

Handout #03

## Appalachian LCC Data Management & Sharing Recommendations

*A draft presented for review by LCC Steering Committee*

Modified by Paul Leonard, GIS / Planning ([Paul.Leonard@fws.gov](mailto:Paul.Leonard@fws.gov), 413-345-0189) from recommendations from Appalachian LCC Data Management Working group, 'Guidance and Requirements for NCCWSC/CSC Data Management Plans', v 1.0-- Prepared by the CSC/NCCWSC Data Management Working Group, 'Data Management Best Practices for LCCs', v. 3.4 – Prepared by LCC Network Data Management Group.

### Introduction

#### Background

The Appalachian Landscape Conservation Cooperative (Appalachian LCC or AppLCC) has been funding projects since 2012. Data Management Plans (DMPs) have not been a requirement with project proposals (e.g., National Climate Change and Wildlife Science Center DMP template<sup>1</sup>). This requirement (as suggested by National LCC guidance<sup>2</sup>) will demand additional workflows, protocols, and/or a infrastructure for the collection, management, and sharing of AppLCC metadata and research products. Such a cohesive system of workflows and infrastructure is necessary to implement the best-practice, scalable concept of the “data life cycle” which describes the data management process from project (or organization) inception to completion (Figure 1).

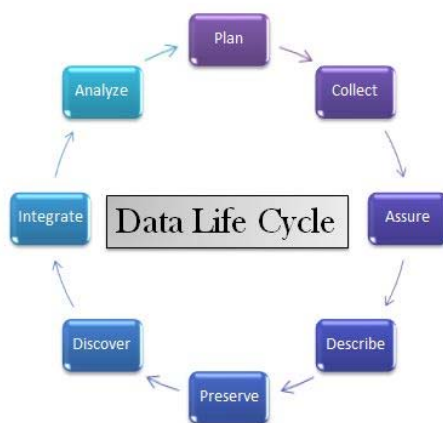


Figure 1. Adapted from: 'Data life cycle' NSF-supported Data Observation Network for Earth (DataONE).

<sup>1</sup> NCCWSC Data Management Plan Guidance template accessed February 10, 2014 at <https://nccwsc.usgs.gov/?q=node/15>.

<sup>2</sup> Data Management Best Practices for Landscape Conservation Cooperatives. Accessed February 10, 2014 at <https://my.usgs.gov/confluence/download/attachments/411074828/LCC->

[DataMgmt\\_Best\\_Practices\\_Part1\\_v3.4.pdf?version=1&modificationDate=1377820229539&api=v2](#)

To document and disseminate successes and findings of AppLCC-supported science, there is a need for improved sharing of information with the public, stake-holders, and decision-makers. To that end, the App LCC Data Management Working Group has developed this set of Data Management Recommendations which outline specific workflows, protocols and infrastructure recommendations to document and share LCC science with stakeholders and the public.

## **Objectives for Management**

The four primary objectives for App LCC Data Management are development of a plan, workflows, infrastructure, and data delivery. The DWG recommends:

1. Review and ratify the National LCC best practices (Appendix A.) as an App LCC Data Management Plan.
2. Establish workflows for documenting and delivering information about LCC-funded projects and products.
3. Provide a platform for LCC stakeholders to be able to store, view, manipulate, and collaborate on GIS data (for example, using Amazon cloud and/or geonode environment) with training materials.
4. Deliver project metadata and data products to the public.

To achieve these objectives, it is recommended that the LCC leverage existing workflows, protocols and platforms. With respect to data management planning, The National Science Foundation (NSF) and NCCWSC have developed proven data management guidance and templates. NCCWSC has also invested in development of the myUSGS RFP Manager (<https://my.usgs.gov/rfpManager/>) to manage metadata including research interest and intent, proposed work, and data management plans. A myUSGS Data Management Plan Manager is currently under development with guidance from NCCWSC to make it easier for principal investigators to create and submit data management plans in the RFP process.

## **Recommendations for Management**

This section outlines recommended steps to achieve the App LCC data management objectives. These recommendations emphasize proposals, products and resources:

- **Proposals:** Reflect LCC data sharing priorities in proposal scoring. Automate the collection of project metadata from the proposal process. Provide templates and support for well-formed Data Management Plans for approved projects. Build checks and consequences into the process of DMP submission.
- **Products:** Develop and implement standards for documentation and delivery of products.
- **Resources:** Provide support for staff and resources to achieve data management objectives.

## **Proposals: Where data management begins**

Good data management begins with the RFP process. This is an opportunity to select projects with a strong data sharing component to support good science. The metadata acquisition process begins at this stage with collection of project metadata as well as a well-formed data management plan (DMP). The DWG recommends the following related to the RFP process:

- Automate population of project metadata into Geonode Server/Amazon Cloud from RFP projects and workflows. Options include establishing an agreed-upon manual metadata

transfer process initially while exploring options to automate the process.

- Provide support to selected projects in completing a well-formed DMP and allow for review and revision of DMPs in the RFP process. Provide templates and support for well-formed Data Management Plans for approved projects (Appendix B). Build checks and consequences into the process of DMP submission.

## **Products: Metadata**

The purpose of metadata is to make sure data can be translated. It enables re-use and comprehension of the data being described; therefore, every data product should have a metadata record. The DWG recommends the following with respect to metadata:

- Standards development
  - Recommend using the North American Profile of ISO 19115 metadata standard
  - Develop standards for completeness of metadata. Determine which fields are required. For example, require information about data access constraints
  - When delivering GIS products, require information about how to use and symbolize it to prevent misuse or misrepresentation
- Metadata entry
  - For projects, automate metadata entry from the RFP process
- Make tools and training available
  - Make available a list of free tools for metadata creation
  - Provide training and support for metadata creation
- Metadata delivery and linkage with data
  - Require online linkage to product from metadata record or contact information to obtain data
  - Make metadata records available via (1) Project Profile pages on the App LCC website, (2) Geonode Server web services, and (3) Data.gov

## **Resources: Support to implement recommendations**

While existing products and tools will be leveraged to the maximum extent possible to implement the recommendations outlined in this document, investment in dedicated personnel and support will be required to develop and implement data management infrastructure and workflows for App LCC. As a starting point, the following is an initial and by no means comprehensive list of required resources:

- Platform administrators: At least one staff at 50% time who quality check data entry, user accounts, access levels, and overall data curation workflows for each platform.
- Data curators: One staff member at 25% time to check data for relevancy and completeness prior to addition to the AppLCC Data Catalog.
- Training and outreach: One staff member at 25% time to develop documentation and conduct training on a regular basis. Support for a common web platform accessible by staff to conduct online trainings such as a App LCC Webex site.

**Recommendations for Data Sharing:** the primary objective of data sharing is to make products as widely available as possible following these guiding principals:

1. Provide clear guidance for data sensitivities on a case by case basis.
2. Develop contracts for obtaining sensitive data as needed by projects (e.g., Nature Serve) and promote a national-level effort to streamline this process
3. Maintain an archive of data products but not responsible for the maintenance of datasets.
4. Develop map services to share spatial data with other LCCs (e.g., Databasin) and the public.

## Appendix A.

### 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 Riester, and Joel Reynolds

#### Introduction

This document recommends Best Practices for Landscape Conservation Cooperative (LCC) data<sup>1</sup> 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.

---

<sup>1</sup>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.”

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 lifecycle<sup>2</sup> 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 cycle<sup>2</sup>: 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;

---

<sup>2</sup> 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](mailto: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 data<sup>1</sup>, 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

Metadata<sup>3</sup> 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>

---

<sup>3</sup> 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.

---

## Appendix B.

### Appalachian LCC DATA MANAGEMENT PLAN TEMPLATE

*Based on the NCCWSC/CSC Data Management Guidance, February 26, 2013 update.*

#### Notes:

- Each output should have its own table. Make a copy of the table for each distinct data output.
- An example of a well-documented Data Management Plan is located at <https://nccwsc.usgs.gov/sites/default/files/files/Example DMP.pdf>.

Project Title: [Insert Project Title]

#### Data Inputs – Existing Collections

| 1                                       | [Name of Collection]   |
|---|--|
| Description:                            | Describe the information that will be used and the nature and scale (e.g., national, regional, landscape, etc.) of the data. Include a link to the source of the existing data.  |
| Format:                                 | Identify the formats in which the data are maintained and made available.  |
| Quality Checks:                         | Specify the procedures used to evaluate the existing data, including verification, validation, and an assessment of usability.   |
| Source:                                 | Identify the source for the data.  |
| 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.   |
| Access & Sharing:                       | Prior to the completion of the project, specify who should have access to project information/products and what type of access (Public, Read, Write, No Access).   |
| 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. |
| Fees:                                   | Identify any fees associated with acquiring the data.  |
| 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. For example, a new field data collection.)**

| 1                                       | [Provide a brief name to describe new data collection]   |
|---|--|
| Description:                            | Describe the information that will be used and the nature and scale (e.g., national, regional, landscape, etc.) of the data that will be collected.  |
| 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.                                      |
| Format:                                 | Identify the formats in which the data will be generated, maintained, and made available.  |
| Data Processing & Scientific Workflows: | Describe data processing steps or provide a scientific workflow you plan to use to manipulate the data, as appropriate.  |
| Protocols:                              | Identify any standard protocols or methodologies that will be used to collect the data, if available.  |
| Quality Checks:                         | Specify the procedures for ensuring data quality.  |
| 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 Appalachian LCC repository, identify any other repositories where you plan to share your data.  |
| Access & Sharing:                       | Prior to the completion of the project, specify who should have access to project information/products and what type of access (Public, Read, Write, No Access).   |
| Exclusive Use:                          | Project data and associated products should be available publically 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 two years.) |
| 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.                             |
| Citation:                               | Specify how the project’s data should be cited.  |
| Contact:                                | Provide a point(s) of contact if questions arise related to the data and associated products (name, email, and phone number).  |

## Software and Other Needs

| 1             | [Name of Software or Other Need]  |
|---------------|---|
| Description:  | Describe any software or other needs that are required for the project. Software such as Microsoft Office, Adobe, and an Internet Browser do not need to be provided. |
| Restrictions: | Identify any limitations on access or reuse that accompany the software or other needed items.  |
| Fees:         | Identify any fees or other costs associated with acquiring the software or other items.   |
| Source/Link:  | Provide a link or a source for the need if available.   |

## Data Outputs (e.g., Project Deliverables or Products)

| 1                                       | [Name of Output]   |
|---|--|
| Description:                            | Describe the data output.  |
| 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.                                      |
| Format:                                 | Identify the formats in which the data will be generated, maintained, and made available.  |
| Data Processing & Scientific Workflows: | Describe data processing steps or provide a scientific workflow you plan to use to manipulate the data, as appropriate.  |
| Quality Checks:                         | Specify the procedures for ensuring data quality during the project.   |
| 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 Appalachian LCC repository, identify any other repositories where you plan to share your data.  |
| Access & Sharing:                       | Prior to the completion of the project, specify who should have access to project information/products and what type of access (Public, Read, Write, No Access).   |
| Exclusive Use:                          | Project data and associated products should be available publically 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 two years.) |
| 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.                             |
| 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 publically.   |
| Contact:                                | Provide a point(s) of contact if questions arise related to the data and associated products (name, email, and phone number).  |