Southern Rockies Landscape Conservation Cooperative Data/Information Management and Delivery Standards: FY15

Version 1.0, April 2015

1 Introduction

This document specifies the standard for data and information delivery for Southern Rockies Landscape Conservation Cooperative (SRLCC) science providers. It specifies project-level data management practices, data documentation, format standards, and product delivery processes. The standards are designed to ensure and facilitate full and open access to scientific data[1] and data products funded by the SRLCC. A well-developed data management strategy has mutual benefits for the SRLCC and the Principal Investigators (PI) of funded projects. Specifically, the SRLCC gains confidence that products will be delivered in a timely manner and in a format most useful to partners, resource managers, and the public. PIs will gain an understanding of expected product quality, documentation, and delivery format and process. The standards stipulated here are consistent with requirements of the National LCC Network, National Climate Change and Wildlife Science Center (NCCWSC), and the National Science Foundation (NSF).

Investigators developing a proposal for funding from the SRLCC are encouraged to review these standards. These standards are a binding condition for all SRLCC supported projects. Any variation from these standards must be requested in writing by the PI and agreed to by the SRLCC prior to project initiation.

The process for how SRLCC acquires science information is:

1. SRLCC Steering Committee identifies science needs
2. SRLCC makes a request for proposals or a funding opportunity announcement
3. Proposals are submitted by PIs

[1] Data may include "textual information, numeric information, instrumental readouts, equations, statistics, images (whether fixed or moving), diagrams, and audio recordings. It includes raw data, processed data, derived data, published data, physical samples, and archived data. It includes the 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. This definition of data also includes any custom code or applications that were developed to aid in data analysis or transformation and are necessary to understand the data. Code and applications must include adequate documentation and/or within code comments to understand the function."

4. SRLCC evaluates proposals
5. PIs are notified of awards and requirement to submit a Data Management Plan (DMP)
6. SRLCC reviews and approves DMP
7. Funds allocated
8. Science conducted; products generated and documented
9. Products delivered to SRLCC Science Coordinator and Data Manager
10. SRLCC Science Coordinator and Data Manager review and approve products
11. SRLCC disseminates products to partners, resource managers, and the public via the Landscape Conservation Management and Analysis Portal (LC MAP) and Conservation Planning Atlas (CPA)

2 Data Management Plan

When a proposal is identified for funding, a DMP must be delivered to and approved by the SRLCC before funds are provided to the PI. A key tool to foster quality data development and documentation, the DMP helps researchers and data managers:

- Think holistically about their project and data design, data needs, methodology, computational and analytical needs, documentation, data storage/archiving, and product delivery
- Develop realistic budgets regarding project's data management activities
- Implement standard data documentation practices at the outset of the project.

For SRLCC, a DMP:

- Assists with reproducibility and review of research projects because data and methods are well documented
- Helps ensure that data and data products acquired, and then delivered to partners and the public are: of high quality; well documented; discoverable and accessible; and available for the long term
- Is consistent with the best practices of other science funding programs such as the NCCWSC and NSF

Project proposals funded by the SRLCC must deliver a written DMP within 30 days of proposal acceptance. Approval of the DMP occurs before funds are provided to the PI. The DMP will:

- Address all aspects of the data lifecycle[2]: plan, collect, assure, describe, preserve, discover, integrate, analyze

[2] Data lifecycle example, http://www.dataone.org/best-practices

- Describe data inputs acquired from existing sources (provenance, documentation, and use restrictions)
- Anticipate the full array of data products generated using SRLCC funds including primary (i.e., field collected) and secondary (i.e., derived from analysis or modeling) data
- Describe how new data will be collected or existing data will be leveraged or reused including analytical tools and software
- Articulate quality assurance/quality control procedures
- Identify the metadata standard for all datasets
- Identify anticipated data formats and coordinate systems
- Describe a plan for long-term storage of samples and physical collections (if appropriate)
- Specify how and when data will be transferred to SRLCC custody
- If applicable, describe archiving, data delivery, and long-term maintenance measures

The DMP can be satisfied by using the template developed by NCCWSC. Use of the NCCWSC DMP template benefits project PIs and the SRLCC in the following ways:

1. PIs benefit by having a clear understanding of what information is required prior to funds being allocated.
2. The SRLCC benefits by receiving a DMP that can easily be transferred to their project life cycle tracking system.
3. Both PIs and the SRLCC benefit from the DMP being reviewed and approved in a timely manner.

Below are URLs to the NCCWSC DMP template as well as the USGS DMP Checklist that should be used during the process:

- DMP template provided by NCCWSC: https://nccwsc.usgs.gov/?q=node/15
- USGS Data Management Checklist: http://www.usgs.gov/datamanagement/plan/data-management-checklist.pdf

3 Data Development, Documentation, and Delivery

PIs are obligated to submit or make available to the SRLCC a copy of the raw data[1], derived data products, and other supported materials created or gathered in the course of work under SRLCC supported research. Release of data products into the public domain at the conclusion of the project is the *de facto* policy of the SRLCC. PIs are required to preserve and transfer their data and data products to SRLCC in accepted formats (see Data Formats below) needed for long-term science research. These standards do not supersede the legal requirements imposed upon organizations to restrict public access to data (*e.g.* culturally sensitive sites or the

locations of threatened or endangered species). However, such legal requirements restricting information and data access must be clearly stated in the project proposal, scope of work, and DMP.

3.1 Roles and Responsibilities

A. PIs shall be responsible for the quality, completeness, and description of the data, metadata, and associated products prior to final submission to the SRLCC.
B. Raw data shall be provided to the SRLCC as soon as possible after collection. The purpose of the SRLCC raw data archive is to protect against data loss. Raw data is not intended to be accessible by other researchers or the public. Upon transfer of raw data from PIs, the SRLCC becomes responsible for the long-term maintenance and eventual public access to the data.
C. PIs are responsible for delivering a copy of all data, appropriate metadata, and other supporting information to the SRLCC for archiving.
D. Upon transfer of data from investigators to the SRLCC, the SRLCC becomes responsible for providing the long-term maintenance and public access to this data. In cases where SRLCC cannot provide long-term maintenance and public access to SRLCC funded science data (*e.g.* terabytes of climate models), the PI, Science Coordinator, and Data Manager will arrange for data to be made available through a public web site, an institutional archive that is standard to a particular discipline or organization, or other approved repository. Intention to use this alternative approach to making data public and discoverable must be articulated in the DMP.

3.2 Data Delivery

E. All data and derived data products shall be submitted to SRLCC no later than 30 days after the conclusion of the project.
F. Conclusion of the project is the date the project contract ends. Where necessary, final payment will be withheld until all data and proper documentation have been turned over to SRLCC.

3.3 Special Cases

G. For projects producing observation sets greater than 5 years in duration and for long-term (>5 years duration) projects:
    i. The DMP should identify arrangements to make data publically available at intervals throughout the project life span starting in the second year of the project. Where applicable, a data sharing schedule should be described in the DMP.
    ii. The following recommended data sharing schedule has been approved by the SRLCC and could be followed:

- Data collected from January 1 to September 30 of a given year will be made publicly available by March 31 of the following year.
- Data collected from October 1 to December 31 of a given year will be made publicly available by June 30 of the following year.

4 Physical Specimens

PIs are responsible for depositing any samples, genetic material, and/or physical collections associated with their research in a recognized and approved repository or collection within their discipline. Where applicable, a sample or physical collection preservation plan shall be described in the project's DMP.

5 Proprietary Data and Software

PIs that will use proprietary data (*e.g.* data developed by a private organization), such that the terms of information release or types of data use are affected, should clearly state this in their proposal documents. The data restriction(s) need to be clearly documented in the proposal and DMP, which must clearly state what information, data, and conclusions cannot be released to the public upon conclusion of the project.

All data deemed sensitive, privileged, or subject to restricted access should be identified and appropriately labeled by the PI upon submission to the SRLCC. Policies for access to these data should be negotiated between the PIs and the SRLCC Coordinator or Science Coordinator, and documented in writing, prior to project implementation. Legal requirements restricting information and data access must be clearly stated in the project proposal and DMP. PIs need to clearly understand that federally funded projects are subject to federal laws that have been determined to override their personal or professional needs.

6 Data Formats

All project data and data products shall be delivered in formats approved by the SRLCC to ensure that they may be promptly released to the public upon delivery. LC MAP (ScienceBase) and the SRLCC Conservation Planning Atlas (Data Basin) are platforms utilized by the SRLCC to archive and disseminate project materials, so thus formats must be compatible with these systems. All formats shall be described in the DMP. The SRLCC recognizes that exceptions to these formats may be necessary. All exceptions to the following formats must be specified in the DMP and approved by the SRLCC. The following common data formats and spatial coordinate systems are accepted by the SRLCC:

A. Non-spatial Data and Derived Data Products

Deliverable Type	**Data Format**
Report/Publication	.pdf, .doc, .docx, .xml
Map	.pdf, .tif, .jpg
Database	.accdb, .xls, .xlsx, .csv
Image	.gif, .tif, .jpg, .bmp, .pdf

Metadata	.xml
Applications, Tools, Scripts	.py, .dll, .exe, .txt, ArcGIS Toolboxes/Tools/Models, IDL, etc. (describe in DMP)
All other deliverables	Specify in DMP

B. Spatial Data Products

Deliverable Type	**Data Format**
Vector	.shp, .gdb, .mdb, .sd
Raster	.gdb, .mdb, .tif (GeoTiff), ArcInfo Grid, ASCII text, .e00

C. Spatial Coordinate Systems

All data sets must be in a standard coordinate system, either geographic (*e.g.*, GCS) or projected (*e.g.* UTM, Albers, etc.) are accepted. All projections shall also be based on either the WGS84 or NAD83 datums. Coordinate systems for all raw and derived data need to be described in the DMP.

7 Metadata

Metadata[3] is required for all data sets and project products. A complete metadata record is required for the project as a whole (Project Metadata) and for each data product (Dataset Metadata) delivered. Content and format must follow a standard and widely recognized metadata protocol. We recommend the use of either the Federal Geospatial Data Committee Content Standard for Digital Geospatial Metadata (FGDC CSDGM) or International Standards Office (ISO) 19115/19119 protocols. If research reuses or leverages an existing data set, the metadata for research projects should cite the source data reference and provide a link to the data. Some sources for metadata creation and support include:

FGDC Geospatial Metadata Tools: http://www.fgdc.gov/metadata/geospatial-metadata-tools/
USGS Metadata Support: http://www.usgs.gov/datamanagement/describe/metadata.php
USGS Online Metadata Editor: http://mercury.ornl.gov/OME/
USGS Online Metadata Parser: http://geo-nsdi.er.usgs.gov/validation/
USGS Metadata Parser Software: http://geology.usgs.gov/tools/metadata/tools/doc/mp.html
EPA Metadata Editor: https://edg.epa.gov/EME/

[3] Simply defined, metadata is a set of data that describes and gives information about other data. In practice, a metadata record is a file of information, usually presented as an XML document, which captures the basic characteristics of a data or information resource. It defines and describes the who, what, when, where, why and how of the resource.