

Data Management Best Practices for Landscape Conservation Cooperatives

Part 1: LCC Funded Science

Version 3.4, November 2012

LCC Network Data Management Working Group

Sean Finn, Josh Bradley, Emily Fort, Jennifer Jenkins, Pat Lineback, BJ Richardson, Ric Riestler, and Joel Reynolds

Introduction

This document recommends Best Practices for Landscape Conservation Cooperative (LCC) data¹ management and delivery. It is the first of three in a series of Data Management Best Practices documents. The objective of Part 1 is to ensure and facilitate transparent access to scientific data and data products funded by LCCs. Subsequent documents will describe data sharing practices amongst multiple LCCs and interactions with the broader science and information management communities. We suggest the Best Practices within these documents be adopted as standards by all 22 LCCs. We suggest the specific standards identified in this document, *Part 1: LCC Funded Science*, be required of all partners funded entirely or in part by an LCC. A well-developed data management strategy has mutual benefits for the LCC and the Principal Investigators (PI) of funded projects. Specifically, LCCs gain confidence that products will be delivered in a timely manner in a format most useful to partners, resource managers, and the public. PIs will have *a priori* understanding of expected product quality, documentation, and delivery format and process. The practices described herein are consistent with requirements of the National Climate Change and Wildlife Science Center and the National Science Foundation.

These standards should be made available to potential science partners at the time a request for science is released (i.e., language attached to or contained within a Request for Proposals). Compliance with the policy should be a key criteria used by the Steering Committee during proposal evaluation. We recommend these standards be considered a binding condition for all LCC-supported projects.

Situations may occur in which deviations from this policy are required. We recommend that any variation from these *Best Practice* standards must be requested in writing by the PI and agreed to by the LCC Coordinator and Science Coordinator, working on behalf of the Steering Committee, prior to the initiation of the funded project.

¹Data 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.”

Here we describe general Best Practices for LCCs operating within the LCC Network. Individual LCCs should adopt and maintain a specific plan which addresses how that LCC will manage those “internal” data generated by the LCC itself and data generated by funded projects. The plan should address all stages of the data lifecycle² and document the storage procedures, structure and methods associated with managing LCC-generated documents and data. It should outline standard operating procedures for archiving and publishing data and deliver a template for PIs to comply with those standards.

Data Management Plan

A key tool to foster quality data development and documentation is a Data Management Plan (DMP) which 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, and
- Implement standard data documentation practices at the outset of the project.

For LCCs, a Data Management Plan:

- Assists with reproducibility and review of research projects because data and methods are well documented.
- Helps ensure data and data products can be discovered and are accessible and available for the long term.
- Is consistent with the best practices from many science funding programs such as the National Climate Change and Wildlife Science Center (NCCWSC) and National Science Foundation (NSF).

Project proposals funded by the LCC should deliver a written DMP within three (3) months of proposal acceptance. Ideally, the DMP should be delivered and approved before funds are provided to the PI. However, if this process causes delay of project implementation, the Coordinator and PI may negotiate and alternate schedule. The DMP will:

- Address all aspects of the data life cycle²: plan, collect, assure, describe, preserve, discover, integrate analyze;
- Describe data inputs acquired from existing sources (provenance, documentation, and use restrictions);
- Anticipate the full array of data products generated using 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 including analytical tools and software;
- Articulate quality assurance/quality control procedures;

² See data lifecycle example, <http://www.dataone.org/best-practices>

- Define the metadata standard for all datasets;
- 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 LCC custody; and
- If applicable, describe archiving, data delivery, and long-term maintenance measures.

The Data Management Plan can be satisfied by using a number of resources. Example resources include:

- Template provided by National Climate Change and Wildlife Science Center (<https://nccwsc.usgs.gov/?q=node/15>)
- The UC3/DataONE Data Management Plan tool (<https://dmp.cdlib.org/>)
- [Data Delivery Standards and Specifications Template](#) being finalized by an interagency team and managed by USFWS Region 8. (U.S. Fish and Wildlife Service Sacramento, CA. 2012. Data Delivery Standards and Specifications Template; contact Pat_Lineback@fws.gov)

Data Development, Documentation, and Delivery

Principal Investigators are expected to submit or make available to the LCC a copy of the raw data¹, derived data products, and other supporting materials created or gathered in the course of work under LCC-supported research. Release of data products into the public domain at the conclusion of the project should be the *de facto* policy of the LCC. PI(s) are required to preserve and transfer their data and data products to LCCs in commonly accepted formats needed for long-term science research. The recommendations set forth in this document do 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 (where applicable), proposal, scope of work, and DMP.

Roles & 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 LCC.
- b. Raw data should be secured and archived as described in the DMP 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 LCC staff.
- c. PIs are responsible for delivering a copy of all data, appropriate metadata, and other supporting information to the LCC for archiving.
- d. Upon transfer of data from investigators to the LCC, the LCC becomes responsible for providing the long-term maintenance and public access to this data. In cases where the DMP describes a non-LCC website and data repository, it is the LCC's responsibility to approve that the site is appropriate, and to ensure that weblinks to those data portals are current on LCC websites. Intention to use this alternative approach to making data public and discoverable must be articulated in the DMP.

Data Delivery

- e. All data and derived data products shall be submitted to the LCC no later than 90 days after the conclusion of the project.
- f. Conclusion of the project is defined as the date the project contract ends. Where necessary, final payment should be withheld until all data and proper documentation have been turned over to the LCC.

Special Cases

- g. Projects that are inherently tied to a matriculating graduate student may be granted initial periods of exclusive data use. All exclusive data use agreements must be approved in writing at project startup.
 - i. The period of exclusive use may be extended to three (3) years total 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 (5) years total for projects supporting work of a PI or Co-PI who is a matriculated student in a doctoral degree program.
 - ii. The period of exclusive use should not be extended past the student's graduation date.
- h. For projects producing observation sets greater than 5 years in duration and for long-term (>5 years duration) projects:
 - i. Written arrangements should be made to make data publically available at intervals throughout the project life span starting in the second year of the project.
 - ii. The following data sharing schedule is recommended: 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.

Physical Specimens

Principal Investigators should be 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 should be defined in the project's DMP.

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 pre-proposal (where applicable), 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.

All data deemed sensitive, privileged, or subject to restricted access should be identified and appropriately labeled by the PI upon submission to the 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. Legal requirements restricting information and data access must be clearly stated in the project pre-proposal (where applicable), proposal and scope of work.

Metadata

Metadata³ 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, but other commonly used standards (i.e., Ecological Metadata Language [EML]) would be acceptable. 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. Some sources for metadata creation and support include:

FGDC Geospatial Metadata Tools: <http://www.fgdc.gov/metadata/geospatial-metadata-tools/>

LC MAP Metadata Tool: <https://www.sciencebase.gov/catalog/?community=LCCMAP>

USGS Online Metadata Editor: <http://mercury.ornl.gov/OME/>

EPA Metadata Editor: <https://edg.epa.gov/EME/>

mp Metadata Parser: <http://geology.usgs.gov/tools/metadata/tools/doc/mp.html>

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.