranda on Fundamental Science Practices. They constitute a collection of related policies applicable to all and consequences for failure to abide by policy are outlined in detail in these policies. It is the responsibility of USGS technology professionals to continually enhance the discoverability, accessibility, and usability of the Bureau's scholarly publications and data. Compliance to these requirements can be measured and evaluated for continuous improvement.

All Federal financial assistance awards must adhere to the OMB Administrative Requirements and Cost Principles as identified in *OMB 12/26/2013 Omni Circular*. These Federal regulations are currently incorporated by reference in all USGS awards, which also require extramural funded researchers to provide copies of their related publications and data to the agency. Compliance with this requirement will be overseen by the USGS Office of Acquisition and Grants.

In addition, compliance is ensured through progress reporting as required in the funding agreement and the Financial Assistance Monitoring Protocol used by USGS pursuant to 2 CFR 200.205(c)(3). Starting in January 2016, USGS will be required to report such recipients to Federal Awardee Performance and Integrity Information System (FAPIIS) as required by 2 CFR 200.212. Funds are withheld if an awardee is in noncompliance.

11.0 Public Consultation Experience

USGS co-sponsored and attended planning meetings that were open to the public and hosted at the National Academies May 14-17, 2013, entitled "Public Access to Federally-Supported Research and Development Data and Publications." Stakeholders at these meetings included libraries, publishers, users of federally-funded research results, civil society groups, federally funded researchers, and universities.

12.0 Interagency Coordination

The USGS is an active participant in OSTP-sponsored interagency working groups on making Federal publications and data more widely accessible to the public. This includes interagency coordination with the Federation of Earth Science Information Partners (ESIP), the National Science Foundation EarthCube initiative, and the USGS Community for Data Integration (CDI). The USGS also conducts a National Geological and Geophysical Data Preservation Program in

collaboration with State Geological Surveys and other DOI Bureaus for physical geoscience data and materials (e.g., rock cores, samples, paper records, etc.) as well as other data preservation activities for specific collections and data assets.

The USGS also participates in the *Big Earth Data Initiative* (BEDI), a collaborative effort coordinated by OSTP and the U.S. Group on Earth Observations that has been pursued across the agencies operating Earth observation sensors and surveys since FY 2014. The initiative is designed to improve discoverability, accessibility, and usability for earth observation data systems across the civilian agencies. BEDI offers some opportunities to potentially advance the overall response to the public access directives. A potential partnership between ScienceBase and the GeoPlatform holds promise to help bring together important aspects of data management and distribution technologies.

13.0 Public Notice

The USGS does not intend to publish the plan for comment in the Federal Register because the USGS is already providing public access as described above in this plan. This plan will be posted publicly on the USGS Office of Science Quality and Integrity Web page and a link to this plan will be included on the DOI Open Government Web page.

14.0 Update and Re-evaluation of the Plan

This plan will be evaluated every three years to determine progress and to update as necessary. USGS will provide updates on the implementation of this plan to OSTP and OMB on January 1 and July 1 for two years after the effective date of the agency's final plan.

15.0 Timeline for Implementation

All policies, practices, and technologies necessary to meet the objectives outlined in this plan are in place now. Most requirements identified in this plan are currently in place, those requirements identified in this plan as not currently in place will be operational by the effective date. The effective date for full implementation of this plan is October 1, 2016. FY 2017 and beyond will comprise monitoring, and improvement activities.

16.0 Resources

The OSTP memorandum calls for the "identification of resources within the existing agency budget to implement the plan". Both budget and personnel resources will be needed to implement this USGS Plan, and the lack of identified resources is a critical risk. The long-established tradition of public access to USGS scholarly publications and data as well as the activities underway to improve policies, practices, and technologies will support successful implementation of this Plan; however, significant Bureau-wide resources will be necessary to support the increased number of datasets being placed in Bureau repositories. Many elements of this plan are in place, operational, and using existing funding, while other aspects of this plan, such as expansion of digital archives and repositories, modification of IPDS input and output mechanisms and workflows, will require additional resources beyond those currently committed in related budgets. Overall funding needs and priorities for implementation will be reflected in future budget submissions.

17.0 Risks

The USGS identifies several risks that may hinder execution of this plan:

- Lack of identified resources or lack of cohesion and coordination of resources across organizational units within the USGS.
- Lack of established, sanctioned, adequately funded, and fully curated, trusted Bureau-level data repositories.

18.0 References Cited

Burwell, S.M., Mancini, D.J., Park, T., and VanRoekel, S., 2013

Open Data Policy—Managing Information as an Asset: Executive Office of the President, Office of Management and Budget, Memorandum M-13-13 (May 9, 2013), 12 p., accessed October 1, 2014 at <http://www.whitehouse.gov/sites/default/files/omb/memoranda/2013/m-13-13.pdf>.

Faundeen, J.L., Burley, T.E., Carlino, J.A., Govoni, D.L., Henkel, H.S., Holl, S.L., Hutchison, V.B., Martin, Elizabeth, Montgomery, E.T., Ladino, C.C., Tessler, Steven, 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., <http://dx.doi.org/10.3133/ofr20131265>.

Holdren, J.P., 2013

Memorandum for the heads of executive departments and agencies—Increasing Access to the Results of Federally Funded Scientific Research: Executive Office of the President, Office of Science and Technology Policy (February 22, 2013), 6 p., accessed September 25, 2014, at http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf.

Obama, Barack, 2013

Executive Order No. 13642, Making Open and Machine Readable the New Default for Government Information: Federal Register Vol. 48, No. 93, pg. 28111 (May 9, 2013), accessed October 1, 2014 at <https://www.federalregister.gov/articles/2013/05/14/2013-11533/making-open-and-machine-readable-the-new-default-for-government-information>.

Appendix A -- Summary of USGS Policies Relating to Public Access and Open Data

Policy	SM No.	Purpose & Scope
Fundamental Science Practices: Foundation Policy	SM 502.1	<p>"The scientific reputation of the U.S. Geological Survey (USGS) for excellence, integrity, and objectivity is one of the Bureau's most important assets. This reputation for reliable science brings authority to data and findings, creates and protects long-term credibility, and ensures that the public trust is met. This chapter provides the foundation for a set of fundamental principles that are detailed in accompanying policy chapters. These principles collectively are USGS Fundamental Science Practices (FSP) that underlie USGS science activities, uphold the Bureau's scientific reputation, and underscore its mandate to provide reliable science to address pressing societal issues. The FSP also promote the broad release and communication of results of USGS science activities in information products (SM 1100.1). This policy applies to all employees and science activities funded by the USGS."</p>
Fundamental Science Practices: Planning and Conducting Data Collection and Research	SM 502.2	<p>"The USGS has a legacy as the producer of long-term datasets for multiple uses, many of which are geographically extensive. Part of the value of these datasets is dependent on USGS scientists describing and documenting the methods used to collect data and making these data accessible in information products. Proper documentation (including appropriate metadata) and broad dissemination ensure that USGS data and research can be interpreted appropriately, meet the highest scientific standards of excellence, and can be used broadly by the scientific community. This chapter updates the Fundamental Science Practices (FSP) policy for planning and conducting data collection and research to ensure that scientific goals are achievable and are appropriate to the mission of the USGS; the proposed methods have a reasonable likelihood of achieving the desired results; and where appropriate, methods conform to accepted standards and procedures."</p>
Fundamental Science Practices: Peer Review	SM 502.3	<p>"Peer review, as a cornerstone of scientific practice, validates and ensures the quality of published USGS science. This policy updates the Fundamental Science Practices (FSP) requirements for peer review of USGS information products and applies to all USGS scientific and technical information, whether published by the USGS or an outside entity."</p>

Policy	SM No.	Purpose & Scope
<p>Fundamental Science Practices: Review, Approval, and Release of Information Products</p>	<p>SM 502.4</p>	<p>"As a public agency, the U.S. Geological Survey (USGS) has a responsibility to make the results of its scientific investigations widely available to the public in the form of timely, technically sound, and professionally presented information products. USGS scientists are encouraged to publish their data and findings in ways that contribute to the most effective release of USGS science and best enhance the Bureau's reputation for reliable science. This chapter provides the requirements and responsibilities for the appropriate review and approval of information products prior to release. This policy applies to all Bureau science information products (SM 1100.1), whether they are published by the USGS or an outside entity. The following information products are excluded from this policy: news releases, letters to the editor (not to scientific journals), and opinion pieces or op-eds; poster sessions and presentation materials used as single-use representation of USGS work at scientific meetings, briefings, conferences, and hearing testimony; real-time hazards data and alerts; and satellite data."</p>
<p>Fundamental Science Practices: Safeguarding Unpublished U.S. Geological Survey Data, Information, and Associated Scientific Materials</p>	<p>SM 502.5</p>	<p>"This chapter provides the requirements for safeguarding unpublished U.S. Geological Survey (USGS) science data and information, including unpublished deliberative and predecisional information, proprietary data and information, nonproprietary USGS data and information, and associated scientific materials (for example, physical samples). The term "unpublished" in this chapter refers to draft, interim, or background information and materials developed or collected and used to finalize USGS information products for approval and release."</p>
<p>Fundamental Science Practices: Scientific Data Management Foundation</p>	<p>502.6 (IM OSQI 2015-01)</p>	<p>"This Instructional Memorandum (IM) provides interim policy for establishing a U.S. Geological Survey (USGS) data management foundation following a data lifecycle model."</p>
<p>Fundamental Science Practices: Metadata for Scientific Data, Software, and Other Information Products</p>	<p>502.7 (IM OSQI 2015-02)</p>	<p>"This Instructional Memorandum (IM) provides interim policy on metadata requirements for U.S. Geological Survey (USGS) scientific data, software, and other information products. It also provides guidance for complying with appropriate USGS and other Federal standards, such as the metadata standards endorsed by the Federal Geographic Data Committee (FGDC), the interagency committee that provides metadata guidance for all Federal Government scientific data (geospatial and non-geospatial)."</p>
<p>Fundamental Science Practices: Review and Approval of Scientific Data for Release</p>	<p>502.8 (IM OSQI 2015-03)</p>	<p>"This Instructional Memorandum (IM) provides interim requirements and procedures for review and approval of scientific data prior to release or dissemination. The IM applies to all U.S. Geological Survey (USGS) scientific data that are released to the public (including, but not limited to, data that are made available in datasets, databases, data services, and publications, as well as model outputs</p>

Policy	SM No.	Purpose & Scope
		and derived products."
Fundamental Science Practices: Preservation Requirements for Digital Scientific Data	502.9 (IM OSQI 2015-04)	"This Instructional Memorandum (IM) specifies preservation requirements that apply to all U.S. Geological Survey (USGS) digital scientific data and associated information."
Authority to Approve Information Products	SM 205.18	"This chapter updates the delegations of authority to approve various U.S. Geological Survey (USGS) science and other information products for release (refer to SM 1100.1 and SM 500.5)."
Publishing: Information Product Planning	SM 1100.1	"This chapter establishes policy for information product planning. Each information producer of the U.S. Geological Survey (USGS) must develop and adhere to information product plans. The topics to be addressed in a typical plan are described in general terms. This policy applies to all USGS Disciplines, Offices, and activities and all scientific, educational, and outreach information products of the USGS, with the following exclusions: immediate-turnaround, time-sensitive products (e.g., news releases, letters to the editor, and op-eds) and brief or single-use representations of USGS work (e.g., poster session and informal meeting materials). This policy is to be referred to when planning, reviewing, and executing each information product."
Publishing: Authorship, Acknowledgments, and Credits in USSG Information Products	SM 1100.5	"This policy is issued for guidance to U.S. Geological Survey (USGS) authors, including employees, contractors, cooperators, and volunteers and scientists emeriti who are preparing information products (see SM 1100.1) to be published either by the USGS or by outside organizations or journals."
Scientific Integrity	SM 500.25	"This chapter establishes USGS policy for ensuring scientific integrity in the conduct of scientific activities and procedures for reporting, investigating, and adjudicating allegations of scientific misconduct by USGS employees and volunteers. Volunteers include all scientists working under Scientist Emeritus agreements."
Integrity of Scientific and Scholarly Activities	305 DM 3	"This chapter establishes Departmental policy on the integrity of scientific and scholarly activities the Department conducts and science and scholarship it uses to inform management and public policy decisions. Scientific and scholarly information considered in Departmental decision making must be robust, of the highest quality, and the result of as rigorous scientific and scholarly processes as can be achieved. Most importantly, it must be trustworthy. It is essential that the Department establish and maintain integrity in its scientific and scholarly activities because information from such activities is a critical factor that informs decision making on public policies. Other factors that inform decision making may include economic, budget, institutional, social, cultural, legal and environmental considerations."

Appendix B -- List of Acronyms

AM	Accepted Manuscript
CDI	USGS Community for Data Integration
CSDGM	Content Standard for Digital Geospatial Metadata
CHORUS	Clearinghouse for the Open Research of the United States
DMP	Data Management Plan
DOI (organization)	U.S. Department of the Interior
DOI (technology)	Digital Object Identifier
ESIP	Federation of Earth Science Information Partners
FGDC	Federal Geospatial Data Committee
FSP	Fundamental Science Practices
FY	Fiscal Year
MODS	Metadata Object Description Service
NARA	National Archives and Records Administration
IM	USGS Instructional Memorandum
IPDS	Information Product Data System
ISO	International Organization for Standardization
JATS	Journal Article Tag Suite
OAG	Office of Acquisition and Grants (USGS)
OMB	Office of Management and Budget
OSQI	Office of Science Quality and Integrity (USGS)
OSTP	Office of Science and Technology Policy
PoR	Publication of Record
PDF	Portable Document Format
R&D	Research and Development
RSS	Rich Site Summary/Really Simple Syndication
SM	U.S. Geological Survey Manual
SPN	Science Publishing Network
USGS	U.S. Geological Survey
XML	Extensible Markup Language