



ALASKA SCIENCE CENTER DATA POLICY

RESEARCH DATA MANAGEMENT PLAN GUIDANCE & FORMS FOR PRINCIPAL INVESTIGATORS

March 12, 2015

Version 2.1.0

Point of Contact: Stan Smith, Data Manager for the ASC, stansmith@usgs.gov

CONTENTS

Objective.....	4
ASC Data Policy	5
Research Data Management Planning Process	6
RDMP Workflows.....	7
1. Submit project proposal.....	8
2. Project Selected for Funding	9
3 4. Design Spreadsheet/Database	10
5. Build Database.....	11
6. Design Data Collection Form	12
7. Archive Raw Data	13
8. QC Raw Data Archive Base Data	14
9. Conclude Research Cycle.....	15
10. Package Data For Distribution.....	16
11. Publish Data.....	17
12. Maintenance of Archived & Published Data	18
Research Data Management Plan Forms.....	19
ASC Project Proposal Form.....	21
ASC Project Implementation Form.....	22
ASC Project Forms Table of Contents.....	23
ASC New Data Collection Form	25
ASC Existing Data Form	27
ASC Derived Data Product Form.....	28
ASC Custom Software Form	30
ASC Model Form.....	31
ASC Data Description Form	32
ASC Data Domain Form	33
ASC Data Process Step / Data Quality / Lineage Form	34
ASC Project Contact Form	35
Glossary	36
Archival Formats.....	39
Digital Audio	39
Digital Video	39
Digital Photographs	40
Scanned Text	40
Geospatial Formats	41
Presentation Formats.....	41

Textual Data	41
Structured Data Format	42
Email	43
References	44

OBJECTIVE

The primary objective of the Alaska Science Center Data Management Plan and Data Policy is to fulfill the requirements inherited with acceptance of federal funding and implement these requirements to benefit the Alaska Science Center (ASC), USGS, the greater scientific community, and the public.

Federal data management requirements are stated in a series of federal laws and mandates, DOI and USGS policies, and agency directives which together orchestrate the management of federal research data to ensure that the highest levels of data quality, integrity, and utility are achieved for the benefit of current and future scientists, decision-makers, and the public. USGS researchers must also follow the requirements of our Fundamental Science Practices (FSPs) that clarify how science is conducted and resulting data products are developed, reviewed, approved, and released.

At the federal level there is a growing emphasis on sharing data. In particular, a memorandum issued by the Executive Office of the President, Office of Science and Technology Policy, “Increasing Access to the Results of Federally Funded Scientific Research,” dated February 2013 sets the requirement “...that, to the greatest extent and with the fewest constraints possible and consistent with law and the objectives set out below, the direct results of federally funded scientific research are made available to and useful for the public, industry, and the scientific community.” This memorandum was followed in May of 2013 by a memorandum from the Office of Management and Budget, “Open Data Policy – Managing Information as an Asset” providing specific guidance on the management of federal data assets.

The ASC data policy statements asserted in the *ASC Data Policy* section of this document define the Center’s responsibilities in meeting the objectives outlined in the federal memorandums and are consistent with USGS FSP procedures. Other objectives targeted by the ASC Data Policy to be achieved through formalizing of data management practices include the following:

- Obtain the maximum benefit from our research investment by preparing and preserving our data to a standard sufficient for its long-term use in research
- Realize efficiencies from consistency in our methodologies for documenting, preserving, and delivering data
- Provide the researchers with an easy to follow approach to fulfill their ASC and Project data management responsibilities
- Provide automated tools and personalized support to assist the researchers in meeting their data management responsibilities
- Increase the visibility and awareness of the science being conducted at the Alaska Science Center

Data management at the ASC involves a partnership between science, information technology (IT), and data management (DM) staff. The purpose of this document, whose primary audience is the Principal Investigator and their science staff, is to make clear what data management responsibilities are incurred by the ASC, PI, IT, and DM and what information needs to be made available for a productive interaction between the responsible parties leading to the desired end result.

The cost in effort, time, resources, and money of effective data management will ultimately be justified by the value of data and products that are shared with others. The development and maintenance of accurate data, computational models, software, and documentation are complex and nontrivial matters, and the validity of the science depends upon the documented proof that the work was done professionally. All staff and others working for, or funded by, the ASC will be required to understand the policy statement listed below, adhere to these data policies, and create a Research Data Management Plan (RDMP) for each research project.

The Data Policy for ASC staff and contractors requires that research data, computational models, software and scripts, web-based tools, and information products (collectively “data”) adhere to the following policy statements:

- I. Data will comply with all Federal laws, DOI and USGS policies, directives, and standards.
- II. Data will meet the requirements of the Alaska Science Center.
- III. Data will have a designated Data Steward knowledgeable in the specific topic who is accountable and responsible for the specifications and quality of the dataset.
- IV. Data will have a designated Data Manager who is accountable and responsible for its long-term safe-keeping, description, management, maintenance, and availability.
- V. Established ASC facilities are to be used whenever possible for data archive, storage, and distribution.
- VI. Project data are developed and maintained in an ASC managed workspace where
 - a. proactive protection from loss or corruption are provided through the use of access permission, back-ups, version control, and off-site storage,
 - b. files and other documents are organized and accessible in digital or paper folders in an understandable fashion that would allow a knowledgeable peer, to
 - i. continue the work with minimal startup effort in the absence of the employee, and
 - ii. replicate the work given the same starting material.
- VII. Data are associated with up-to-date documentation, including
 - a. Project Research Data Management Plans (RDMPs) that include project identity information, contacts, ASC relevance, staff roles and responsibilities, individual files and data elements, original source material, temporal and spatial extent, quality assurance and processing performed, limitations of the data, and final disposition in an archive or data repository,
 - b. appropriate data flow diagrams and data models that conceptually or logically describe workflow, processes and data structures, and
 - c. full provenance information that gives credit to original sources, describes how data were acquired for use, and identifies authors and actions for modifications, transformations, and improvements to the data.
- VIII. Data will be finalized in a way that products are
 - a. documented using formal standardized metadata formats such as ISO 19115 or FGDC,
 - b. provided with a valid digital object identifier (doi),
 - c. undergo a review process and receive approval before archiving or release to the public,
 - d. made ready to share with other ASC staff through a designated ASC repository in a timely manner, and
 - e. shared with the public at large within 12 months of completion of the work or product.
- IX. Data made public will
 - a. use machine-readable and open formats,
 - b. be published to the finest possible level of granularity practicable and permitted by law, and
 - c. have open licenses that place no restrictions on copying, publishing, distributing, transmitting, adapting, or otherwise using the information for non-commercial or commercial purposes.
- X. ASC will maintain an inventory of data resources which will in turn update the USGS and DOI data inventories.
- XI. ASC or other USGS organization will maintain the authoritative copy of data and metadata for which it will be responsible for all maintenance updates and distribution to other repositories.

RESEARCH DATA MANAGEMENT PLANNING PROCESS

The ASC Research Data Management Plan (RDMP) documents the data requirements of a project. The RDMP collects information about data inputs, data transformation processes, custom software and scripts, project contacts, and data products. The RDMP uses a set of forms to guide the Principal Investigator (PI) through the process of gathering and documenting the required information. Together, the set of completed forms provides the information needed to prepare ISO or FGDC project and data metadata.

In practice no two projects are likely to follow the same workflow. The RDMP acknowledges the variability in project requirements and the PI's independence in determining the research approach. To accommodate this variability the RDMP Guidance is organized into discrete events which when they occur trigger some data management responsibility on the part of the PI. The RDMP forms are designed to describe single data collections and should be added to the project's RDMP as needed.

While each project is encouraged to proceed using its optimal workflow, the expectation remains that as the project concludes all objectives defined in the ASP Data Policy statements will have been met.

The following table describes in general terms what RDMP requirements at various stages of project development.

Project Stage	RDMP Action Required
Pre-proposal	None.
Proposal	Each new project proposal requires a completed Project Proposal Form accompanied by a separate New Data Collection Form , Existing Data Form , Data Product Form , Custom Software Form , and Model Form for each data collection planned for the project. Note that only the top section of these forms need to be completed for the Proposal stage.
Selected for Funding	Once the project has been selected for funding the " Project Implementation Form " should be completed. This includes meeting with your Information Technology (IT) representative to request work space on an IT managed server appropriate for your project's work and data files. A new project record should be added to the ASC Science Portal for your project containing known information and a status of "In Progress". A short consultation with the Data Manager (DM) assigned to your project is required to acquaint the DM with anticipated data documentation, archival, and deployment needs. This is also an opportunity for the PI to learn how DM may be able to help with database design, RDMP completion, data preservation, metadata preparation, and data publication.
Research	As your research proceeds the data, software, and model forms initiated in the Proposal stage should gradually be completed. This includes adding new forms for data, software, and models not anticipated during the Proposal stage and dropping forms for discarded data. The project record in the ASC Science Portal should also be periodically updated. Raw data collected will be documented and archived as it is collected.

Project Stage	RDMP Action Required
Completion	As the project nears completion the DM will use the completed RDMP forms to prepare metadata records for base data required to be archived, derived product data for internal USGS use, and data to be released outside USGS. Document Object Identifiers (doi) will be obtained for data products selected for publication. The project record in the ASC Science Portal should be updated. DM will work with the PI to publish data products through the ASC Portal and distribute data and metadata to additional repositories as required by the USGS FSP.

RDMP WORKFLOWS

The project workflow described in this section is intended to lead a researcher through the steps necessary to fulfill the Alaska Science Center data management requirements. Note that only steps that are the responsibility of the PI are identified in the workflow. Additional actions taken by members of the Data Management staff are not presented in this document, such as handling the long-term preservation, exposure, and delivery of data.

The workflow is expressed as an event model. These events do not always occur in the same order and may occur multiple times or not at all within a particular project cycle. When one of the following events does occur, the PI should proceed to the corresponding section of the workflow to learn what data management actions he/she should take.

Project events requiring data management action:

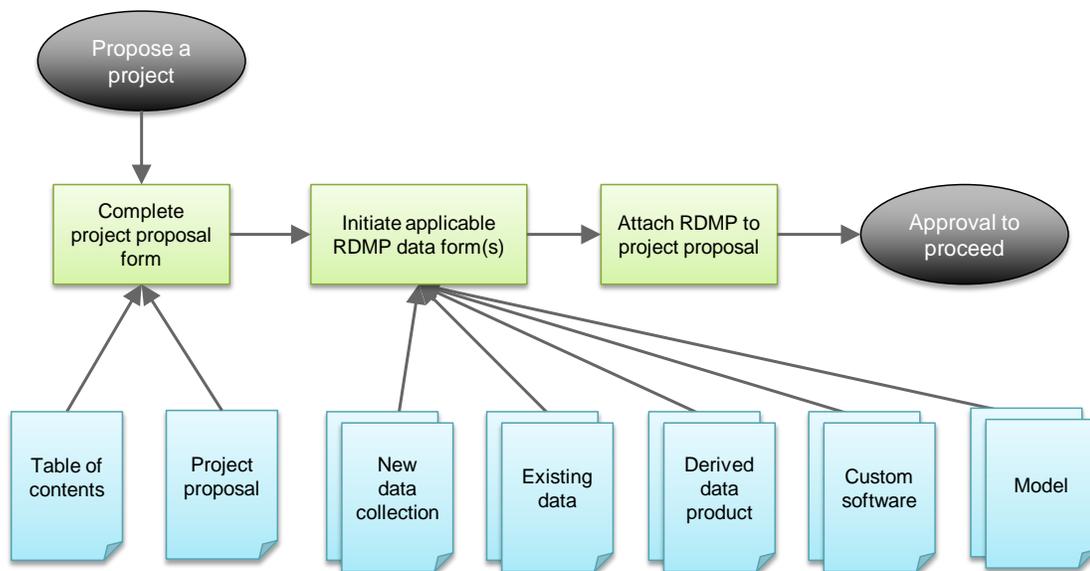
1. [Submit project proposal](#)
2. [Project selected for funding](#)
3. [Design spreadsheet](#)
4. [Design database](#)
5. [Build database](#)
6. [Design data collection form](#)
7. [Archive raw data](#)
8. [QC raw data | archive base data](#)
9. [Conclude research cycle](#)
10. [Package data for distribution](#)
11. [Publish data](#)
12. [Maintenance of archived & published data](#)

The PI is encouraged to consult with or request assistance from DM staff members as needed to interpret or complete any data management action related to the above events or to complete RDMP forms.

1. SUBMIT PROJECT PROPOSAL

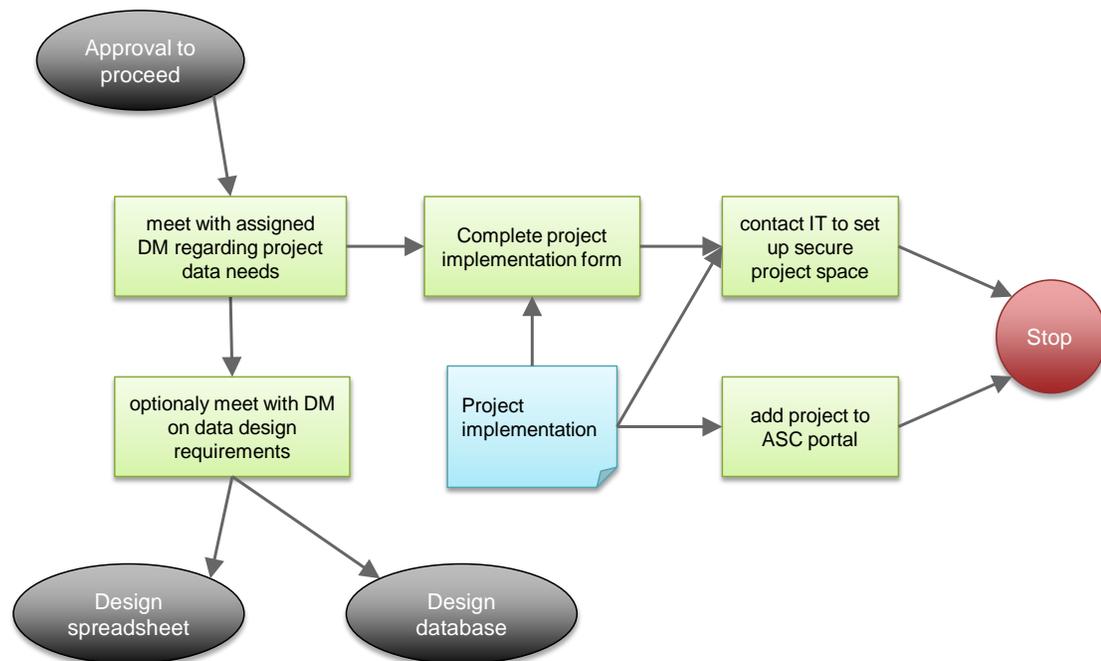
Each new project proposal is required to attach a brief Research Data Management Plan (RDMP). The proposal-level RDMP outlines the project's anticipated data requirements and serves as a skeleton for full a RDMP should the project be funded. Details will be added to these RDMP forms throughout the research cycle provided the project is approved.

- a. Complete project proposal form: The proposal-level RDMP requires a completed [Project Proposal Form](#) accompanied by a [Table of Contents](#).
- b. Initiate data forms: Separate [New Data Collection](#), [Existing Data](#), [Derived Data Product](#), [Custom Software](#), and [Model](#) planned for the project. Note that data, software, and model forms are divided into sections corresponding to phases of the research project. Only the top section of each form needs to be completed for the project proposal
- c. Attach RDMP to project proposal: Attach the RDMP forms to the project proposal for submission to management for consideration.



2. PROJECT SELECTED FOR FUNDING

- a. Consultation with Data Manager: A short consultation with the Data Manager (DM) assigned to your project is required to acquaint the DM with anticipated data documentation, archival, and deployment needs. This is also an opportunity for the PI to learn how DM may be able to help with completion of RDMP forms, database design, data preservation, metadata preparation, and data publication. The PI and other project staff are encouraged to make use of these services whenever the need arises.
- b. Complete the Project Implementation Form: Complete the [Project Implementation](#) form.
- c. ASC Science Portal record: Add your project to the ASC Science Portal with a status of “In Progress”. Much of the needed information can be copied from your completed Project Implementation form.
- d. Project Folder: Coordinate with your Information Technology (IT) representative to have a secure space defined on an IT managed server to hold your project files and data. IT will apply access permissions to this folder according to the schedule you documented on the Project Implementation Form.
- e. Data Design Consultation: You have the option to meet with DM to discuss data design requirements for your project dataset(s) and database(s). This is not a required step but it is an optimal time to discuss your data design options.



3 | 4. DESIGN SPREADSHEET/DATABASE

Design Spreadsheet

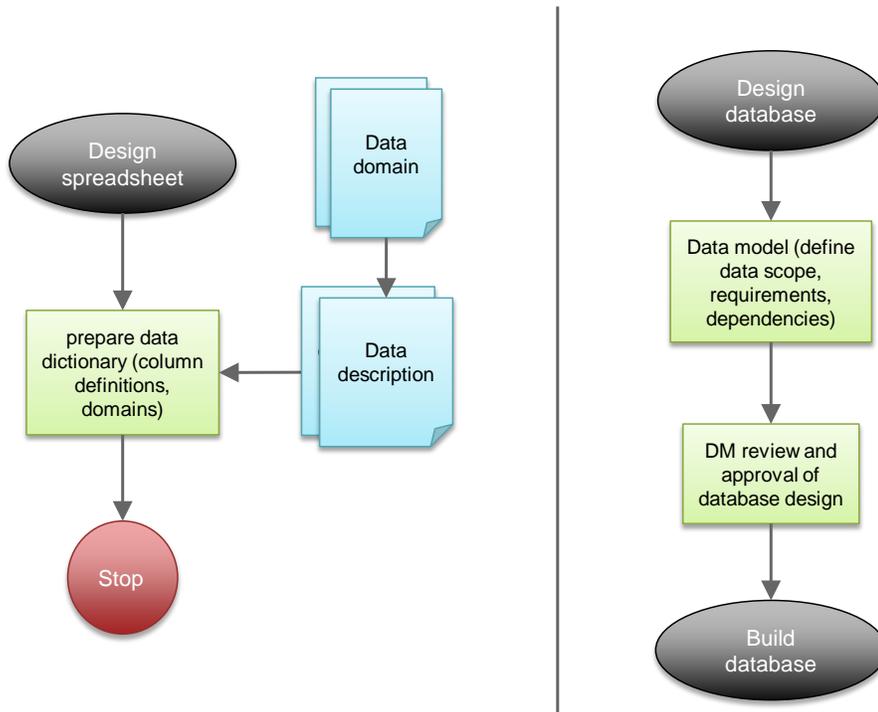
If the spreadsheet is anticipated to be archived, used in subsequent research projects, published, or become available to the public it will require proper metadata including a data dictionary. Spreadsheets that are used only as intermediate work steps do not need to be documented.

- a. **Add data dictionary:** Complete a [Data Description](#) form to your RDMP for each sheet in the spreadsheet. Remember to attach a [Data Domain](#) form for each domain used by a spreadsheet column.

Design Database

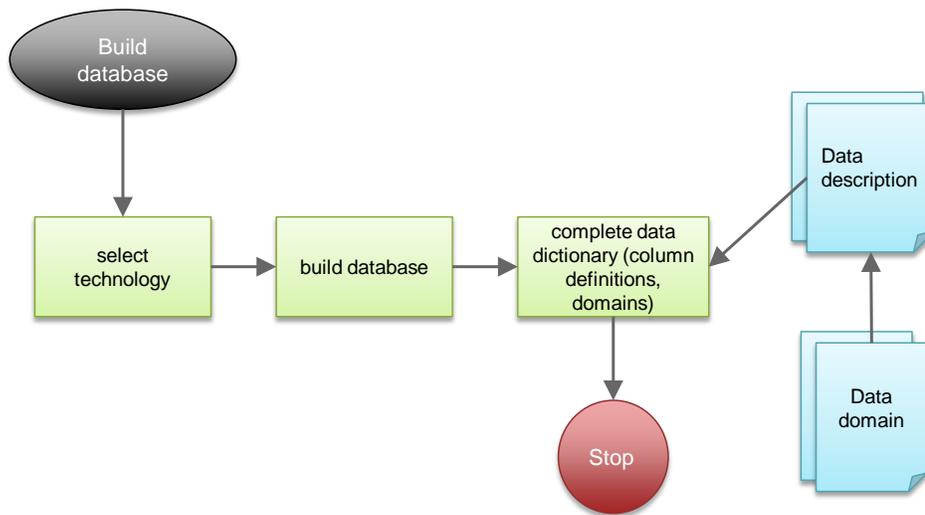
All new database designs need to be reviewed and approved by the project's Data Manager prior to building the database. It is not required that the Data Manager perform the database design work or build the database, although these options are available and encouraged. The DM can alternatively provide consultation during the database design and implementation process.

- a. **Design Database:** Define the database scope, entities and attribution, dependencies between entities, and attribute domains. Data models (entity relationship diagrams) are an ideal tool for documenting and communicating this information. Data models are encouraged for all databases and are required for large and complex databases. Your DM will help make the determination whether a data model is required.
- b. **Review & Approval:** Submit the database design to the DM assigned to the project for final review and approval. The DM may suggest some changes to improve database integrity, performance, and integration with other ASC data.



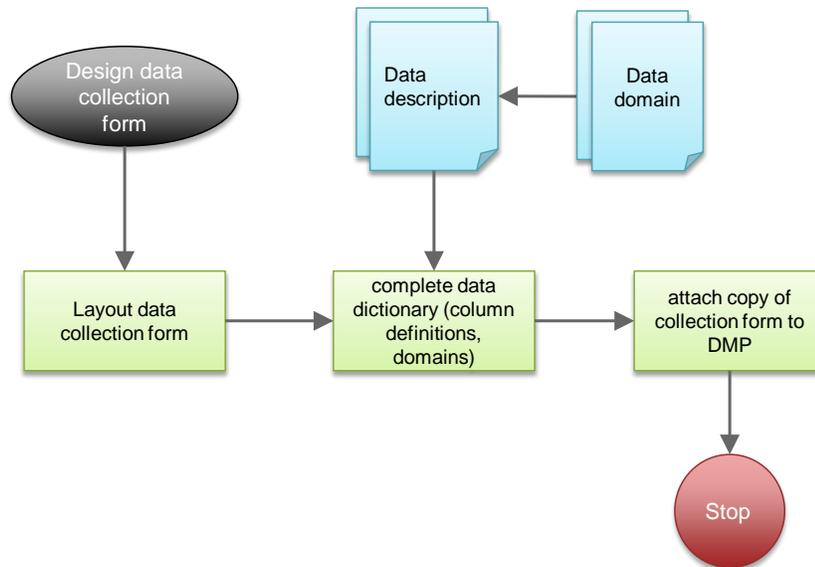
5. BUILD DATABASE

- a. **Select Database Technology:** Select the appropriate database technology for implementing the database. Selection of a technology commonly used and supported at the ASC is encouraged. Your assigned Data Manager can help you make this choice taking into consideration size, user community, features, and Project team's experience.
- b. **Build the Database:** Write the database schema definitions in the database language selected.
- c. **Document the Database:** For each table in the database, add a [Data Description Form](#) to your RDMP. Complete the form and remember to attach a separate [Data Domain Form](#) for each domain used.



6. DESIGN DATA COLLECTION FORM

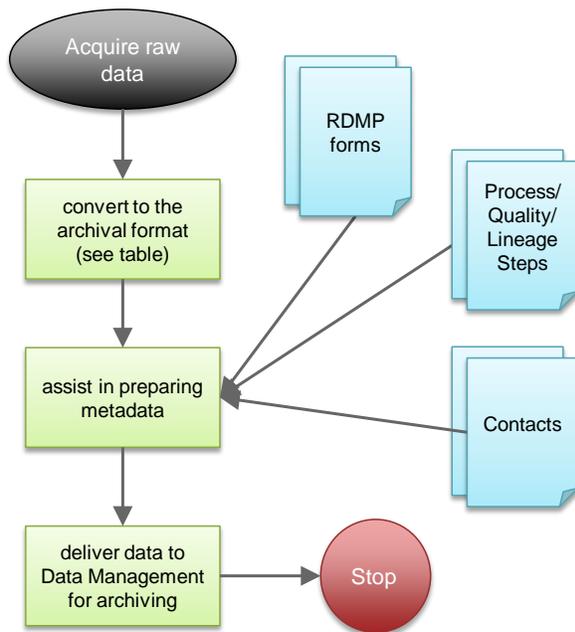
- Layout Data Collection Form:** Layout your field or lab data collection form. Remember to include specific fields to record the observer name, date, and some unique identifier or sequence number on each form.
- Document the Form:** For each data collection form, add a [Data Description Form](#) to your RDMP to define the form fields. Also complete a separate [Data Domain Form](#) for each domain used on the form field.
- Add Form to RDMP:** Add a copy of the form to your RDMP to complete the documentation.



7. ARCHIVE RAW DATA

Raw data should be archived as soon as possible after collection. This practice will preserve the raw data in its truest state and reduce the possibility of alteration or loss. Archived raw data will not be released beyond the boundaries of the immediate Project team members without authorization from the PI.

- Convert Raw Data for Archive:** Convert the raw data to the most appropriate archival format as described in the table below. If the preferred archival format is impractical or your raw data format is not represented in the table, see the appendix on [Archival Formats](#) for more information. Contact your Project’s Metadata Specialist if you need assistance.
- Prepare Metadata:** All archived raw data files will have complete metadata records. The Metadata Author should work with the Project’s Metadata Specialist to prepare the metadata for generation of an FGDC or ISO record. The project and data RDMP forms prepared earlier will be used in this step. Additional [Data Process/Quality/Lineage Steps](#) and [Contacts](#) forms may need to be added. If the necessary forms were not completed in earlier steps, that task will need to be added here.
- Archive Raw Data:** Deliver the fully documented archive package(s) to the Data Manager for archiving. Request READ access for Project team members that will need access to the archive folder.

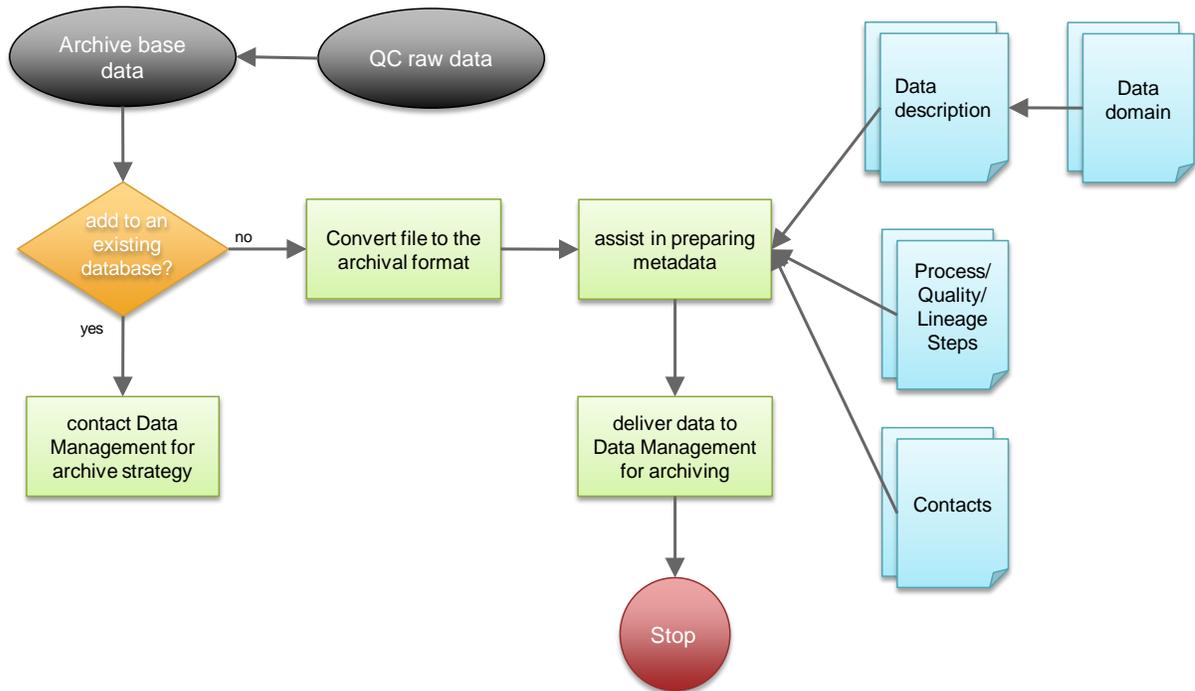


raw data format	archival format
field form, field notes, lab report, paper document	scan to PDF/A
e-document	save as PDF/A
structured data	convert to CSV, ASCII, JSON
photo	TIFF, DNG, JPEG
video	MPEG4, MOV, AVI
instrument data (binary)	process to ASCII
instrument data (ASCII)	ASCII
remote sensed data	contact Data Management

8. QC RAW DATA | ARCHIVE BASE DATA

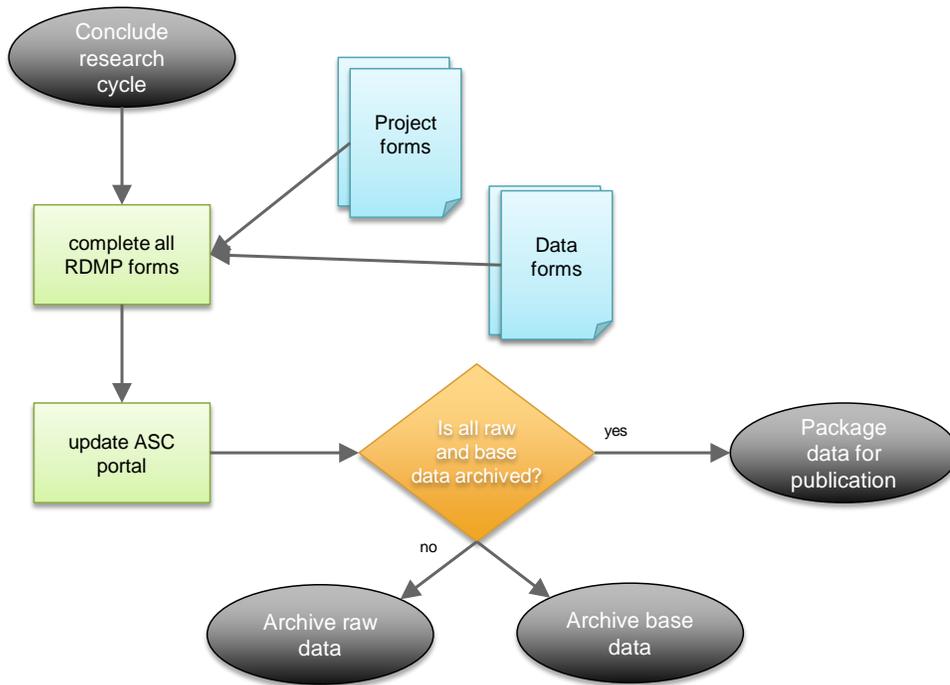
Perform quality checks (QC) of raw data. Raw data that has been QC'd and verified as ready for use in research is termed 'base' data. Base data is also archived to the same standards as raw data.

- Convert Base Data for Archive:** Convert the base data to the most appropriate archival format. See the appendix on [Archival Formats](#) for more information. Contact your Project's Data Manager if you need assistance.
- Prepare Metadata:** All base data archive files will have comprehensive metadata records. The PI should work with the Project's Metadata Specialist to prepare the metadata and archive package. The project and data RDMP forms prepared earlier will be used in this step. Additional [Data Process/Quality/Lineage Steps](#) and [Contacts](#) forms may need to be added. If the necessary forms were not completed in earlier steps, that task will need to be added here.
- Archive Base Data:** Deliver the fully documented archive package(s) to the Data Manager for archiving. Request READ access for Project team members that will need access to the archive folder. At the conclusion of the research project the folder will become accessible to all USGS researchers.
- Add Base Data to Existing Database:** If the base data is to be added to an existing database or archive, contact your Project's Data Manager for determining next steps.



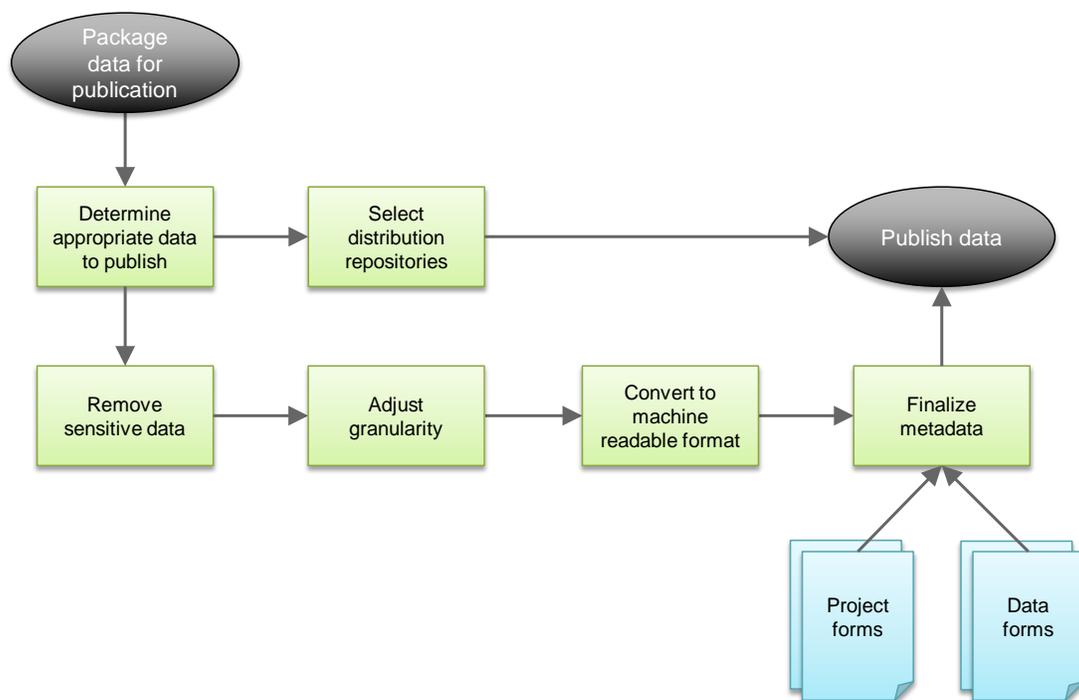
9. CONCLUDE RESEARCH CYCLE

- a. Complete RDMP Data Forms: Complete all necessary RDMP forms not completed earlier as these will be needed to create your project and data metadata records prior to publication.
- b. Update the ASC Science Portal: Complete the Project's record in the ASC Science Portal. Remember to update the record status and release the record to the public.
- c. Archive Raw and Base Data: Archive any raw and base data not already archived.



10. PACKAGE DATA FOR DISTRIBUTION

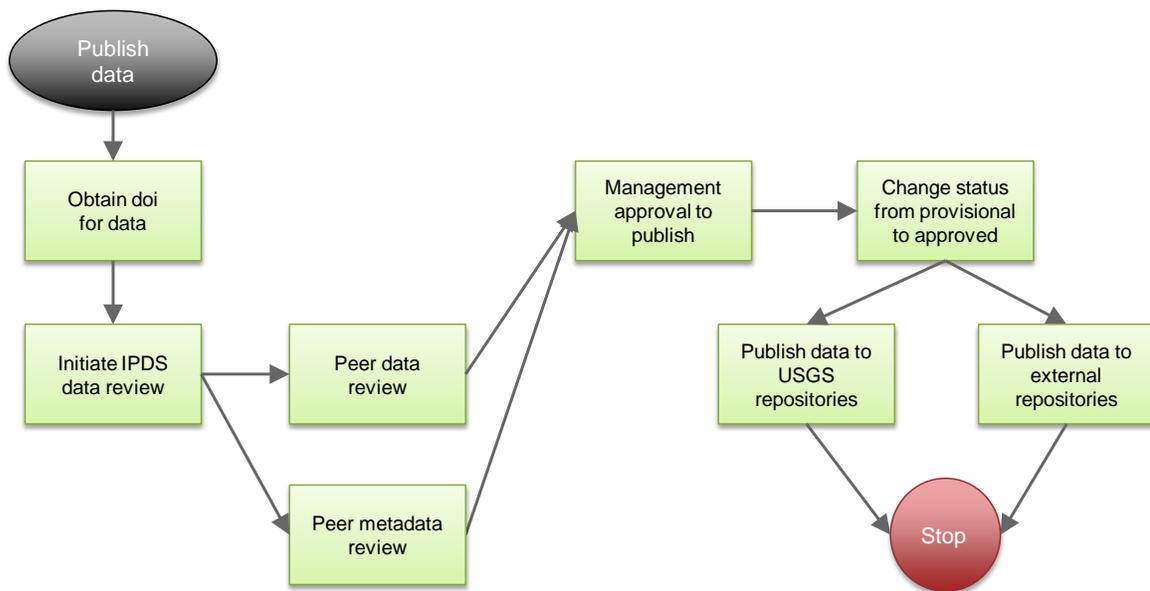
- a. Determine Appropriate Data to Publish: Determine the appropriate project data to publish. Keep in mind the federal requirement is to make all data which is useful to science or of interest to the public freely available. This requirement extends beyond the summary data generally published with a journal article.
- b. Remove Sensitive Data: Data or records inappropriate for public viewing or use should be removed from the dataset. This includes Personally Identifiable Information (PII), sensitive data regarding threatened and endangered species, data related to ongoing legal action, and data classified as a risk to national security.
- c. Adjust Granularity: The federal requirement is to publish data to ‘the maximum granularity practicable’. This may exceed the granularity referenced in the journal article. The PI should determine the maximum useful granularity to science and the Data Manager can help to identify delivery options. Together they should set the maximum practicable granularity.
- d. Convert to a Machine Readable Format: If the data is not already in a machine readable format, it should be converted to a format which is easily be consumed by computer applications. Note that not all data can be made machine readable.
- e. Finalize Metadata: Adjust or prepare metadata for the distribution package. Much of the metadata will already have been collected in previous steps. Adjust this metadata to match the final distribution package.
- f. Select Distribution Repositories: Select the repository(ies) desired for distributing this data. The ASC will be one distribution site but other thematic portals and repositories may be more appropriate for some types of data. Contact your project DM for clearance to use an external repository as the USGS Office of Science Quality and Integrity maintains a list of disapproved repositories.



11. PUBLISH DATA

All data is considered provisional (preliminary) until it has completed the peer review processes for both the data and its metadata. After a successful reviews of the data and metadata an ASC manager may change the data status from 'provisional' to 'approved' and clear the data for distribution outside USGS.

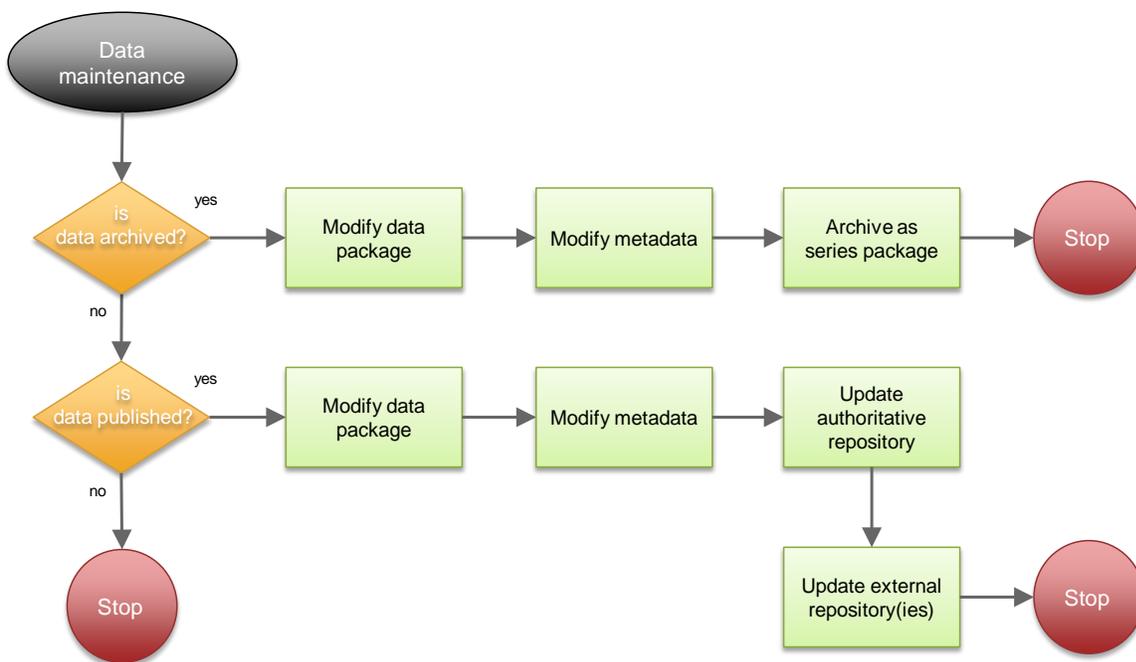
- a. Obtain a doi for the Data: All data intended for release outside of USGS must be accompanied by a digital object identifier (doi). Your DM will be able to help you obtain and register a doi for each separate data collection.
- b. Initiate IPDS for Data: Data products intended for release are to use the USGS Information Product Data System (IPDS) to track the review and approval process.
- c. Peer Data Review: Data reviews are performed to check that the data meet standards for quality and completeness and requires the reviewer to have specialized scientific knowledge of the particular type of data being reviewed. Refer to the [data checklist](#) for more detail.
- d. Peer Metadata Review: Metadata reviews are performed to check that they meet standards for quality and completeness and require the reviewer to have specialized knowledge of the metadata standard being used. Refer to the [metadata checklist](#) for more detail.
- e. Management Approval to Publish: Once the data and metadata reviews have successfully completed management approval can be requested to release the data.
- f. Change Status to Approved: After management has signed off on the reviews and approved data for release the data disclaimer in the metadata record should be changed from 'provisional' to 'approved'.
- g. Publish to USGS Repository: In most instances a USGS data repository will be the authoritative data repository. ASC will host small to medium data sets. Contact your DM to locate suitable repositories for larger data sets.
- h. Publish to External Repositories: Publishing copies of data to external thematic repositories is encouraged.



12. MAINTENANCE OF ARCHIVED & PUBLISHED DATA

Correction or modification of data previously published or archived needs to be documented and applied to the archive and distribution package. Note that newly acquired data or extensions to data do not qualify as corrections or maintenance.

- Modify the Data Package:** Migrate the corrections or modifications into the archive or distribution data package.
- Modify the Metadata:** Modify the metadata to indicate who, when, and what was modified from the original archive or distribution package.
- Archive as a Series:** The amended archive package should be archived as a new archive package without disrupting the original package. The archive will contain all versions that have been archived.
- Update Authoritative Repository:** Replace the distribution package in the authoritative USGS repository.
- Update External Repositories:** To the extent possible, replace the distribution package in external repositories chosen to distribute this data.



RESEARCH DATA MANAGEMENT PLAN FORMS

This section contains templates for the project and data forms. Please copy the form tables as needed and paste them into your project's Research Data Management Plan before entering information into the form fields.

1. [ASC Project Proposal Form](#)
This form is to be completed and submitted with the project proposal. It gathers very general information about the project that will contribute to ASC Science Portal entries and project metadata and identification of responsible parties for the various data management roles. The form requires the signature of a Program Manager.
2. [ASC Project Implementation Form](#)
This form is to be completed after the project has been selected for funding. The form gathers some additional project detail for the ASC Science Portal and records the project's top folder location and access permissions. The form also requires a signature from the project's Data Manager to indicate that the PI has met with the DM to discuss anticipated data requirement for the project.
3. [ASC Project Forms Table of Contents](#)
This form should be added to the RDMP at the proposal stage. Add entries to the table of contents (TOC) for each form attached to the RDMP (other than the Project Proposal and Project Implementation forms). The TOC form also has checkboxes that provide a fast over view of the RDMP status by data collection.
4. [ASC New Data Collection Form](#)
Add this form to the RDMP for each new data collection ([see Glossary](#)) originated by the project. For the project proposal, only the top section needs to be completed. The remaining fields pertinent to the data collection should be completed during the research cycle as the information becomes known.
5. [ASC Existing Data Form](#)
Add this form to the RDMP for each existing data collection ([see Glossary](#)) used by the project. For the project proposal, only the top section needs to be completed. The remaining fields pertinent to the data collection should be completed during the research cycle as the information becomes known.
6. [ASC Derived Data Product Form](#)
Add this form to the RDMP for each derived data product ([see Glossary](#)) produced by the project. For the project proposal, only the top section needs to be completed. The remaining fields pertinent to the derived data product should be completed during the research cycle as the information becomes known.
7. [ASC Custom Software Form](#)
Add this form to the RDMP for each non-trivial custom software application, script, and web tool developed by the project as a product or that is needed to fully understand the scientific process applied. For the project proposal, only the top section needs to be completed. The remaining fields pertinent to the custom software should be completed during the research cycle as the information becomes known.
8. [ASC Model Form](#)

Add this form to the RDMP for each computational model developed by the project. For the project proposal, only the top section needs to be completed. The remaining fields pertinent to the model should be completed during the research cycle as the information becomes known.

9. [ASC Data Description Form](#)

Add this form to the RDMP for each sheet in a spreadsheet, table in the database, and field collection form which is to be archived, distributed, or otherwise requires metadata prepared. Add a row to the Data Description Form for each attribute of the sheet, table, or form. Note that not all fields will be appropriate for each attribute. Fill in Name and Definition for each attribute and complete the other fields as appropriate.

For attributes that take their values from domains complete an ASC Data Domain Form, attach it to the RDMP, and place the domain form's ID on the Data Description Form.

For attributes that received quality control checks complete an ASC Data Process Steps / Data Quality / Lineage Form, attach it to the RDMP, and place the process steps form's ID on the Data Description Form.

10. [ASC Data Domain Form](#)

Add this form to the RDMP for each data domain used. Use the domain's form ID on Data Description Forms for each attribute associated with the domain. Note that once a domain has been documented it may be referenced by multiple other forms.

11. [ASC Data Process Steps / Data Quality / Lineage Form](#)

Add this form to the RDMP to describe processing steps for New Data Collection, Existing Data, and Data Product preparation or Quality Control steps for Data Description forms. Only describe the non-trivial steps and keep the definitions brief, these descriptions are for metadata.

12. [ASC Project Contact Form](#)

Add this form to the RDMP for each contact. The contact may be an individual or an organization. Place the contact form's ID in the appropriate field for the form requiring contact information. Note that once a contact has been documented on a contact form it may be referenced by multiple other forms.

ASC PROJECT PROPOSAL FORM

This form is to be completed during the project proposal phase.

PROJECT [10T]	
Title:	<i>Enter full project name</i>
Funding Source:	USGS Program: Major Initiative:
Financial System IDs:	FBMS: 10T Basins+: 10T
Abstract:	10T 10T
Project Start Date:	10T
Project End Date:	10T
Spatial Description:	10T
Temporal Description:	10T
Responsible parties:	Program Manager: 10T Principal Investigator(s): 10T Researcher(s): 10T Data Steward: 10T Data Manager: Metadata Author: 10T Metadata Specialist: Web Developer: 10T
Hardware environment:	10T
Software environment:	10T
ASC manager sign-off:	Program Manager signature/date:

ASC PROJECT IMPLEMENTATION FORM

This form is to be completed during early implementation of the project.

DM consultation sign-off:	Data Manager signature/date:
Short abstract for web:	10T
Project location:	Map point: 10T 10T Bounding box: 10T 10T 10T 10T Boundary polygon: 10T Common Place Names: 10T
Keywords:	10T
ASC Portal information:	Project ID: 10T Contact ID: 10T
Project folder:	10T 10T Permissions: Management: write Project team: Data Management: USGS ASC: USGS all: Public:
Responsible parties:	Funding Organization(s): 10T Point of Contact(s): 10T Principal Investigator(s): 10T Researcher(s): 10T
Other contributors:	10T

ASC PROJECT FORMS TABLE OF CONTENTS

Use this form is to list the ASC Data Management Forms attached to your project. Duplicate a section for each additional form as needed.

New Data Collections:	Form Name: 10T <input type="checkbox"/> field form design <input type="checkbox"/> raw metadata prep <input type="checkbox"/> base metadata prep	ID: 10T <input type="checkbox"/> data collected <input type="checkbox"/> archive raw data <input type="checkbox"/> archive base data	<input type="checkbox"/> RDMP form complete <input type="checkbox"/> QC raw data <input type="checkbox"/> Publish
Existing Data Collections:	Form Name: 10T <input type="checkbox"/> data acquired	ID: 10T <input type="checkbox"/> RDMP form complete	
Project Data Products:	Form Name: 10T <input type="checkbox"/> RDMP form complete <input type="checkbox"/> archive data <input type="checkbox"/> publish	<input type="checkbox"/> metadata prep <input type="checkbox"/> prep data for public	ID: 10T <input type="checkbox"/> data review <input type="checkbox"/> web development
Custom Software:	Form Name: 10T <input type="checkbox"/> document software	<input type="checkbox"/> RDMP form complete	ID: 10T <input type="checkbox"/> archive software
Models:	Form Name: 10T <input type="checkbox"/> archive data	<input type="checkbox"/> prep data for public	ID: 10T <input type="checkbox"/> publish
Project Contacts:	Form Name: 10T		ID: 10T
Data Description:	Form Name: 10T		ID: 10T
Data Domain:	Form Name: 10T		ID: 10T
Data Lineage:	Form Name: 10T		ID: 10T



ASC NEW DATA COLLECTION FORM

Complete this form for data that does not currently exist and will be collected or generated during the project cycle, such as a new field data collection. This form will follow new data as it matures from raw to base to publication. Remember to add the form Name: and ID: to the Project Table of Contents.

ID [10T]	
----- To be completed for project proposal -----	
Name:	10T
Description:	10T
----- To be completed for raw data -----	
Format:	10T
Data collection template:	Link to template: 10T Template Data Description Form: 10T
Data collection protocols:	10T
Spatial reference system:	10T
Spatial scope & scale:	Description: 10T Scale: 10T
Temporal scope & scale:	Description: 10T Scale: 10T
Data volume estimate:	10T
Restrictions:	Use: 10T Legal: 10T
Non-USGS Partners:	Form ID: 10T
Authoritative Repository:	10T
Contact:	Form ID: 10T
----- To be completed for base data -----	
Format:	10T
Data description:	Form ID: 10T
Data Process/CQ Steps:	Form ID: 10T
Keywords:	10T
Data volume estimate:	10T

Restrictions:	Use: 10T Legal: 10T
Contact:	Form ID: 10T
----- To be completed when data is made public -----	
Exclusive use period:	Period: Justification: 10T
Data Distribution Repository(s):	10T
Citation:	10T
Digital object identifier:	10T
Contact:	Form ID: 10T

ASC EXISTING DATA FORM

Complete this form for data acquired from existing collections that is used in project research for creating data products or project deliverables. Remember to add the form Name: and ID: to the Project Table of Contents.

ID [10T]	
----- To be completed for project proposal -----	
Name:	10T
Description:	10T
Source:	Name: 10T Link: 10T
Restrictions:	Use: 10T Legal: 10T
Agreement(s):	10T
Fees:	10T
Format:	10T
----- To be completed after project has been approved for funding -----	
Data Process Steps:	Form ID: 10T
Spatial reference system:	10T
Spatial scope & scale:	Description: 10T Scale: 10T
Temporal scope & scale:	Description: 10T Scale: 10T
Volume Estimate:	10T
Acquisition Date:	10T
Citation:	10T
Digital object identifier:	10T
Contact:	Form ID: 10T

ASC DERIVED DATA PRODUCT FORM

Complete this form for each derived data product that will be published internally (USGS – ASC) or publically. Remember to add the form Name: and ID: to the Project Table of Contents.

ID [10T]	
----- To be completed for project proposal -----	
Name:	10T
Description:	10T
Completion Date:	10T
----- To be completed after project has been approved for funding -----	
Purpose:	10T
Format:	10T
Data description:	Form ID: 10T
Spatial reference system:	10T
Spatial scope & scale:	Description: 10T Scale: 10T
Temporal scope & scale:	Description: 10T Scale: 10T
Keywords:	10T
Data Derivation Steps:	Form ID: 10T
Volume Estimate:	10T
Maintenance frequency:	
Restrictions:	Use: 10T Legal: 10T
Non-USGS Partners:	Form ID: 10T
Authoritative Repository:	10T
Contact:	Form ID: 10T
----- To be completed when data is made public -----	
Data Refinement Steps: (steps to prepare data for public distribution)	Form ID: 10T

Deployment considerations:	Distribution: 10T Web page: 10T
Exclusive use period:	Period: Justification: 10T
Data Distribution Repository(s):	10T
Citation:	10T
Digital Object Identifier:	10T
Contact:	Form ID: 10T

ASC CUSTOM SOFTWARE FORM

Describe each custom software package written for use in this project. Use a separate form for each code package. Remember to add the form Name: and ID: to the Project Table of Contents. Remember to add the form Name: and ID: to the Project Table of Contents.

ID [10T]			
----- To be completed for project proposal -----			
Name:	10T		
Description:	10T		
Project Type:	10T		
----- To be completed after project has been approved for funding -----			
Version:	Major Release: 10T	Minor Release: 10T	Revision: 10T
Purpose:			
Languages:	<input type="checkbox"/> C# <input type="checkbox"/> jQuery <input type="checkbox"/> Rails 10T	<input type="checkbox"/> Java <input type="checkbox"/> Python <input type="checkbox"/> Ruby	<input type="checkbox"/> JavaScript <input type="checkbox"/> R <input type="checkbox"/> SAS
Available Packages:	10T		
Link to Documentation:	10T		
Minimum Configuration:	10T		
Dependencies:	10T		
Software Repository:	10T 10T 10T		
Maintenance & Support:	Frequency: Support Plan: 10T		
Environment:	10T		
Digital Object Identifier:	10T		
Contact:	Form ID: 10T		

ASC MODEL FORM

Describe the function and methodology used for any models that are part of the project. Remember to add the form Name: and ID: to the Project Table of Contents. Remember to add the form Name: and ID: to the Project Table of Contents.

ID [10T]	
----- To be completed for project proposal -----	
Name:	10T
Description:	10T
----- To be completed after project has been approved for funding -----	
Model version:	10T
Source link:	10T
Model input(s):	10T
Model output(s):	10T
Spatial reference system:	10T
Spatial scope & scale:	Description: 10T Scale: 10T
Temporal scope & scale:	Description: 10T Scale: 10T
Calibration details:	10T
Contact:	Form ID: 10T

ASC DATA DESCRIPTION FORM

This form is to be completed for each sheet in a spreadsheet, table in a database, and field collection form which is to be archived, distributed or otherwise require metadata to be prepared. Copy-paste the column definition section in the form table as needed to complete a definition for each column. Remember to add the form Name: and ID: to the Project Table of Contents.

ID [10T]	
Name:	10T
Description:	10T
Location:	10T
Key Columns:	10T
Column(s): (repeat for each column – two column sections are shown)	Name: 10T Definition: 10T Datatype: 10T Units: 10T Min Value: 10T Domain ID: 10T QC Tests: 10T Use Limits: 10T Legal Const: 10T Other: 10T Max Value: 10T
	Name: 10T Definition: 10T Datatype: 10T Units: 10T Min Value: 10T Domain ID: 10T QC Tests: 10T Use Limits: 10T Legal Const: 10T Other: 10T Max Value: 10T

ASC DATA DOMAIN FORM

This form is to be completed for each data domain used in spreadsheets, data tables, and field collection forms which are to be archived, distributed or for otherwise need metadata prepared. Copy-paste the domain value section in the form table as needed to complete a definition for each domain value. Remember to add the form Name: and ID: to the Project Table of Contents.

ID [10T]	
Name:	10T
Description:	10T
Domain value(s): (repeat for each domain value – two value sections are shown)	Value: 10T Definition: 10T
	Value: 10T Definition: 10T

ASC DATA PROCESS STEP / DATA QUALITY / LINEAGE FORM

This form is to document non-trivial data processing steps for raw data collection and handling, pre-processing of existing data, and preparation of product data. Copy-paste additional process steps in the form table as needed to complete a definition of each significant process step. Remember to add the form Name: and ID: to the Project Table of Contents.

ID [10T]	
Name:	10T
Data acquisition methodology:	10T
Report conformance with process steps:	10T
Process Steps(s): (repeat for each step – two step sections are shown)	Step ID: 10T Description: 10T Contact ID: 10T
	Step ID: 10T Description: 10T Contact ID: 10T

ASC PROJECT CONTACT FORM

This form is to be completed for each project participants. Remember to add the form Name: and ID: to the Project Table of Contents.

ID [10T]	
Contact type:	
Organization role(s):	<input type="checkbox"/> funding <input type="checkbox"/> distributor <input type="checkbox"/> contributor <input type="checkbox"/> custodian <input type="checkbox"/> owner
Individual role(s):	<input type="checkbox"/> PI <input type="checkbox"/> contributor <input type="checkbox"/> custodian
Organization:	10T
Name:	First Name: 10T Middle Name: 10T Last Name: 10T Preferred Name: 10T
Title:	10T
Address (mailing):	Address Line 1: 10T Address Line 2: 10T City: 10T, State: 10T, Zip: 10T
Phone(s):	Office: 10T Mobile: 10T SMS: 10T
eMail address(s):	10T 10T
Web link(s):	10T 10T

- *base data*

'base' data refers to raw data that has been reviewed, edited for quality when necessary, and certified for general use. Being certified for use is the primary defining criterion of base data. Activities that transition raw to base data include flagging values for quality issues; removing outliers; handling missing data as appropriate; standardizing field names, units and formats; rearranging data; aggregating data into collections with similar data; and making collections of data accessible. In some cases well-formed raw data may simply become certified as base quality and undergo no change at all. In other cases high-frequency instrument-generated data may be reduced to a surrogate value appropriate for scientific research. The result of these activities produces the high-quality, accepted, base facts used in all higher level interpretations of the original data. Base data may exist in many forms, such as discrete facts in relational databases, processed scenes from satellites, or LIDAR delivered by a contractor.
- *data*

Data includes "... textual information, numeric information, instrument readouts, equations, statistics, images (whether fixed or moving), diagrams, maps, and audio and video recordings. It includes raw data, processed data, published data, and archived data. It includes data generated by experiments, by models and simulations, and by observations of natural phenomena at specific times and locations. It includes data gathered specifically for research as well as information gathered for other purposes that is then used in research."
(ref: National Academy of Sciences, 2009, pg. 22).
- *Data Manager*

The Data Manager is the highest-level manager of Data Management services in the Program. The Data Manager is the manager most directly responsible for data management, including coordinating data governance, data stewardship activities, overseeing data management projects, and supervising data management professionals.
- *data product*

'data product' is any data information product produced and published from base data, derived data, a project database, or any combination of these sources.
- *Data Steward*

A Data Steward is a person knowledgeable in a particular area or topic who is assigned accountability for data specifications and data quality for a specific Project or dataset. A Data Steward will draft, review, and refine data definitions; review and approve data model specifications; define data quality requirements and business rules; review and approve project and data level metadata; identify and help resolve data issues; and provide input to data policy, standards, and procedure for their assigned area.
- *doi*

A digital object identifier (DOI) is a character string used to uniquely identify an object such as an electronic document. Metadata about the object is stored in association with the DOI name and this metadata may include a location, such as a URL, where the object can be found. The DOI for a document remains fixed over the lifetime of the document, whereas its location and other metadata may change. Referring to an online document by its DOI provides more stable linking than simply referring to it by its URL, because if its URL changes, the publisher need only update the metadata for the DOI to link to the new URL.

- *existing data*
The term ‘existing data’ refers specifically to data whose origin is external to the research Project. Existing data can be obtained from any source.
- *FGDC*
The Federal Geographic Data Committee (FGDC) is an interagency committee that promotes the coordinated development, use, sharing, and dissemination of geospatial data on a national basis. In the context of metadata, FGDC is best known for its 1993 Content Standard for Digital Geospatial Metadata (CSDGM) most often referred to as FGDC metadata. FGDC metadata is still the widely used in the U.S. but has little international following.
- *ISO*
ISO (International Organization for Standardization) is an independent, non-governmental membership organization and the world's largest developer of voluntary International Standards. In the context of metadata, the ISO standards are gradually replacing the FGDC standard as global scale science is requiring data from around the world. ISO’s entry into geospatial metadata was the 19115 standard in 2003. The standard is updated approximately every 5 years with 19115-2 standard issued in 2009 and 19115-1 in 2014.
- *Metadata*
Metadata describes USGS scientific data and how they were collected and processed. Metadata include specific attributes such as spatial coverage, scale, collection methods, citation, abstract, purpose, and physical format.
- *Metadata Author*
A Metadata Author is responsible for writing the content for the metadata record by insuring the completeness and accuracy of the ASC RDMP forms. The metadata author is not responsible for transforming the ASC RDMP forms into any specific metadata standard.
- *Metadata Specialist*
A Metadata Specialist is a Data Management professional responsible for the preparation, integration, control, and delivery of metadata including metadata repositories.
- *new data*
The term ‘new data’ refers specifically to data whose origin is the research Project. New data can result from observations, lab analysis, instrumentation placed by a project, or other means.
- *quality control (QC)*
Quality control (QC) is a procedure or set of procedures intended to ensure that a data product adheres to a defined set of quality criteria or meets the requirements of the researcher. The procedures must be designed by a person knowledgeable about what constitutes quality in the dataset. Procedure can be visual or automated and can check for missing values, invalid data types, values not in acceptable range, values not represented in a domain, and other criteria as needed. Quality assurance steps are to be performed at all stages of the science data lifecycle.
- *process step*
Steps performed to verify, organize, transform, repair, integrate, and produce data in an appropriate form for subsequent use.
- *raw data*

'raw' data refers to data which are not modified as obtained from their original source except as necessary to allow for their long-term storage in an electronic format. Raw data may exist as discrete values recorded by a field worker, data streamed from instruments, photographs, strip-charts, lab reports, or remotely sensed data. The classification of raw data is also extended to data obtained from an existing external source that has not yet been reviewed. The critical indicator is that the data values are unaltered as collected at the source. Raw data are not usually made available for direct use by researchers. However, raw data are a high priority for stable, long-term storage because they are the original source of data at all other levels.

- *RDMP*

The Research Data Management Plan (RDMP) defines how data will be handled during the research cycle. It is also a record of actions taken by the research team during the research cycle that become the basis for the long-term care and preservation of the data including archival, publication, and distribution.

ARCHIVAL FORMATS

Information on archival formats is provided by the National Archives.

DIGITAL AUDIO

The Digital audio category encompasses formats used to encode recorded sound as machine readable files by converting acoustic sound waves into digital signals. Digital audio formats are generally composed of both a wrapper format, usually the common name associated with the file extension, and an encoding method or codec.

General requirements for digital audio records:

- Digitize to standards appropriate for the accurate preservation of the original audio, when converting analog material (e.g., audio cassettes, record albums, and reel-to-reel audio tapes).
- Transfer digital audio at a minimum of 16 bits per sample, but 24 bits per sample is encouraged; and
- Transfer digital audio at a minimum sample rate of at least 44.1 KHz, but sampling at 96 KHz is encouraged.

Preferred Formats	Codecs	Specifications
Audio Interchange Format (AIFF)	Linear Pulse Code Modulation Audio (LPCM)	Audio Interchange File Format: "AIFF" A Standard for Sampled Sound Files Version 1.3, Apple Computer, Inc.: (http://www-mmsp.ece.mcgill.ca/Documents/AudioFormats/AIFF/Docs/AIFF-1.3.pdf)
Waveform Audio File Format (Wave)	Linear Pulse Code Modulation Audio (LPCM)	Multimedia Programming Interface and Data Specifications 1.0: (http://www-mmsp.ece.mcgill.ca/Documents/AudioFormats/WAVE/Docs/riffmci.pdf)

DIGITAL VIDEO

Digital moving images consist of bitmap digital images or "frames" displayed in rapid succession at a constant rate, giving the appearance of movement.

General requirements for digital moving image records:

- Agencies must digitize to standards appropriate for accurate preservation of the original video and audio components, when converting analog material.

Preferred Formats	Codecs	Specifications
Audio Video Interleaved Format (AVI)	Uncompressed 4:2:2	Multimedia Programming Interface and Data Specifications 1.0: (http://www.kk.iij4u.or.jp/~kondo/wave/mpidata.txt)
QuickTime Media Video 9 File Format (MOV)	Uncompressed 4:2:2	Apple QuickTime File Format Specification (ISO/IEC 14496-14:2003): (https://developer.apple.com/library/mac/documentation/QuickTime/QTFF/QTFFPreface/qtffPreface.html#//apple_ref/doc/uid/TP40000939)
MPEG4	H.264	ISO/IEC 14496-10:2003. Information technology -- Coding of audio-visual objects -- Part 10: Advanced Video Coding (formal name) MPEG-4, Advanced Video Coding: (http://www.iso.org/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=37729)

DIGITAL PHOTOGRAPHS

Digital photographs include still photographs of natural, real-world scenes or subjects produced by digital cameras, and scanned images of photographic prints, slides, and negatives. The guidance applies to master image files of digital photographs created using medium to high quality resolution settings appropriate for continued preservation.

General requirements for digital photographic records:

- Agencies should use appropriate, professional quality, dedicated photographic equipment when capturing images;
- When converting analog material (photographic prints, glass plate negatives, slides, etc.), agencies must digitize to standards appropriate for the accurate preservation of the original image;
- Agencies must digitize analog originals at a minimum resolution of 3,000 pixels across the long dimension; and
- Images that are uncompressed or which make use of lossless compression are preferred.

Preferred Formats	Specifications
Tagged Image File Format (TIFF)	TIFF Revision 6.0 Final — June 3, 1992 Adobe Systems Incorporated: (http://partners.adobe.com/public/developer/en/tiff/TIFF6.pdf)
JPEG File Interchange Format	ISO/IEC 10918-5 Information technology – Digital Compression and coding of continuous-tone still images: JPEG Interchange File Format: (http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=54989) ISO/IEC 10918-1:1994 Information technology – Digital Compression and coding of continuous-tone still images: Requirements and guidelines: (http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=18902)
Digital Negative (DNG)	Adobe Digital Negative (DNG) Specification Version 1.4.0.0: (http://www.images.adobe.com/www.adobe.com/content/dam/Adobe/en/products/photoshop/pdfs/dng_spec_1.4.0.0.pdf)

SCANNED TEXT

Scanned text is a photograph of a printed page produced either by a digital camera or scanner.

General requirements for scanned text include the following:

- Agencies must digitize to standards appropriate for the accurate preservation of the information on the printed page. When converting analog or film based material (microfilm, microfiche, slides, etc.), agencies must digitize to standards appropriate for the accurate preservation of the original image.
- Bitonal (1-bit black and white) images must be scanned at 300-600 ppi. Scanning at 600 ppi is recommended. This is appropriate for documents that consist exclusively of clean printed type possessing high inherent contrast (e.g., laser printed or typeset on a white background);
- Gray scale (8-bit) must be scanned at 300-400 ppi. Scanning at 400 ppi is recommended. This is appropriate for textual documents of poor legibility because of low inherent contrast, staining or fading (e.g., carbon copies, thermofax, documents with handwritten annotations or other markings), or that contain halftone illustrations or photographs; and
- Color (24-bit RGB [Red, Green, Blue]) must be scanned at 300-400 ppi. Scanning at 400 ppi is recommended. Color mode (if technically available) is appropriate for text containing color information important to interpretation or content.

Preferred Formats	Specifications
Portable Document Format/Archival (PDF/A)	ISO 19005-1:2005 Electronic document file format for long-term preservation – Part 1: Use of PDF 1.4 (PDF/A-1): (http://www.iso.org/iso/catalogue_detail?csnumber=38920)
Tagged Image File Format (TIFF)	TIFF Revision 6.0 Final — June 3, 1992 Adobe Systems Incorporated: (http://partners.adobe.com/public/developer/en/tiff/TIFF6.pdf)

GEOSPATIAL FORMATS

Geospatial records include digital cartographic data files and aerial photography that are created and processed in Geographic Information Systems (GIS) or other software applications for spatial analysis.

Preferred Formats	Specifications
Geospatial Tagged Image File Format (GeoTIFF)	Geo TIFF Format Specification: (http://www.remotesensing.org/geotiff/spec/geotiffhome.html)
Geographic Markup Language	ISO 19136:2007 & Version 3.2, OGC document 07-036: (http://www.opengeospatial.org/standards/is)

PRESENTATION FORMATS

Presentation formats are used to convey graphical information to audiences in the form of a slide show. Presentation formats are not acceptable for use as transfer containers for permanent digital still images.

Preferred Formats	Specifications
OpenDocument Presentation Format (ODP)	ISO/IEC 26300:2006 Information technology -- Open Document Format for Office Applications (OpenDocument) v1.0: (http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=43485)
Portable Document Format Archival (PDF/A)	ISO 19005-1:2005 Electronic document file format for long-term preservation – Part 1: Use of PDF 1.4 (PDF/A-1): (http://www.iso.org/iso/catalogue_detail?csnumber=38920)

TEXTUAL DATA

The textual data category refers to two general content types: unformatted (plain text) or formatted. Unformatted plain text (defined in MIME as text/plain) contains basic character information and control or non-printing characters but lacks styling information. Formatted text files include all of the attributes of plain text files but have extended formatting capabilities, for “stylized” or “rich” text features including italics, bold, colors, hyper-linking, etc.

Preferred Formats	Specifications
ASCII Text 7 bit	ISO/IEC 646:1991 Information technology -- ISO 7-bit coded character set for information interchange: (http://www.iso.org/iso/catalogue_detail.htm?csnumber=4777)
OpenDocument Presentation Format (ODP)	ISO/IEC 26300:2006 Information technology -- Open Document Format for Office Applications (OpenDocument) v1.0: (http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=43485)
Portable Document Format Archival (PDF/A)	ISO 19005-1:2005 Electronic document file format for long-term preservation – Part 1: Use of PDF 1.4 (PDF/A-1): (http://www.iso.org/iso/catalogue_detail?csnumber=38920)

STRUCTURED DATA FORMAT

Structured data comprises the broad category of data that is stored in defined fields. Categories for structured data are as follows:

- Database formats are organized collections of associated data that conform to a logical structure. Database formats are determined by “data models” that describe specific data structures used to model an application and generally include navigational, relational, and hybrid models;
- Spreadsheets are tables made up of columns and rows and which contain cells of data. Relationships between cells can be pre-defined as mathematical formulas;
- Statistical data is the result of quantitative research and analysis. Statistical data formats contain collections of data presented in both tabular and non-tabular form; and
- Scientific data refers to research data collected by instrumentation tools during the scientific process. Scientific data formats are either domain specific within a single field of study, or are multi-domain formats used for transfer of scientific data between domains.

General requirements for structured data include the following:

- Agencies must transfer structured data that is both well-formed according to the syntactical conventions of the format, and valid according to the structural rules defined in any associated schemas or document type definitions (DTDs);
- Value Separated Files, e.g. CSV or comma separated value files, may use a character other than the comma. The pipe ‘|’ or caret ‘^’ are recommended delimiters because they are not commonly found in free text fields. Alternatively, text files encoded with ASCII characters and where each field is a fixed width, is also an acceptable transfer format for use with structured data, even though ASCII is technically a data encoding type. ASCII text files must be accompanied by complete documentation of the record lengths and field widths;
- Data files and databases shall be transferred as flat files or as rectangular tables, that is, as two-dimensional arrays, lists or tables. All records in a database, or rows (tuples) in a relational database, should have the same logical format. Each data element within a record should contain only one data value. A record should not contain nested repeating groups of data items; and
- Structured data must be transferred together with any associated files necessary to verify the validity of the data, e.g., DTDs, schemas, and data dictionaries.

Preferred Formats	Specifications
Comma Separated Values (CSV)	Common Format and MIME Type for Comma-Separated Values (CSV) Files: (http://tools.ietf.org/html/rfc4180)
ASCII Text 7 bit	ISO/IEC 646:1991 Information technology -- ISO 7-bit coded character set for information interchange: (http://www.iso.org/iso/catalogue_detail.htm?csnumber=4777)
JavaScript Object Notation (JSON)	The application/json Media Type for JavaScript Object Notation (JSON): (http://www.ietf.org/rfc/rfc4627.txt?number=4627)
Extensible Markup Language	Extensible Markup Language (XML) 1.1 (Second Edition): (http://www.w3.org/TR/2006/REC-xml11-20060816/)

EMAIL

Email is defined as discrete electronic communications transmitted over the Simple Mail Transfer Protocol (SMTP), between two or more people or entities, in compliance with applicable IETF’s Request for Comments (RFC) specifications. Email does not include other functions commonly available via email programs such as calendars, tasks, appointments, newsgroups, or instant messaging. In order for information in a calendar, contact list, address book etc. to be transferred to NARA, it must be scheduled as a separate item.

Please note that NARA considers email attachments to be a component of the email record and does not require that unseparated email attachments meet the transfer standards specified by the format category under which the attachment alone would fall.

General requirements for email:

- Transfers of email records must consist of an identifiable, organized body of records (not necessarily a traditional series);
- Email messages should include delimiters that indicate the beginning and end of each message and the beginning and end of each attachment, if any. Each attachment must be differentiated from the body of the message, and uniquely identified;
- Email messages transferred as XML files must be accompanied by any associated document type definitions (dtds), schemas, and/or data dictionaries;
- Labels to identify each part of the message (Date, To [all recipients, including cc: and bc: copies], From, Subject, Body, and Attachment) including transmission and receipt information (Time Sent, Time Opened, Message Size, File Name, and similar information, if available). To ensure identification of the sender and addressee(s), agencies that use an email system that identifies users by codes or nicknames, or identifies addressees only by the name of a distribution list should include information with the transfer-level documentation; and
- Email converted to formats not natively used by the email program, and which do not maintain header information (such as RTF or Word documents) are not accepted. Printouts of emails are also not accepted under this Bulletin.

Preferred Formats	Specifications
Internet Message Format (EML)	Internet Message Format: (http://www.ietf.org/rfc/rfc2822.txt)
MBOX Email Format (MBOX)	MBOX Email Format: (https://tools.ietf.org/html/rfc4155)

REFERENCES

OSTP, 2013, Increasing Access to the Results of Federally Funded Scientific Research, Office of Science and Technology Policy, February 22,

2013. http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf.

OMB, 2013, Open Data Policy – Managing Information as an Asset, Office of Management and Budget, May 9,

2013. <http://project-open-data.github.io/policy-memo/>

Stan Smith, Steven Tessler, Michael McHale, Thomas Burley, 2012, Data Management Planning Framework (DMPf), Part III - Research Data Management Plan (RDMP) – Enterprise, version 2.2, U.S. Geological Survey. Accessed on March 15, 2013, at http://tx.cr.usgs.gov/data-management/DMPf/DMPf_Part_III_Enterprise_v2_2_1.pdf

Emily Fort, Stan Smith, et al, 2014, National Climate Change & Wildlife Science Center & Climate Science Centers Data Management Plan Guidance, April 17, 2014. https://nccwsc.usgs.gov/sites/default/files/images/DMPGuidance_v3_0.pdf

Steve Tessler, National Water Census Data Management Planning Guidance, version 0.5, February 2014, unpublished.

Stan Smith, Joshua Bradley, Dennis Walworth, Allison Gaylord, Chris Turner, Drew Ignizio, Alaska Data Integration working group (ADIwg), Open Source ISO Metadata Toolkit, 2012-2014. <http://www.adiwg.org/projects/>

National Archives, Records Management, Policy, Appendix A: Table of File Formats, April

2014. <http://www.archives.gov/records-mgmt/policy/transfer-guidance-tables.html>