Alaska Science Center
Data Policy

Research Data Management Plan Guidance & Forms for Principal Investigators

October 21, 2014

Version 1.3

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 a project proposal 8
2. Project Selected for Funding 9
3. Design Spreadsheet/Database 10
4. Build Database 11
5. Design Field Form 12
6. Acquire & Archive Raw Data 13
7. Complete QC of Raw Data & Archive Base Data 14
8. Conclude Research Cycle 15
9. Publish Data 16
10. Publish Paper 17

Research Data Management Plan Forms 18

ASC Project Proposal Form 20

1. ASC Project Implementation Form 21
2. ASC Project Forms Table of Contents 22
3. ASC New Data Collection Form 23
4. ASC Existing Data Form 25
5. ASC Derived Data Product Form 26
6. ASC Custom Software Form 28
7. ASC Model Form 29
8. ASC Data Description Form 30
9. ASC Data Domain Form 31
10. ASC Data Process Step / Data Quality / Lineage Form 32
11. ASC Project Contact Form 33

Glossary 34

Archival Formats 37

1. Digital Audio 37

2. Digital Video 37
3. Digital Photographs 38
4. Scanned Text 38
5. Geospatial Formats 39
6. Presentation Formats 39
7. Textual Data 39
8. Structured Data Format 40
9. Email 41
References 42

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 research funding and implement these requirements to benefit the Alaska Science Center (ASC), USGS, the greater scientific community, and the public.

The data management requirements are stated in a series of federal laws and mandates, DOI 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 must also follow the requirements of our Fundamental Science Practices (FSPs) that clarify how science is conducted and resulting 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.

ASC DATA POLICY

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).

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, DOI, and USGS laws, 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 for the project.
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 databases are to be used whenever possible for data archive, storage, and distribution.
VI. Project data are developed 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 made public will
    a. be published to the finest possible level of granularity practicable and permitted by law,
    b. use machine-readable and open formats, 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.
VIII. 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.
IX. Data will be finalized in a way that products are
    a. documented using formal standardized metadata formats such as ISO 19115-2 or FGDC,
    b. provided with a valid digital object identifier (doi),
    c. ready to share with other ASC staff through a designated ASC repository in a timely manner,
    d. shared with the public at large within 12 months of completion of the work or product, and
    e. undergo a review process and receive approval before archiving or release to the public.
X. ASC will maintain an inventory of data resources which will in turn update the USGS data inventory.
XI. ASC will maintain a listing of public data which will in turn be listed at data.usgs.gov in human- and machine-readable formats.

XII. ASC will maintain the authoritative copy of data and metadata for public distribution. Copies distributed to other distribution centers will refer to ASC for maintenance and updates.

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 preponderance of information needed to prepare ISO or FGDC project and data metadata.

In practice no two projects are likely follow the same workflow. The RDMP acknowledges this variability in project requirements and a 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 may 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 the forms for new data collections, existing data, data products, custom software, and models only require the Name: and Description: fields to be filled in for the project proposal. The remaining form fields can be ignored until after the project has been approved for funding. A short consultation with a Data Management staff member is required to acquaint Data Management (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.
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".

Project Stage	RDMP Action Required
Research	As research progresses, 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.
Completion	At project completion, DM will use the completed RDMP forms to prepare metadata records for raw and base data required to be archived, product data for internal USGS use, and data to be published. DM will add raw and base data to the ASC data archives. Document Object Identifiers (doi) will be obtained for data products selected for publication. The project record in 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 PI.

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 steps to be taken by members of the Data Management staff to handle the long-term preservation, exposure, and delivery of data are not presented.

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 lean what data management actions he/she should take.

Project events requiring data management action:

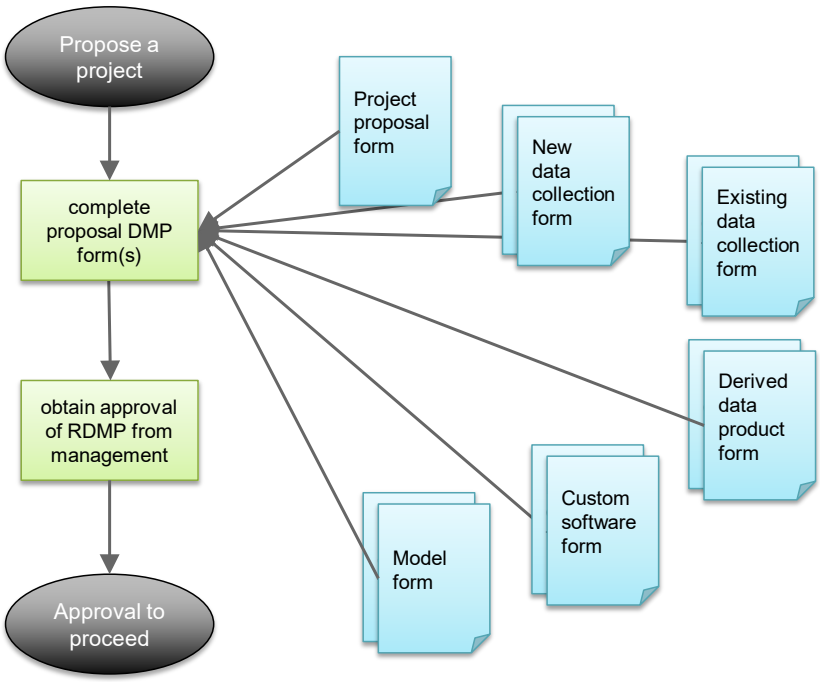
1. Submit a project proposal
2. Project selected for funding
3. Design a spreadsheet
4. Design a database
5. Build a database
6. Design a field data collection form
7. Acquire & archive raw data
8. Complete QC of raw data & archive base data
9. Conclude research cycle
10. Publish data
11. Publish paper

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 complete a RDMP forms.

1. SUBMIT A PROJECT PROPOSAL

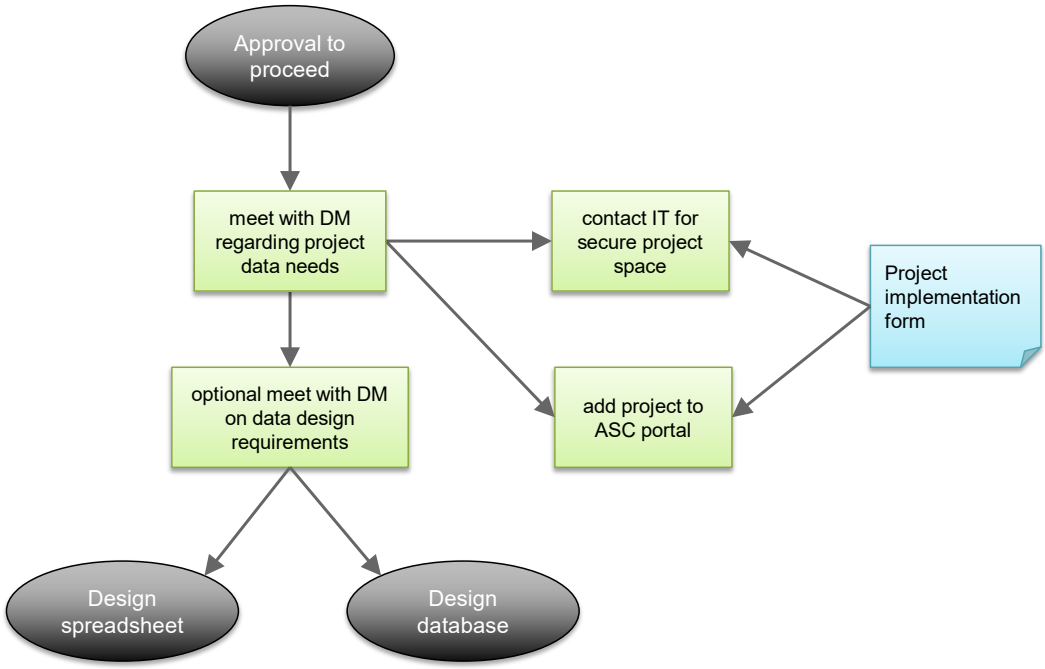
a. Research Data Management Plan -> Each new project proposal is required to attach a brief Research Data Management Plan (RDMP). The proposal-level RDMP outlines the project's anticipated data requirements and serves as a skeleton for full RDMP should the project be funded. Details will be added to the RDMP through the research cycle once the project is approved and as research progresses. The proposal-level RDMP requires a completed Project Proposal Form accompanied by a Table of Contents and separate New Data Collection Forms, Existing Data Forms, Derived Data Product Forms, Custom Software Forms, and Model Forms for each data, software, or model item planned for the project. Note that for the proposal-level RDMP only the Name: and Description: fields are required for the data forms. The remaining fields can be ignored until after the project has been approved and research has begun.

b. Approval -> After completing the RDMP obligation for a proposed project you will need to obtain an ASC Manager sign-off of the Project RDMP to submit a qualified project proposal.



2. PROJECT SELECTED FOR FUNDING

a. Consultation -> A short consultation with a Data Management staff member is required to acquaint Data Management (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. Research Data Management Plan -> Once your project has been selected for funding you should add the Project Implementation Form to your RDMP and complete the form.
c. ASC Science Portal -> Add your project record to the ASC Science Portal with a status of "In Progress".
d. Project Folder -> Coordinate with your Information Technology (IT) representative to have a secure space on an IT managed servers set up to hold your project files and data. IT will apply access permissions to this folder according to the schedule you prepared 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. This is not a required step but it is an optimal time to discuss your data design options.



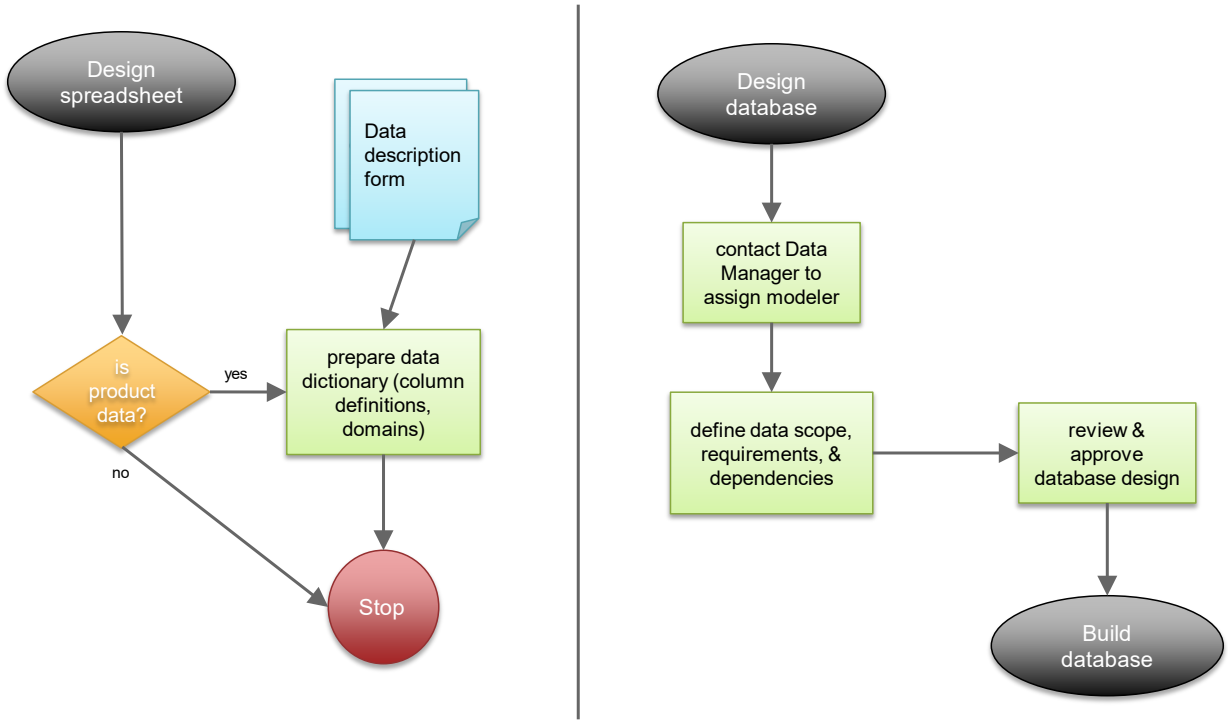
3|4. DESIGN SPREADSHEET/DATABASE

Design Spreadsheet

a. Research Data Management Plan -> If the spreadsheet will be become a data product (i.e. be archived, used in future research, published with a paper, or become available to the public) add a Data Description Form to your RDMP for each sheet in the spreadsheet. Complete the form(s) and remember to attach a Data Domain Form for each domain used by a spreadsheet column. Spreadsheets that are used only as intermediate work steps do not need to be documented.

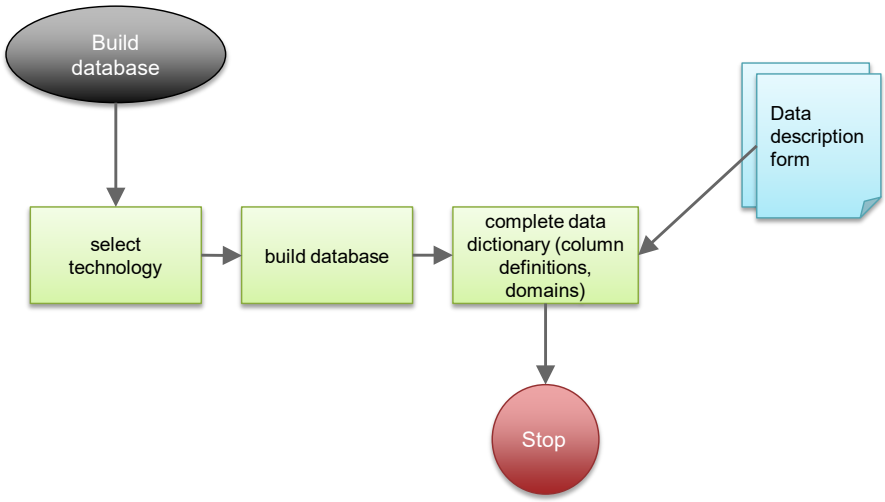
Design Database

a. Assign Data Modeler -> All new database designs will need to be reviewed and approved by the project's Data Manager prior to building the database. Contact your Project's Data Manager to have a Data Modeler assigned to your database project. It is not required that the Data Modeler perform the database design work or build the database, although these options are open and encouraged. The Data Modeler can alternatively provide consultation during the design process and will be responsible for the final review and approval of any design.
b. Design Database -> Define the database scope, entities and attribution, dependencies between entities, and attribute domains.
c. Review & Approval -> Submit the database design to the Data Modeler assigned to the project for final review and approval by the Data Manager. The Data Modeler may suggest some changes to improve database integrity and/or performance.



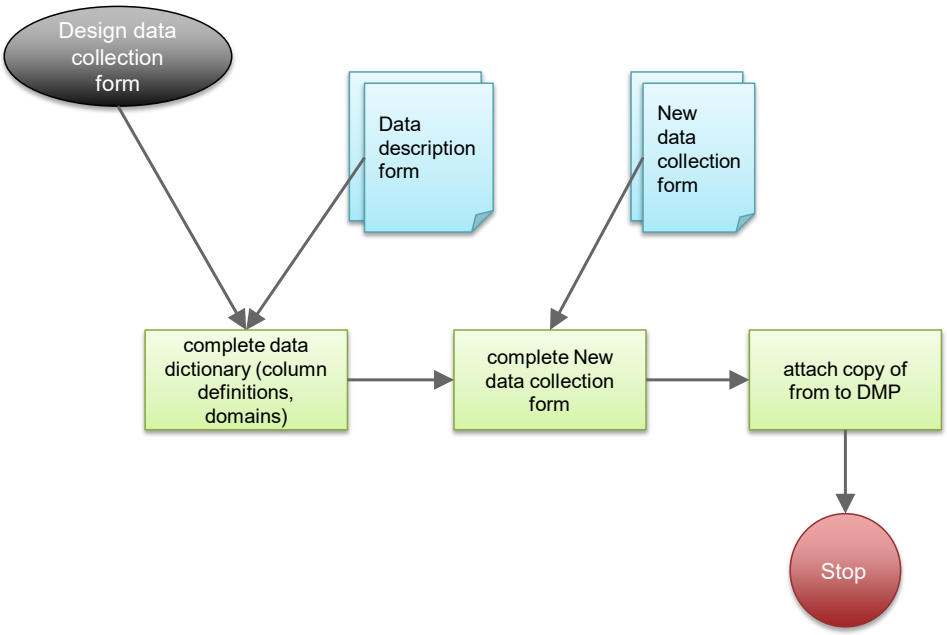
5. BUILD DATABASE

a. Select Database Technology -> Select the database technology for implementing the database. Selection of a technology commonly used at the ASC is encouraged. Your assigned Data Modeler can help you make this choice taking into consideration size, user community, features, and Project team member 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 by the table(s).



6. DESIGN FIELD FORM

a. Design Data Collection Form -> Design your field or lab data collection form. Remember to include fields to record the observer name, date, and some unique identifier or sequence number on each form.
b. Document the Form -> For each form, add a Data Description Form to your RDMP. Complete the form and remember to attach a separate Data Domain Form for each domain used by a form field.
c. Add Form to RDMP -> Add a copy of the form to your RDMP to complete the documentation.

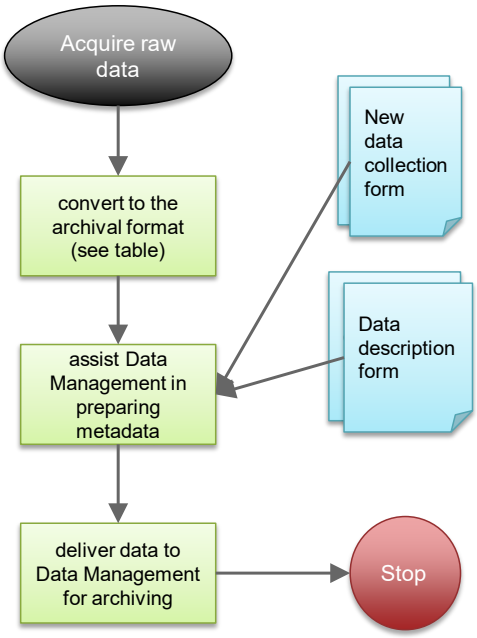


7. ACQUIRE & ARCHIVE RAW DATA

a. Prepare Raw Data for Archive -> Raw data should be archived as soon as possible after collection. This is practice will preserve the raw data in its truest state and reduce the possibility of loss. Raw data will not be released beyond the immediate Project team members without authorization from the PI. Convert the raw data to the most appropriate archival format as depicted 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 or contact your Project's Data Manager.

b. Prepare Metadata -> All archived raw data files will have comprehensive metadata records and will be prepared to the Library of Congress standard 'bagit 0.9'. The PI should work with the Project's Metadata Specialist to prepare the metadata and archive package. Data Description Forms prepared earlier will be used by Metadata Specialist to complete this task. If the necessary forms were not completed earlier, that task will need to be added here.

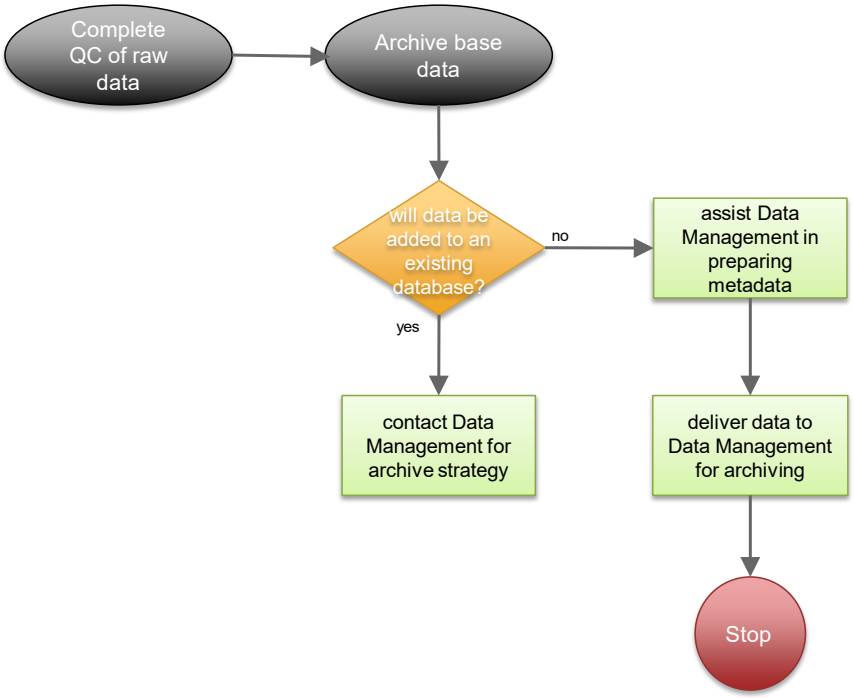
c. Archive Raw Data -> Deliver the fully documented archive package(s) to the Metadata Specialist for archiving. Request READ access for Project team members to access the archive folder.



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

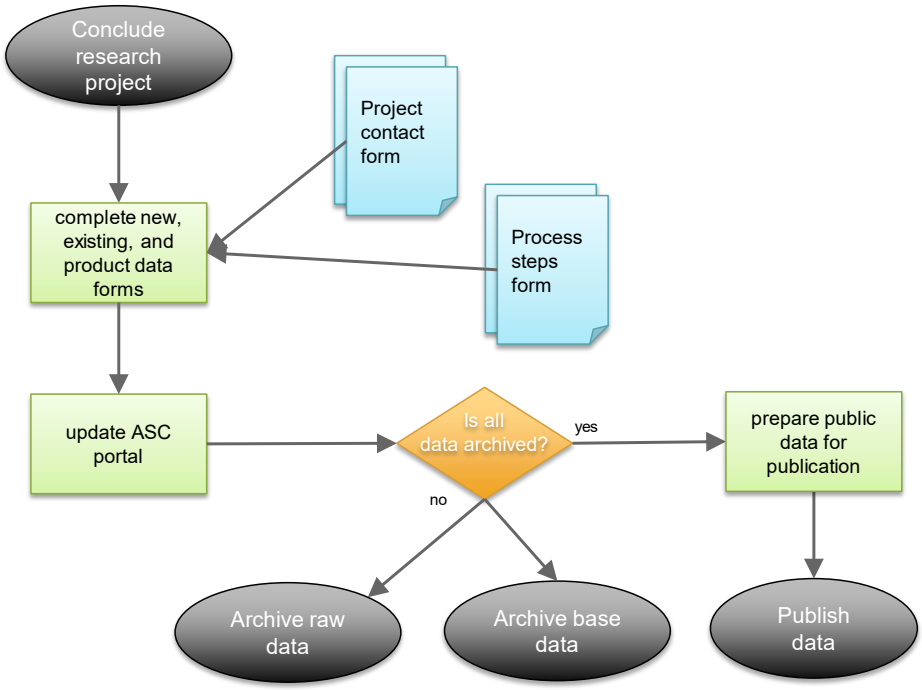
8. COMPLETE QC OF RAW DATA & ARCHIVE BASE DATA

a. QC Raw Data -> Perform quality control (QC) of raw data. Raw data that has been QC'ed and verified ready for use in research is termed 'base' data. Base data is also archived to the same standards as raw data.
b. Prepare Metadata -> All base data archived files will have comprehensive metadata records and will be prepared to Library of Congress standards 'bagit 0.9'. The PI should work with the Project's Metadata Specialist to prepare the metadata and archive package. Data Description Forms prepared for the raw data archive will be used by Metadata Specialist to complete this task. If the necessary forms were not completed earlier, that task will need to be added here.
c. Archive Base Data -> Deliver the fully documented archive package(s) to the Metadata Specialist for archiving. Request READ access for Project team members to access the archive folder. At the conclusion of the research project the folder will become accessible to all USGS researchers.
d. Add Base Data to Existing Database -> If the base data is to be added an existing database or archive, contact your Project's Data Manager for determining next steps.



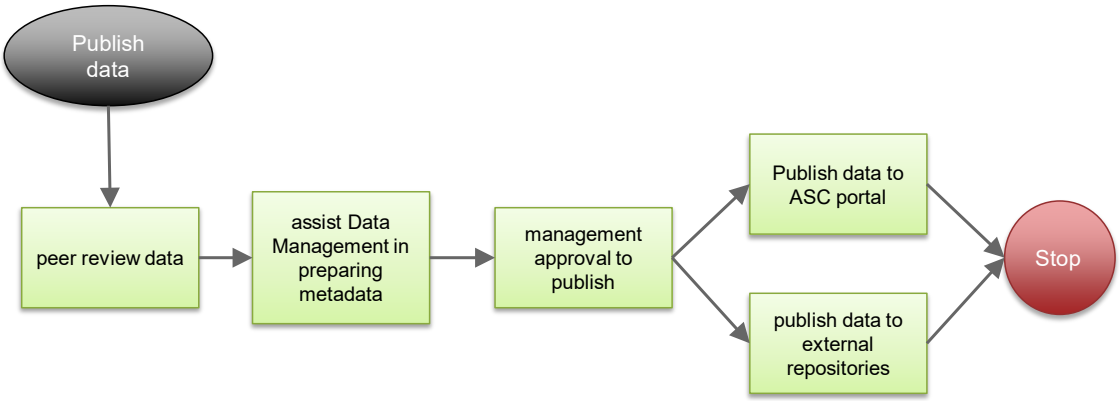
9. CONCLUDE RESEARCH CYCLE

a. Complete RDMP Data Forms -> Complete any RDMP New Data Collection Forms, Existing Data Forms, Data Product Forms, Custom Software Forms, and Model Forms not completed earlier as these will be needed to create your project and data metadata records prior to publication. This includes Project Contact Forms and Process Step Forms not already completed for the Project's data forms.
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.
d. Prepare Public Data -> Meet with your Project's Data Manager to determine which of your project data is appropriate for publication, adjust its granularity, and determine potential USGS and external repositories.



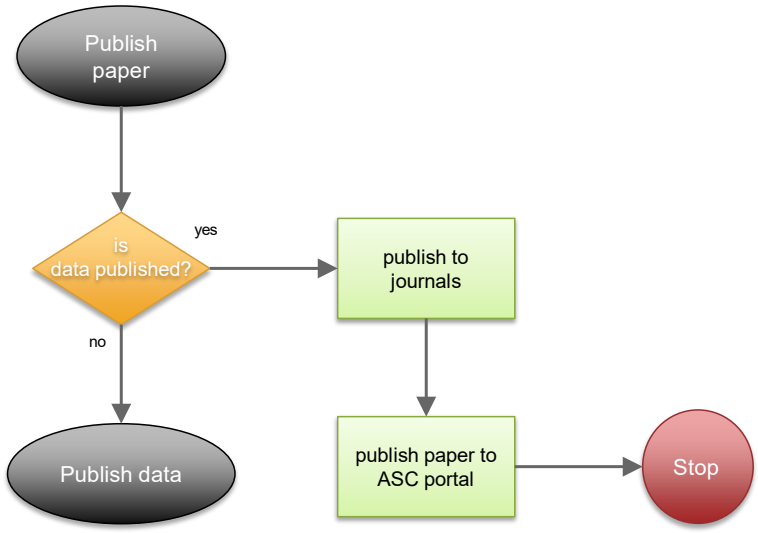
10. PUBLISH DATA

a. Peer Review Data -> All data destined for publication is to be peer reviewed. The Project's Data Manager will be a member of the review team in addition to any reviewers named by the PI.
b. Prepare Metadata -> All public data files will require comprehensive metadata records and will be prepared to FGDC Content Standard for Digital Geospatial Metadata (CSDGM), ISO 19115-2:2009, or ISO 19115-1:2014 standards. Engage with your Project's Metadata Specialist to prepare the metadata for your data publications. Data Product Forms and Data Description Forms prepared earlier will be required by Metadata Specialist to complete this task. If the forms were not completed, that task will need to be added here.
c. Manager Approval to Publish -> Results of the peer review, completed data distribution packages, metadata, and forms are to be submitted to the Project's Manager for approval to publish.
d. Publish to Portal -> All data to be made public and/or to accompany a publication will be accessible through the ASC Science Portal. The Project's Data Manager will assist the PI with getting this data online.
e. Publish to Other Repositories -> The PI is free to publish data approved for release to the public through other thematic portals or repositories as are appropriate.



11. PUBLISH PAPER

a. Publish Data -> It is a good practice – and required by some publishers – to publish your Project's data prior to the publication of paper. This practice ensures web links and Digital Object Identifiers (doi) included in the journal article are accurate and those reading the article have immediate access to the supporting data. If you have not published the data supporting this research paper, that step should be inserted here.
b. Publish to Journal -> The PI should follow all normal and established USGS protocols for publishing science articles.
c. Publish to Portal -> All ASC publication for Alaska science are to be made accessible through the ASC Science Portal. If the article is also to be published in a science journal, publication in the ASC Portal should follow publication of the journal. The Project's Data Manager is available to assist the PI with getting the paper online.



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 denoting 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 Name: and Description: fields need 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 Name: and Description: fields need 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 data product (see Glossary) produced by the project. For the project proposal, only the Name: and Description: fields need 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/or web tool developed by the project as a product or required to fully understand the scientific process applied. For the project proposal, only the Name: and Description: fields need 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 Name: and Description: fields need 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 or 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 which are 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 which receive 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. Attach the domain's form ID to the appropriate Data Description Form.

11. ASC Data Process Steps / Data Quality / Lineage Form
    Add this from 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. Full descriptions of the processing steps are best left to the project's paper.

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 requesting 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/planning phase.

	PROJECT [*Enter Project short name*]
Title:	*Enter full project name*
Funding Source:	USGS Program: *Select a program.* Major Initiative: *Select an initiative.*
Abstract:	*Enter abstract.* *or Enter abstract path/file name.*
Project Start Date:	*Enter project start date.*
Project End Date:	*Enter project end date.*
Spatial Description:	*Enter the general geospatial scope of the project or N/A if none.*
Temporal Description:	*Enter the temporal scope of the study or N/A if none.*
Responsible parties:	Program Manager: *Enter name of Program Manager.* Principal Investigator(s): *Enter name(s) of Principal Investigator.* Researcher(s): *Enter name(s) of Researcher.* Data Steward: *Enter name of Data Steward.* Data Manager: *Select Data Manager.* Metadata Specialist: *Select Metadata Specialist.*
Hardware environment:	*List computing and field hardware requirements.*
Software environment:	Describe software requirements.
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:	*Enter short abstract.*
Project location:	Map point: *Latitude.* *Longitude.* Bounding box: *North Latitude.* *East Longitude.* *South Latitude.* *West Longitude.* Boundary polygon: *Enter boundary polygon path/file name.* Common Place Names: *Enter commonly known names for project location.*
Keywords:	*Enter keyword list.*
ASC Portal information:	`Project ID:` *Enter project portal ID.* `Contact ID:` *Enter maintenance Contact ID:*
Project folder:	*Enter server name.* *Enter project path/folder.* Permissions: Management: `write` Project team: *Select access level.* Data Management: *Select access level.* USGS ASC: *Select access level.* USGS all: *Select access level.* Public: *Select access level.*
Responsible parties:	Funding Organization(s): *Enter Contact Form ID(s) for Funder.* Point of Contact(s): *Enter Contact Form ID(s) for Point of Contact.* Principal Investigator(s): *Enter Contact Form ID(s) for Principal Investigator.* Researcher(s): *Enter Contact Form ID(s) for Researcher.*
Other contributors:	*Enter name(s) of other contributor(s).*

ASC PROJECT FORMS TABLE OF CONTENTS

Use this form is to list the ASC Data Management Forms attached to your project.

New Data Collections:	Form Name: *Enter form Name:*		ID: *Enter Form ID:*
	☐ field form design	☐ data collected	☐ RDMP form complete
	☐ raw metadata prep	☐ archive raw data	☐ QC raw data
	☐ base metadata prep	☐ archive base data	☐ Publish
Existing Data Collections:	Form Name: *Enter form Name:*		ID: *Enter Form ID:*
	☐ data acquired	☐ RDMP form complete	
Project Data Products:	Form Name: *Enter form Name:*		ID: *Enter Form ID:*
	☐ RDMP form complete	☐ metadata prep	☐ data review
	☐ archive data	☐ prep data for public	☐ web development
	☐ publish		
Custom Software:	Form Name: *Enter form Name:*		ID: *Enter Form ID:*
	☐ document software	☐ RDMP form complete	☐ archive software
Models:	Form Name: *Enter form Name:*		ID: *Enter Form ID:*
Project Contacts:	Form Name: *Enter form Name:*		ID: *Enter Form ID:*
Data Description:	Form Name: *Enter form Name:*		ID: *Enter Form ID:*
Data Domain:	Form Name: *Enter form Name:*		ID: *Enter Form ID:*
Data Lineage:	Form Name: *Enter form Name:*		ID: *Enter Form ID:*

ASC NEW DATA COLLECTION FORM

Complete this form for data that does not currently exist and will be collected or generated during the project; for example, a new field data collection. This form will follow new data as it matures from raw to base to published data.

ID [*Enter ID*]	
Name:	*Enter new data collection name.*
Description:	*Enter description for new data collection.*
---------- To be completed after project has been approved for funding ----------	
---------- To be completed for raw data ----------	
Format:	*Select format.* *If other, enter format.*
Data collection template:	`Link to template:` *Enter template path/file name.* `Template Data Description Form:` *Enter Data Description Form ID:*
Data collection protocols:	*Enter data collection protocols.*
Spatial reference system:	*Select a spatial reference system.* *If other, enter reference system.*
Spatial scope & scale:	`Description:` *Enter description of spatial scope.* `Scale:` *Enter spatial scale.*
Temporal scope & scale:	`Description:` *Enter description of temporal scope.* `Scale:` *Enter temporal scale.*
Data volume estimate:	*Enter number.* *Select units.*
Restrictions:	`Use:` *Enter use limitations.* `Legal:` *Enter legal constraints.*
Contact:	`Form ID:` *Enter Contact Form ID of contact for this data collection.*
---------- To be completed for base data ----------	
Format:	*Select format.* *If other, enter format.*
Data description:	`Form ID:` *Enter Data Description form ID:*
Data Process/CQ Steps:	`Form ID:` *Enter Process Steps form ID:*
Keywords:	*Enter comma separated keywords.*
Data volume estimate:	*Enter number.* *Select units.*
Restrictions:	`Use:` *Enter use limitations.* `Legal:` *Enter legal constraints.*

ID [*Enter ID*]	
Contact:	`Form ID:` *Enter Contact Form ID of contact for this data collection.*
---------- To be completed when publishing to the public ----------	
Exclusive use period:	`Period:` *Select period of exclusive use.* `Justification:` *Enter justification for period of exclusive use.*
Repository(s) for data:	*Enter data repositories other than ASC.*
Citation:	*Enter data citation.*
Digital object identifier:	*Enter digital object identifier (doi).*
Contact:	`Form ID:` *Enter Contact Form ID of contact for this data collection.*

ASC EXISTING DATA FORM

Complete this form for existing collections include data that will be used for the purposes of creating the final data products and/or project deliverables. Remember to add the form Name: and ID: to the Project Table of Contents.

ID [*Enter ID*]	
Name:	*Enter existing data collection name.*
Description:	*Enter description for existing data collection.*
Source:	Name: *Enter name of data source.* Link: *Enter link to data source.*
Restrictions:	Use: *Enter use limitations.* Legal: *Enter legal constraints.*
Fees:	*Enter data fees.*
Format:	*Select format.* *If other, enter format.*
---------- To be completed after project has been approved for funding ----------	
Data Process Steps:	Form ID: *Enter Process Steps form ID:*
Spatial reference system:	*Select a spatial reference system.* *If other, enter reference system.*
Spatial scope & scale:	Description: *Enter description of spatial scope.* Scale: *Enter spatial scale.*
Temporal scope & scale:	Description: *Enter description of temporal scope.* Scale: *Enter temporal scale.*
Volume Estimate:	*Enter number.* *Select units.*
Acquisition Date:	*Enter date data was acquired from source.*
Citation:	*Enter data citation.*
Digital object identifier:	*Enter digital object identifier (doi).*
Contact:	Form ID: *Enter Contact Form ID of contact for this data collection.*

ASC DERIVED DATA PRODUCT FORM

Complete this form for each derived data product that will be published internally (USGS – ASC) or publically.

ID [*Enter ID*]	
Name:	*Enter data product name.*
Description:	*Enter description of data product.*
---------- To be completed after project has been approved for funding ----------	
Purpose:	*Enter summary of intentions for which the dataset was developed.*
Format:	*Select format.* *If other, enter format.*
Data description:	`Form ID:` *Enter Data Description form ID:*
Spatial reference system:	*Select a spatial reference system.* *If other, enter reference system.*
Spatial scope & scale:	`Description:` *Enter description of spatial scope.* `Scale:` *Enter spatial scale.*
Temporal scope & scale:	`Description:` *Enter description of temporal scope.* `Scale:` *Enter temporal scale.*
Keywords:	*Enter comma separated keywords.*
Data Derivation Steps:	`Form ID:` *Enter Process Steps form ID:*
Volume Estimate:	*Enter number.* *Select units.*
Maintenance frequency:	*Select maintenance frequency period.*
Restrictions:	`Use:` *Enter use limitations.* `Legal:` *Enter legal constraints.*
Contact:	`Form ID:` *Enter Contact Form ID of contact for this data collection.*
---------- To be completed when publishing to the public ----------	
Data Refinement Steps: (steps to prepare data for public distribution)	`Form ID:` *Enter Process Steps form ID:*
Deployment considerations:	`Distribution:` *Enter data distribution considerations and requirements.* `Web page:` *Enter web page consideration and requirements.*
Exclusive use period:	`Period:` *Select period of exclusive use.* `Justification:` *Enter justification for period of exclusive use.*
Repositories for data:	*Enter data repositories other than ASC.*

Citation:	*Enter data citation.*
Digital Object Identifier:	*Enter digital object identifier (doi).*
Contact:	`Form ID:` *Enter Contact Form ID of contact for this data collection.*

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.

ID [*Enter ID*]	
Name:	*Enter software product name.*
Description:	*Enter description of software product.*
---------- To be completed after project has been approved for funding ----------	
Source file link:	*Enter link to data file.*
Dependencies:	*Enter software dependencies.*
Maintenance & support:	Frequency: *Select maintenance frequency period.* Support Plan: *Enter software support plan.*
Languages:	☐ ASP.Net ☐ C# ☐ ESRI API ☐ Java ☐ JavaScript ☐ jQuery ☐ Python ☐ Rails ☐ Ruby Enter other language(s).
Environment:	*Select environment.* *If other, enter environment.*
Contact:	Form ID: *Enter Contact Form ID of contact for this data collection.*

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.

ID [*Enter ID*]	
Name:	*Enter model name.*
Description:	*Enter description of model.*
---------- To be completed after project has been approved for funding ----------	
Model version:	*Enter version of model.*
Source link:	*Enter link to data file.*
Model input(s):	*List model inputs.*
Model output(s):	*List model outputs.*
Spatial reference system:	*Select a spatial reference system.* *If other, enter reference system.*
Spatial scope & scale:	Description: *Enter description of spatial scope.* Scale: *Enter spatial scale.*
Temporal scope & scale:	Description: *Enter description of temporal scope.* Scale: *Enter temporal scale.*
Calibration details:	*Enter calibration details.*
Contact:	Form ID: *Enter Contact Form ID of contact for this data collection.*

ASC DATA DESCRIPTION FORM

This form is to be completed for each sheet in the spreadsheet, table in the database, and/or field collection form which is to be archived, distributed, or otherwise needs metadata prepared. Copy-paste 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 [*Enter ID*]	
Name:	*Enter dataset name.*
Description:	*Enter description of dataset.*
Location:	*Enter link to data source.*
Key Columns:	*Enter key columns for dataset.*
Column(s): (for each column)	Name: *Enter column name.* Definition: *Enter column definition.* Datatype: *Select datatype.* Other: *If other, enter datatype.* Units: *Enter data units.* Min Value: *Acceptable data minimum.* Max Value: Acceptable data maximum. Domain ID: *Enter Data Domain form ID*: QC Tests: *List QC tests performed.* Use Limits: *Enter use limitations.* Legal Const: *Enter legal constraints.* Name: *Enter column name.* Definition: *Enter column definition.* Datatype: *Select datatype.* Other: *If other, enter datatype.* Units: *Enter data units.* Min Value: *Acceptable data minimum.* Max Value: Acceptable data maximum. Domain ID: *Enter Data Domain form ID*: QC Tests: *List QC tests performed.* Use Limits: *Enter use limitations.* Legal Const: *Enter legal constraints.*

ASC DATA DOMAIN FORM

This form is to be completed for each data domain used in spreadsheets, data tables, and/or forms which are to be archived, distributed or for other reasons need metadata prepared. Copy-paste 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 [*Enter ID*]	
Name:	*Enter domain name.*
Description:	*Enter domain description..*
Domain value(s):	Value: *Enter domain value.* Definition: *Enter definition for domain value.* Value: *Enter domain value.* Definition: *Enter domain value.*

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 process step section 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 [*Enter ID*]	
Name:	*Enter procedure step name.*
Data acquisition methodology:	*Enter general methodology for obtaining data.*
Report conformance with process steps:	*Enter description of adherence to process steps.*
Process Steps(s):	Step ID: Enter procedure step ID. Description: Enter description of procedure. Contact ID: Enter Contact Form ID: Step ID: Enter procedure step ID. Description: Enter description of procedure. Contact ID: Enter Contact Form ID:

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 [*Enter ID*]	
Contact type:	*Select contact type.*
Organization role(s):	☐ funder ☐ contributor ☐ owner ☐ distributor ☐ custodian *Enter other role(s).*
Individual role(s):	☐ PI ☐ contributor ☐ custodian *Enter other role(s).*
Organization:	*Enter organization name.*
Name:	First Name: *Enter contact first name.* Middle Name: *Enter contact middle name.* Last Name: *Enter contact last name.* Preferred Name: *Enter contact preferred name.*
Title:	*Enter contact title.*
Address (mailing):	Address Line 1: *Enter first address line.* Address Line 2: *Enter second address line.* City: *Enter city.*, State: *Enter state.*, Zip: *Enter postal code.*
Phone(s):	Office: *Enter office phone number.* Mobile: *Enter mobile phone number.* SMS: *Enter SMS phone number.*
eMail address(s):	*Enter email address.* *Enter email address.*
Web link(s):	*Enter web links for contact.* *Enter web links for contact.*

GLOSSARY

- *base data*
  'base' data refers to raw data which have 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, and audio 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 product produced and published from base, derived, data from 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 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.

- *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/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://wwwimages.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