



NATIONAL CLIMATE CHANGE & WILDLIFE SCIENCE CENTER & CLIMATE SCIENCE CENTERS DATA MANAGEMENT PLAN GUIDANCE

Prepared by:

NCCWSC/CSC Data Management Working Group
US Geological Survey

April 21, 2014

Version 3

More Information: [NCCWSC Data Policies](#)

Point of Contact: Emily Fort, Data and IT Coordinator for the NCCWSC & CSCs, efort@usgs.gov

CONTENTS

Background	3
Data Management Plan Process	3
Why Prepare a Data Management Plan?.....	4
Data Management Plan Description and Guidance.....	4
Data Management Plan	5
Data Inputs – Existing Collections	5
Data Inputs – New Collections	6
Models	7
Custom Software/Code and Web Tools.....	7
Data Products (e.g., Deliverables).....	8
Example – Data Management Plan.....	9
Data Inputs – Existing Collections	9
Data Inputs – New Collections	11
Models	12
Custom Software/Code and Web Tools.....	12
Data Products (e.g., Deliverables).....	13

BACKGROUND

In order to provide guidance and support for data sharing, the NCCWSC/CSC Data Management Working Group has developed three documents. The first document, a [NCCWSC/CSC Data Sharing Policy](#), requires the development of a data management plan and also requires that data and associated products be publicly shared. The second document (this document), [NCCWSC/CSC Data Management Plan Guidance](#), provides details to researchers on what should be included in the data management plan that is required as part of the proposal and for funded projects. Finally, the third document, the [NCCWSC/CSC Data Management Manual](#) is intended to provide CSC Data Stewards and researchers funded by the NCCWSC or a CSC information on a variety of topics including roles and responsibilities, metadata, file formats, and data standards. Each CSC has a [Data Steward](#) to work with funded PIs on the development of the Data Management Plan, answer any questions related to data management, and assist with the transition of data and information products into the NCCWSC/CSC repository.

DATA MANAGEMENT PLAN PROCESS

The Data Management Plan (DMP) includes information about data used in projects including: 1) data inputs – existing data collections, 2) data inputs – new collections (e.g., new field data collections, new data generated, etc.), 3) models, 4) custom software/code and web tools, and 5) data products - expected datasets, code/software, and other products.

Project Stage	DMP Action Required
Statement of Intent	No DMP required.
Full Proposal	When a Statement of Intent (SOI) results in a request for a full proposal, a DMP is required to be submitted along with the proposal. The DMP requirements during this stage are limited to information needed to understand the proposal from a data perspective. The minimal required information is identified in the Data Management Plan template of this document in the non-shaded fields. Information in the shaded fields may also be provided if desired but is not required.
Receipt of Award	<p>If the proposal is selected for funding, all of the fields in the DMP must be completed within one month of the lead researcher receiving funding, including both shaded and non-shaded fields. Fields requiring information that is not known early in the project should be marked “TBD” (To Be Determined).</p> <p>The fully completed DMP includes details to ensure that data products are consistently formatted and documented during development, and that data deliverables are provided with maximum utility. The DMP will help the CSC Data Stewards assist the researcher with data management activities and make the project’s data and derived products available after project completion. The DMP should be periodically reviewed</p>

Project Stage	DMP Action Required
	and enhanced as the project evolves.
Project Completion	At project completion, the research team will work with their Data Steward to transition data (new data collections as well as final data products) and other products to the NCCWSC repository. All data and complete metadata descriptions must be provided to the Data Steward at project completion. For information the NCCWSC previously agreed not be made publicly available for a period of time or that is undergoing review (such as the USGS Fundamental Science Practices), data will be placed in a secure area until the period of exclusivity or review period has expired.

WHY PREPARE A DATA MANAGEMENT PLAN?

- Assists with visibility, reproducibility, and validity of research projects because data are well documented, including approach and methodology.
- Helps ensure data and data products are accessible and available for the long term.
- Consistent with the best practices. Initiates the process of gathering metadata and documentation throughout the project life cycle.

DATA MANAGEMENT PLAN DESCRIPTION AND GUIDANCE

Descriptions of the data in the DMP are divided into the following categories:

1. Data Inputs – Existing Collections;
2. Data Inputs – New Collections (e.g., data collected from the field, new data generated);
3. Models
4. Custom Software/Code and Web Tools; and
5. Data Products– Expected datasets, code/software, and other products.

Information should be provided for each element as appropriate. If a section is not appropriate (i.e., no new data are being collected), that should be noted and no additional information is necessary. If information is not known at the level of detail requested, as much information as possible should be provided. For multiple inputs and products, the data description table should be copied as many times as needed so that each table represents one input or one product.

The information provided in the DMP will be reviewed as part of the proposal process. When proposals are funded, information in the proposal will be used to create initial records for the project in the NCCWSC and CSC project tracking tool and data repository – ScienceBase. A CSC Data Steward will contact the principal investigator of each funded project to assist with the development of the DMP.

DATA MANAGEMENT PLAN

Notes:

1. Fields shaded gray are not required for proposals. If a project is funded, all fields are required.
2. Each input, model, custom software/code, or product should have its own table. Make a copy of the table for each distinct data input, model, custom software/code, or product.
3. In addition to the DMP, Data Stewards and NCCWSC/CSC Staff will work with the PI to obtain additional project information such as a project footprint, images, and publications.

Project Title: [Insert Project Title]

Lead PI:

Data Point(s) of Contact (if different):

DATA INPUTS – EXISTING COLLECTIONS

Existing collections include data that will be used for the purposes of creating the final data products and/or project deliverables. The collection(s) have already been obtained and have not been collected during the course of the project. Examples: PRISM data, USGS water data, remote sensing, etc.

1	[Enter Name of Existing Collection]
Description:	Describe the information that will be used, including its characteristics, temporal scope and scale, and geographic scope and scale, when available.
Source:	Identify the source for the data; include a link and digital object identifier (DOI) if available.
Restrictions:	Identify any limitations on access or reuse (e.g., sensitive data, restricted data, software with license restrictions, etc.) and provide justification for restriction. Provide citation or documentation describing limitations if due to policies or legal reasons.
Format:	Identify the formats in which the data are maintained and made available.
Fees:	Identify any fees associated with acquiring the data.
Quality Checks:	Identify the procedural steps used to evaluate the existing data, including verification, validation, and an assessment of usability.
Data Processing & Scientific Workflows:	Describe any data processing steps or provide a scientific workflow you plan to use to manipulate the data, as appropriate.
Backup & Storage:	Describe the approach for backup and storage of the information associated with the research project during the project.
Volume Estimate:	Estimate the volume of information that will be generated: megabyte (MB), GB, TB, or PB.
Citation:	Provide citation for data product. If the data product can be found online, provide a URL.

DATA INPUTS – NEW COLLECTIONS

Data that does not currently exist and will be collected or generated during the course of the project for the purposes of creating the final data products and/or project deliverables, for example, a new field data collection. New data collections must be delivered as a project deliverable at project completion and do not need to be added as a data product in the DMP.

1	[Provide a brief name to describe new data collection]
Description:	Describe the information that will be collected, including its characteristics, temporal scope and scale, and geographic scope and scale, when available.
Data Management Resources:	Describe the proposal resources allocated for data management activities for the new data collected as a level of effort, total dollars allocated, or as a percentage of the total project's cost. Resources could include people's time or proposal funding.
Exclusive Use:	Project data and associated products should be available publicly at the end of the project. If a request to limit access for a period of time after project completion is needed, please identify the length of time and the reason for the extension. (Request cannot be more than one year.)
Restrictions:	Identify any limitations on access or reuse (e.g., sensitive data, restricted data, software with license restrictions, etc.) and provide justification for restriction. Provide citation or documentation describing limitations if due to policies or legal reasons.
Format:	Identify the formats in which the data will be generated, maintained, and made available.
Protocols:	Identify any standard protocols or methodologies that will be used to collect the data, if available.
Quality Checks:	Identify the procedural steps for ensuring data quality.
Data Processing & Scientific Workflows:	Describe data processing steps or provide a scientific workflow you plan to use to manipulate the data, as appropriate.
Metadata:	Identify the metadata standard that will be used to describe the document (FGDC, ISO, EML, etc.)
Volume Estimate:	Estimate the volume of information generated: megabyte (MB), GB, TB, or PB.
Backup & Storage:	Describe the approach for backup and storage of the information associated with the research project during the project.
Repository for Data:	In addition to the NCCWSC repository (ScienceBase), identify any other repositories where you plan to share your data. Indicate if data will be integrated into an existing collection or offered as a new collection.
Citation:	Specify how the project's data should be cited.
Digital Object Identifier (DOI)/Link:	Provide a digital object identifier (DOI)/link to the data when available publicly.
Lifespan of Data	At some point, datasets may be archived. Choose one of the following options to indicate how long you anticipate this data will be of value to other researchers. Less than 5 years, 5-10 years, 10-20 years, 20-50 years, 50+ years.

MODELS

Describe the function and methodology used for any models that are part of the project. Any code developed to execute the model (if any was/will be developed by the project) should be described in the custom software/code section.

1	[Name of Model]
Description	Provide a brief description of the model and its purpose.
Model Version	Identify the version of model used.
Source/Link:	Provide a link or a source for the model.
Model Input(s)	Enter the types of input data required for driving, calibrating, or validating the model. For proposals, summary information is all that is needed. For funded projects, these should be described in detail in the existing or new collection, data inputs section.
Model Output(s)	Enter the types of output data the model will produce. For proposals, summary information is all that is needed. For funded projects, provide more details as known. If the model output is a generated dataset that is a project deliverable, describe it in detail in the data product section.
Calibration Details	Briefly describe the calibration/validation approach being taken.

CUSTOM SOFTWARE/CODE AND WEB TOOLS

Describe any custom software or code used as part of this project. If a web tool (e.g., visualization, decision support, etc.), is a project deliverable that should be included in this section.

1	[Name of Software or Other Need]
Description:	Describe any custom software or code developed or used, and/or any web tools being developed as part of the project.
Source/Link:	If the custom software or code can be accessed via an online repository, provide a link.
Restrictions:	Identify any limitations on access or reuse.
Maintenance and Support for the Web Tool	If a web tool is developed as part of the project, is there a strategy for the ongoing support and maintenance of the web tool after the project is complete? If so, briefly describe it.
Languages:	Identify the computing language/framework that was used (e.g., Java, .Net, Ruby, Rails, SQL, etc.)
Environment:	Identify the operating system environment (e.g., Windows, Linux, MacOS X, etc.)

DATA PRODUCTS (E.G., DELIVERABLES)

Identify project deliverables and data products that were developed as a result of the project's funding.

1	[Name of Data Product]
Description:	Describe the information that will be produced, including its characteristics, temporal scope and scale, and geographic scope and scale, when available.
Data Management Resources:	Describe the proposal resources allocated for data management activities for the data products as a level of effort, total dollars allocated, or as a percentage of the total project's cost. Resources could include people's time or proposal funding.
Format:	Identify the formats in which the data will be generated, maintained, and made available.
Exclusive Use:	Project data and associated products should be available publicly at the end of the project. If a request to limit access for a period of time after project completion is needed, please identify the length of time and the reason for the extension. (Request cannot be more than one year.)
Restrictions:	Identify any limitations on access or reuse (e.g., sensitive data, restricted data, software with license restrictions, etc.) and provide justification for restriction. Provide citation or documentation describing limitations if due to policies or legal reasons.
Quality Checks:	Identify the procedural steps for ensuring data quality during the project.
Data Processing & Scientific Workflows:	Describe data processing steps or provide a scientific workflow you plan to use to manipulate the data, as appropriate.
Metadata:	Identify the metadata standard that will be used to describe the data and products (FGDC, ISO, EML, etc.)
Volume Estimate:	Estimate the volume of information generated: megabyte (MB), GB, TB, or PB.
Backup & Storage:	Describe the approach for backup and storage of the information associated with the research project during the project.
Repository for Data:	In addition to the NCCWSC repository (ScienceBase), identify any other repositories where you plan to share your data.
Citation:	Specify how the project's data should be cited.
Digital Object Identifier (DOI)/Link:	Provide a digital object identifier (DOI)/link to the project when available publicly.
Lifespan of Data	At some point, datasets may be archived. Choose one of the following options to indicate how long you anticipate this data will be of value to other researchers. Less than 5 years, 5-10 years, 10-20 years, 20-50 years, 50+ years.

EXAMPLE – DATA MANAGEMENT PLAN

Project Title: Impacts of Climate Change and Precipitation Changes on Critter X in the Southwest (example)

Lead PI: Sarah Joe, sjoe@example_u.edu

Data Point(s) of Contact (if different): Jane Smith, jsmith@agency.gov

DATA INPUTS – EXISTING COLLECTIONS

Existing collections include data that will be used for the purposes of creating the final data products and/or project deliverables. The collection(s) have already been obtained and have not been collected during the course of the project. Examples: PRISM data, USGS water data, remote sensing, etc.

1	PRISM 4x4 km Gridded Monthly Climate Data
Description:	4x4 km gridded monthly Tmax, Tmin, and precipitation data that spans the US.
Source:	http://www.prism.oregonstate.edu/
Restrictions:	<p>The data and other information available on the prism.oregonstate.edu website are subject to the conditions stated below.</p> <p>Map Graphics. Map graphics may be freely reproduced and distributed. The PRISM Climate Group copyright notice, URL, and the date of map creation must be reproduced in full when the image is reproduced in any form, whether electronically or in hard copy. If an image is not provided with such a notice, one must be added in the following format: Copyright © <year>, PRISM Climate Group, Oregon State University, http://prism.oregonstate.edu Map created <date>.</p> <p>Data. All data (gridded, polygon, tabular, graphical) retrieved from the website or otherwise provided on the website may be freely reproduced and distributed for non-commercial purposes. When referring to the data, the source must be clearly and prominently stated and include, at a minimum, our name, URL, and the date of data creation. For example: PRISM Climate Group, Oregon State University, http://prism.oregonstate.edu, created 4 Feb 2004.</p>
Format:	http://prism.nacse.org/historical/ - various options for format including AcrGIS, Grid, etc.
Fees:	None
Quality Checks:	Review to make sure data downloaded accurately and completely.
Data Processing & Scientific Workflows:	Using VisTrails - http://www.vistrails.org/index.php/Main_Page
Backup & Storage:	During the project, data are backed up to lab server daily and to an external hard drive weekly
Volume Estimate:	100 GB
Citation:	PRISM Climate Group, Oregon State University, http://prism.oregonstate.edu , created 4 Feb 2004

2	Downscaled CMIP5 Climate and Hydrologic Data
Description:	Monthly downscaled (12x12 km gridded) temperature and precipitation data for the western US from select CMIP5 models historical simulations and 21 st century projections
Source:	http://gdo-dcp.ucllnl.org/downscaled_cmip_projections/
Restrictions:	Make appropriate citations
Format:	NetCDF or ASCII text
Fees:	None
Quality Checks:	Review to make sure data downloaded accurately and completely.
Data Processing & Scientific Workflows:	Model input
Backup & Storage:	Backed up to network server. Backed up weekly to external hard drive.
Volume Estimate:	10 TB (estimate)
Citation:	TBD

3	USGS Land Cover Data 2001, 2006, and 2011
Description:	NLCD describes the land cover of each 30-meter cell of land in the conterminous United States and identifies which ones have changed since the last update.
Source:	http://www.mrlc.gov/eva/viewer.html
Restrictions:	None
Format:	TIFF
Fees:	None
Quality Checks:	Review to make sure data downloaded accurately and completely.
Data Processing & Scientific Workflows:	Model input
Backup & Storage:	Backed up to network server. Backed up weekly to external hard drive.
Volume Estimate:	100 KB (estimate)
Citation:	TBD

DATA INPUTS – NEW COLLECTIONS

Data that does not currently exist and will be collected or generated during the course of the project for the purposes of creating the final data products and/or project deliverables, for example, a new field data collection. New data collections must be delivered as a project deliverable at project completion and do not need to be added as a data product in the DMP.

1	Population Monitoring of Critter X
Description:	Critter X is located in the mountain regions of Arizona and New Mexico. Every 3 months over the span of 2 years field data will be collected in 10 locations across the study region that reflect daily counts of critter X sightings.
Data Management Resources:	Salary of master’s student to collect the data. 50% of salary of a doctoral student to manage and analyze the data.
Exclusive Use:	None
Restrictions:	None
Format:	Spreadsheet
Protocols:	Following the protocols described here: http://www.fws.gov/sacramento/es/Survey-Protocols-Guidelines/es_survey.htm for Critter X.
Quality Checks:	Data entered by student who collects data. Data are reviewed and compared to field notebooks by another student. Both students check for outliers, consistency and enter basic metadata about data collection. Doctoral student conducts analysis containing quality checks and spots checks records against field notebooks.
Data Processing & Scientific Workflows:	Using VisTrails - http://www.vistrails.org/index.php/Main_Page
Metadata:	FGDC
Volume Estimate:	10 MB
Backup & Storage:	Backed up to server in lab and to external hard drive weekly.
Repository for Data:	In addition to the CSC Repository, ScienceBase, we plan to deposit our data in Dryad, http://datadryad.org/
Citation:	Joe, S. and J. Smith. 2016. <i>The Critter X Data Set. Version 1</i> . USGS ScienceBase. http://dx.doi.org/10.xxxx/notfoo.547983 . Accessed 1 May 2018.
Digital Object Identifier (DOI)/Link:	TBD
Lifespan of Data	50+ years

MODELS

Describe the function and methodology used for any models that are part of the project. Any code developed to execute the model (if any was/will be developed by the project) should be described in the custom software/code section.

1	SAHM VisTrails
Description	Software for Assisted Habitat Modeling Package for VisTrails (SAHM: VisTrails) v.1
Model Version	v.1
Source/Link:	https://www.fort.usgs.gov/products/23403
Model Input(s)	Climate, land use change, population survey of critter x
Model Output(s)	Habitat requirements of Critter X and predictions for its potential distribution based on habitat suitability
Calibration Details	Briefly describe the calibration/validation approach being taken.

CUSTOM SOFTWARE/CODE AND WEB TOOLS

Describe any custom software or code used as part of this project. If a web tool (e.g., visualization, decision support, etc.), is a project deliverable that should be included in this section.

1	Python Code
Description:	Python code developed as part of VisTrails work to ingest data inputs
Source/Link:	Posted to project git hub repository - https://github.com/
Restrictions:	None
Maintenance and Support for the Web Tool	N/A
Languages:	Python
Environment:	Windows

DATA PRODUCTS (E.G., DELIVERABLES)

Identify project deliverables and data products that were developed as a result of the project's funding.

1	Habitat Maps for Critter X
Description:	For Critter X's current and projected habitat, maps will be developed based on model output to show how critter x's habitat is projected to change due to climate and land use changes.
Data Management Resources:	50% of masters student's time to develop maps, 10% of post-doc to review and quality check
Format:	Shapefiles and web map services
Exclusive Use:	None
Restrictions:	None
Quality Checks:	Models and maps will be evaluated using standard criteria. Species distribution maps will be reviewed by project personnel and potentially wildlife managers.
Data Processing & Scientific Workflows:	Using VisTrails
Metadata:	FGDC
Volume Estimate:	1 TB
Backup & Storage:	Backed up to server in lab and to external hard drive weekly.
Repository for Data:	ScienceBase + Western Governor's CHAT http://www.westgovchat.org/map
Citation:	Joe, S. and J. Smith. 2016. Critter X Habitat Changes. Version 1. USGS ScienceBase. http://dx.doi.org/10.xxxx/notfoo.547983 . Accessed 1 May 2018.
Digital Object Identifier (DOI)/Link:	TBD
Lifespan of Data	50+ years.