California Landscape Conservation Cooperative

Information Management, Delivery, and Sharing Standards

April 2014 Version

This document provides draft guidance on data and information delivery for the California Landscape Conservation Cooperative (CA LCC) science providers. It defines project-level data management practices, data documentation 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 CA LCC. This guidance is based on Draft Recommended Data Management Best Management Practices for Landscape Conservation Cooperatives and the guidelines developed by the Great Northern LCC.

These guidelines are being provided as part of the request for full proposals for the CA LCC and are currently under review by CA LCC subcommittees and subject to change. Upon approval, these policies may be considered a binding condition upon all CA LCC-supported projects. Compliance with this policy should be a key criteria used by the Review Team during proposal evaluation.

1. Data Management Plan

The purpose of a Data Management Plan is to:

- Assist with reproducibility of research projects because data are well documented.
- Help ensure data and data products are accessible and available for the long term.
- Be consistent with the best practices from many funding agencies such as the National Climate Change and Wildlife Science Center (NCCWSC) and National Science Foundation (NSF).

When a proposal is identified for funding by the CA LCC Steering Committee, a written Data Management Plan must be delivered and approved by the Coordinator and Science Coordinator within 3 months of proposal acceptance. The Data Management Plan (DMP) will:

- Address all aspects of the data life cycle;

[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."

- Describe proposed source data including its source, documentation, and use restrictions
- Anticipate the full array of data products generated using CA LCC 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;
- Articulate quality assurance/quality control procedures;
- Define the metadata standard for the data;
- Identify anticipated data formats;
- Describe plan for long-term storage of samples and physical collections (if appropriate);
- Specify how and when the data will be transferred to CA LCC custody; and
- If applicable, describe archiving, data delivery, and long-term maintenance measures.

The CA LCC has provided a Data Management Plan tool, adapted from the NCCWSC template, for use by all CA LCC funded projects. Each funded project must be described on the California Climate Commons, and must use the Data Management Plan tool on that site to create an associated DMP. The CA LCC data managers will assist with creating and maintaining the DMP, but primary responsibility resides with the PI(s). The process of creating, approving, and updating the DMP is described at http://climate.calcommons.org/article/CA-LCC-data-management; PI's should familiarize themselves with these expectations and budget for them.

2. Data Development, Documentation, and Delivery

Principal Investigators are expected to submit or make available to the CA LCC the raw data[1], derived data products, and other supporting materials created or gathered in the course of work under CA LCC-supported research. Release of these materials into the public domain at the conclusion of the project required. PI(s) are required to preserve and transfer their data and data products to the CA LCC in commonly accepted standards needed for long-term science research.

a. PI's shall be responsible for the quality, completeness, and description of the data, metadata and associated products prior to submitting to the CA LCC.

b. Raw data should be secured and archived as described in the Data Management Plan as soon as possible after its collection. The purpose of a raw data archive is to protect against data loss so the archive should have a tracking method and means of accessing those data by both PI and CA LCC staff. The intent is to ensure that those data are not destroyed or lost if the PI retires or changes jobs. If raw data is transferred from investigators to the CA LCC, the CA LCC becomes responsible for maintaining the raw data archive. The PI may choose to archive the data elsewhere; the replicate archival process the PI intents to use shall be described in the DMP or receive the approval of the CA LCC Science Coordinator.

c. All data and derived data products should be submitted to the CA LCC no later than 90 days after the conclusion of the project. Conclusion of the project is defined as the date

the project contract ends. Where necessary, final payment may be withheld until all data and proper documentation have been turned over to the CA LCC.

d. For those projects in which PIs have been granted initial periods of exclusive data use (see below), data should be made publically available as soon as possible, but no later than two years after the conclusion of the project. The period of exclusive use may be extended to three years for projects supporting work of a PI or Co-PI who is a matriculated student in a master's degree program or up to five years for projects supporting work of a PI or Co-PI who is a matriculated student in a doctoral degree program. The period of exclusive use should not be extended past the student's graduation date.

e. For projects producing observation sets greater than 5 years in duration and for long-term (>5 years duration) projects, arrangements should be made to make data publically available at intervals throughout the project life span starting in the second year of the project. 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.

f. Upon transfer of data from investigators to the CA LCC, the CA LCC becomes responsible for providing the long-term maintenance and public access to this data.

g. For data which has constraints such as file sizes or data types not supported by the capacity of the CA LCC, an alternative information clearinghouse may be arranged. In such cases the PI should arrange for data to be made available through a public web site, an institutional archive that is standard to a particular discipline or university, or through other approved repositories. If an alternative information clearinghouse is used, the PIs remain responsible for providing long-term maintenance and support for the data. In all cases, the PIs should still be responsible for delivering a copy of all data, appropriate metadata, and other supporting information to the CA LCC for archiving (ensuring that the CA LCC retains access to the information in the event of insolvency of the alternative information clearinghouse chosen by the PI to serve project data. Intention to use this alternative approach to making data public and discoverable must be indicated in the DMP.

3. Physical Specimens

Principal Investigators are responsible for depositing any samples, genetic material, and/or physical collections associated with their research in a recognized repository or collection within their discipline. Where applicable, a sample or physical collection preservation plan should be defined in the project's data management plan.

4. Proprietary Data and Software

Principal Investigators that will use or create proprietary data such that the terms of information release or types of data use are affected should clearly state this in their proposal documents. The requirements of data restriction should be documented in the proposal and data management plan, and must clearly state what information, data, and conclusions cannot be released to the public upon conclusion of the project.

a. All data deemed sensitive, privileged, or subject to restricted access should be identified and appropriately labeled by the PI upon submission to the CA LCC. Policies for access to these data should be negotiated between the PIs and the LCC Coordinator or Science Coordinator, and documented in writing, prior to project implementation.

b. This policy does not supersede the legal requirements imposed upon organizations to restrict public access to data. However, such legal requirements restricting information and data access must be clearly stated in the project pre-proposal, proposal and scope of work.

5. Metadata

Metadata should be required of all data sets. Metadata content and format will be determined on a project by project basis, and should be FGDC or ISO compliant. If research reuses or leverages an existing data set, the metadata for research projects should cite the source data reference and link to the data.

6. Exceptions

Deviations from this policy are possible, but any variation must be requested in writing by the PI and agreed to by the CA LCC Coordinator and Science Coordinator, working on behalf of the Steering Committee, prior to the implementation of the funded project.