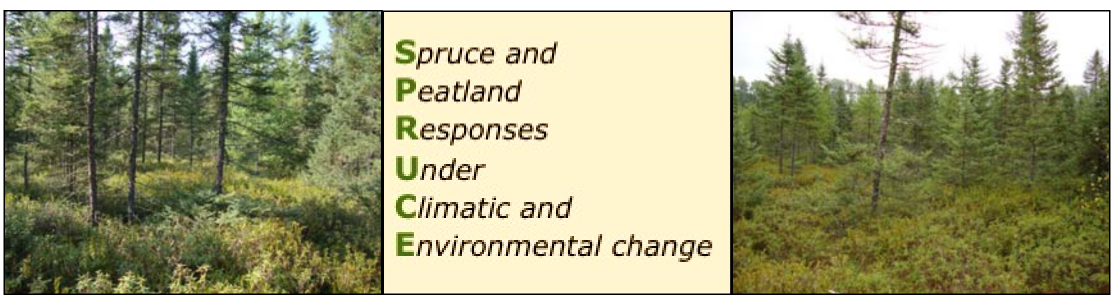
Data Management Plan for SPRUCE

Spruce and Peatland Responses under Climatic and Environmental Change

Prepared by Les A. Hook

Environmental Data Science & Systems Group
Environmental Science Division
Oak Ridge National Laboratory

Oak Ridge, Tennessee



Research sponsored by the Office of Biological and Environmental Research within the U.S. Department of Energy's Office of Science

The SPRUCE experiment is a multi-year cooperative interaction among scientists of the Oak Ridge National Laboratory operated by UT-Battelle, LLC and the U.S. Forest Service, Northern Research Station, Marcell Experimental Forest

Contents

- Creating Archive Documentation
- Providing Metadata to Searchable Indexes and Clearinghouses
- Assigning Descriptive Data Set Titles

Data Systems Management

- Day-to-Day Operation of Data Management Systems
- Data Management System and Software Configuration Control Guidelines

Appendices

Appendix XW: Tabular data file format guidance

Appendix XX: Reporting Photos -- Forms

Appendix XY: SPRUCE Task Data Descriptions -- Metadata and Documentation

Appendix XZ: Example CDIAC "Mercury" Search Metadata Report

Data Management Plan for SPRUCE

Contact: Les Hook, hookla@ornl.gov, 865-241-4846

Data Management Plan Revision Log

Date	Version	Action
2010/03/31	**0.1**	**Initial release of draft plan for review by SPRUCE Team.**
2010/05/10	**0.2**	**Second release of draft plan for review by SPRUCE Team.. Includes updates to Data Policy.**
2010/06/01	**1.0**	**Release of Version 1.0 as operational version.**

Data Management Plan for SPRUCE

Introduction

Overview

The SPRUCE project (Spruce and Peatland Responses under Climatic and Environmental Change) is implementing an experimental platform for the long-term testing of the mechanisms controlling the vulnerability of organisms, ecosystems, and ecosystem functions to increases in temperature and exposure to elevated CO_2 treatments within the northern peatland high-carbon ecosystem.

This experiment is a major long-term investment in field facilities, staff, measurements, and observations over the next 10 to 12 years by the Terrestrial Ecosystem Science Program of the DOE Office of Science as a major component of its climate change science initiatives.

As such, the implementation of the experiment and application of the results to answer the science questions will have high visibility within DOE and the broader climate change science community. It is one of the main research activities of the ORNL Climate Change Science Institute.

It is important to all parties that this collaborative effort between ORNL (UT-Battelle) and the U.S. Forest Service, Northern Research Station, at the Marcell Experimental Forest is a success. Additional external investigators will join this Team as the effort matures.

Goals for SPRUCE Data Management

- Ensure the fidelity of and accessibility of SPRUCE data to the participants to facilitate all the pertinent science questions;
- Minimize the amount of time research personnel need to spend on data management activities while achieving high quality data and metadata; and
- Ensure that the data and metadata can be located and used by project personnel (initially) and the broader scientific community and public when appropriate quality checked data are available.

Approach to Data Management Planning

This data management plan will provide a structured framework to capture the project-defined requirements / needs for maintaining data quality and consistency and controlling data processing from collection through archiving and public access. The plan will provide data management guidance and best practices, but it's up to the ORNL SPRUCE research group, the Task Leaders in particular, and Forest Service Staff, to reach a consensus about what needs to be controlled, to provide processing details, and to establish who is responsible for implementation. Accountability is key.

Planning Considerations

- This plan identifies the resources and data management topics needed to support SPRUCE data collection activities. The plan provides a framework to support field sampling, measurements, monitoring, and analyses that will follow over the experimental study years.
- During the pre-experimental period in 2010 and 2011, field work will have an exploratory emphasis, leading to final plans for sampling and data collections starting with time zero observations in the summer of 2011. Experimental measurements will be planned, sampling methods, and analytical techniques will be evaluated during this period.
- Data management information collected pre-experiment will inform the “final” experimental data management processes described in general terms in this document.
- The planning process will accommodate SPRUCE tasks that are subject to change or modification, and must allow for added tasks. The experimental technology will evolve over the duration of this multi-year study, and the data management plan will have to be flexible and updated as needed with version control.
- The data management plan is being prepared as a web accessible document, with pages/sections suitable for printing.

Initial Data Management Planning

Planning has focused on the following key elements. Details have been addressed to varying degrees as the SPRUCE plans and needs have been identified. Items not dealt with in this data management plan release will be taken up in subsequent revisions as information becomes available. Key steps in the development of this plan include:

- defining the data collection tasks,
- identifying existing site monitoring, GIS, and remote sensing data,
- identifying and ensuring that site characterization data are properly maintained,
- identifying and ensuring that spatial environmental monitoring data are properly maintained,
- ensuring that experimental data are properly archived and distributed according to the SPRUCE Data Policy,
- identifying possible coordination with other ORNL SFAs and projects, and
- resolving critical informatics knowledge gaps identified in the requirements definition.

It is important to recognize during the planning process and subsequent implementation that

- SPRUCE data collection activities will naturally evolve over the life of the project,
- collaborations with other Climate SFAs and ORNL projects will evolve as their data needs change, and
- data management and informatics technology components may need to evolve to meet these needs.

Planning and Implementation Resources

During the data management planning process SPRUCE is providing funds to allow expertise and tools of the Environmental Data Science and Systems (EDSS) Group to be applied to SPRUCE for the development of data quality and archiving protocols, web site development, and metadata creation. As the SPRUCE project archive develops, the Carbon Dioxide Information and Analysis Center (CDIAC) will ensure that observational and automated data are archived and publicly accessible, and promote data accessibility, use, and analysis. Further details can be found in Annex C of the Science Plan.

SPRUCE Project Overview

SPRUCE Project Description

SPRUCE is an experiment to assess the response of northern peatland ecosystems to increases in temperature and exposures to elevated atmospheric CO_2 concentrations.

The SPRUCE experiment is a key experimental component of ORNL's Climate Change Science Institute. The experiment focuses on terrestrial ecosystems and the mechanisms that underlie their responses to climatic change. The experimental work is to be conducted in a *Picea mariana* [black spruce] – *Sphagnum* spp. bog forest in northern Minnesota, 40 km north of Grand Rapids, in the USDA Forest Service Marcell Experimental Forest (MEF). The site is located at the southern margin of the boreal peatland forest. It is an ecosystem considered especially vulnerable to climate change, and anticipated to be near its tipping point with respect to climate change. Responses to warming and interactions with increased atmospheric CO2 concentration are anticipated to have important feedbacks on the atmosphere and climate, because of the high carbon stocks harbored by such ecosystems. Experimental work in the 8.1-ha S1 bog will be a climate change manipulation focusing on the combined responses to multiple levels of warming at ambient or elevated CO2 (eCO2) levels. The experiment provides a platform for testing mechanisms controlling the vulnerability of organisms, biogeochemical processes and ecosystems to climatic change (e.g., thresholds for organism decline or mortality, limitations to regeneration, biogeochemical limitations to productivity, the cycling and release of CO2 and CH4 to the atmosphere).
The manipulation will evaluate the response of the existing biological communities to a range of warming levels from ambient to +9°C, provided via large, modified open-top chambers. The ambient and +9°C warming treatments will also be conducted at eCO2 (in the range of 800 to 900 ppm). Both direct and indirect effects of these experimental perturbations will be analyzed to develop and refine models needed for full Earth system analyses.

The complete SPRUCE Science Plan (http://mnspruce.ornl.gov/CCP_ResponseSFA_PlanExp.pdf) and additional information are available on the SPRUCE website: http://mnspruce.ornl.gov/ .

Research Timeline

Following is a general timeline for implementation of the Science Plan

RESEARCH TIMELINE

(Bold items are those that need to be considered in the schedule for Data Management Planning.)

FY 2010 (with some scoping activities in FY 2009)
Summer 2009 – **Inventory *Picea mariana* basal area across S1, conduct a peat depth survey, initiate water-level observations to characterize local water table geometry, heterogeneity and seasonal variation to enable the determination of optimum locations for treatment blocks within the S1 bog.**

Oct 2009 – Finalize the ORNL/USFS Memorandum of Understanding.
Nov 2009 – Initiate National Environmental Policy Act (NEPA) process.
Jan 2010 – Peat depth surveys under snow cover.
Dec 2009 to June 2010 – **Establish and test operational aboveground 12-m prototype at ORNL.**
Winter 2009 through 2010 – **Evaluate pre-treatment Picea mariana characteristics**.
April 2010 – **Initiate continuous environmental monitoring on ambient plots.**
May to Aug 2010 **– Initiate the collection of baseline vegetation data**.
Jun 2010 to Jan 2011 – **Conduct allometric evaluations on Picea and shrubs**.
Sep 2010 – Complete NEPA Process.
Sep 2010 – Complete experimental engineering plans and diagrams.

FY 2011
Oct - December 2010 Bring electrical power to the S1 site.
Dec 2010 through March 2011 – Install main access boardwalks for each treatment block.
November through September 2011 – Establish and test a belowground corral prototype in a non-critical area of the S1 bog following NEPA approvals.
October 2010 to Mar 2011 – Experimental construction.
Jun 2011 – Lease and locate movable office/storage space.
Jul 2011 – Add concrete pad for the CO2 and propane storage tanks.
All year – **Continue environmental data collection.**
Apr to Oct 2011 – **Collect pretreatment biological observations.**

FY 2012
Dec 2011 to Mar 2012 – Experimental construction.
May 2012 – Complete construction of all above- and belowground infrastructure.
Jun 2012 – Bring in CO2 and propane tanks and test all systems.
Apr to Oct 2011 – **Collect pretreatment biological observations.**
Sep 2012 – Initiate temperature and CO2 treatment.

SOURCE: Response SFA Science Plan. Last updated 2010/04/29

Data Management Planning

Organization

- Data Policy
- Data Flow
- Project Name Information
- Identifying Measurement and Sampling Sites

Data Policy

The development of the policy involved current project participants and data users as well as potential data users through long-term data archive representation and planning. A clear statement of the importance of the data collection effort and of the flow of data and information before, during, and after the current activities was formulated in the broadest possible context. It is a shared responsibility of all participants to implement the data policy.

Components of the Data Policy are represented in the Data Flow diagram that follows.

Version 1.1 20100430

SPRUCE Data Policy: Archiving, Sharing, and Fair-Use

The open sharing of all SPRUCE experiment data among researchers, the broader scientific community, and the public is critical to advancing the mission of DOE's Program of Terrestrial Ecosystem Science.

SPRUCE is implementing an experimental platform for the long-term testing of the mechanisms controlling the vulnerability of organisms, ecosystems, and ecosystem functions to increases in temperature and exposure to elevated CO2 treatments within the northern peatland high-carbon ecosystem. All data collected at the SPRUCE facility, all results of any analysis or synthesis of information, and all model algorithms and codes developed in support of SPRUCE will be submitted to the SPRUCE Data Archive in a timely manner such that data will be available for use by SPRUCE researchers and, following publication, the public.

This policy is applicable to all SPRUCE participants including the SPRUCE Research Group at the Oak Ridge National Laboratory (ORNL), the U.S. Forest Service, cooperating independent researchers, and to the users of SPRUCE data products (see the Data Fair-Use Statement).

SPRUCE data policies are consistent with the sponsoring U.S. DOE Program for Terrestrial Ecosystem Science Data Policy and with the Memorandum of Understanding between the U.S. Forest Service and UT-Battelle.

Data Archiving and Discovery

The SPRUCE Data Archive will be maintained by the Carbon Dioxide Information Analysis Center (CDIAC) at ORNL. CDIAC provides the long-term system stability and archive longevity required for the SPRUCE multi-year project, and reliable public data access.

The Archive will maintain two levels of data accessibility. The first is for sharing recently collected, derived, and processed data products among SPRUCE participants including the ORNL research group, Forest Service, and related SFAs. The second is for access to mature data products by the broader scientific community and public. Public access will be concurrent with open literature or web site publication of SPRUCE results.

The discovery (identifying and finding) of SPRUCE data sets, derived products, synthesis results, and models (inputs, outputs, and codes) by the scientific community and public will be facilitated through the compilation of descriptive companion metadata records and their inclusion in searchable metadata databases and clearinghouses.

The Carbon Dioxide Information Analysis Center (CDIAC) is the primary climate-change data and information analysis center of the U.S. DOE. CDIAC is located at DOE's Oak Ridge National Laboratory.

Data Sharing

Timeliness of Data Availability

SPRUCE researchers will actively process, quality assure, and document environmental measurements, experimental data, observations, and modeling results and submit them to the Data Archive in a timely manner. Initially these data products will be shared among SPRUCE participants and subsequently with the public.

The diverse set of planned measurement tasks vary greatly in their temporal measurement frequency, ranging from, for example, 30 minute averages of 1 minute air temperature measurements to annual aboveground vegetation measurements. The amount of processing and analysis effort needed to create a given product varies accordingly.

To identify a reasonable time for processing and quality assurance of data while maintaining accountability for submitting data, each SPRUCE Task Leader will define a schedule for submitting data to the Archive for their given measurements.

Suggested guidelines for submitting data to the Archive for sharing among SPRUCE participants.

These are representative tasks and suggested timeframes:
- Environmental measurements (automated instruments) -- 30 days after the completion of a month of measurements
- Annual surveys and seasonal measurement efforts -- 120 days from the completion of the survey
- Laboratory analyses of vegetation nutrient concentrations -- 60 days from completion of analyses

Suggested guidelines for submitting data to the Archive for public access.

These are representative tasks and suggested timeframes:
- Environmental measurements (automated instruments) -- annual updates
- Annual surveys and seasonal measurement efforts -- With publication of papers.
- Laboratory analyses of vegetation nutrient concentrations -- With publication of papers.

Quality Assurance of Data

Data products will be submitted to the Archive, initially for sharing among SPRUCE participants, and subsequently for access by the scientific community and public.

The level of quality assurance needed for sharing newly collected, derived, and processed data among SPRUCE participants for information and confirmation purposes is typically not as great as that for publication and public access. By defining the quality level of the data needed for each type of access, it will be clear when data products are ready to be shared and the users will know what quality checks have been performed.

Each SPRUCE Task Leader will define the quality assurance checks to be performed prior to data sharing among SPRUCE participants (Quality Level 1) and then prior to public access (Quality Level 2). When data products have been updated as a result of additional quality checks or discovery of errors, the data should be resubmitted to the Archive and the quality level documentation changed (e.g., to Level 2).

Suggested guidelines for defining data Quality Levels.

Level 1: Indicates an internally consistent data product that has been subjected to quality checks and data management procedures. For example:
- Site documentation has been reviewed for completeness.
- Procedures and protocols were reviewed for compliance.
- Calibrations and quality control samples have been evaluated and necessary corrections made.
- The data have been adjusted for "zero drift" (continuous measurements), or for "blank bias" (lab analyses) as appropriate.
- Consistency checks have been performed with other measurements within the same data file. These internal consistency checks might include diurnal analyses to look for expected patterns, or time series analyses to detect outliers, extreme values, or time periods with too little or too much variation.

Level 2: Indicates a complete, externally consistent data product that has undergone interpretative and diagnostic analysis by the SPRUCE participants. For example, in addition to Level 1 procedures:

- Data have been closely examined by the data manager and/or data users for external consistency when compared to other related data.
- External checks might include correlation by scattergram, comparison of data with other similar data for the same time period, and comparison of a measurement made by two different methods.
- If comparisons were not within the precision of the measurements, then measurement records and other information have been reviewed.
- More? Other?

For completeness, **Quality Level 0** data are products of unspecified quality that have been subjected to minimal processing in the field and/or in the laboratory (e.g., raw data, photos, hardcopy data sheets, scanned data sheets, notebooks, etc.). This may, for example, be data from an instrument logger expressed in engineering units or using nominal calibrations, or high resolution data before aggregating to a selected interval. These products should be submitted to the Data Archive for long-term storage but will not be shared.

Data Fair-Use Statement

The SPRUCE data provided on the public archive are freely available and were furnished by the SPRUCE Research Group at ORNL, U.S. Forest Service, and cooperating independent researchers who encourage their use.

- Please inform (e-mail is appropriate) the SPRUCE scientist(s) of your use the archived data and of any publications that result from your use of the data. Contact information is provided on the Project web site.
- We advise users to check the SPRUCE Data Archive frequently to ensure that you are using the latest version of the data.
- Please acknowledge (1) data products as a citation as provided in the data archive documentation, (2) web site information downloads as a bibliographic web citation, or (3) general SPRUCE information as an acknowledgment or personal communication if no other citation form is applicable.
- When publishing original analyses and results using these data, please acknowledge the agency or organization that supported the collection of the original data.
- Please include these terms as publication keywords as applicable: SPRUCE Experiment, ORNL, U.S. DOE Office of Science, Marcell Experimental Forest, Northern Research Station, U.S. Forest Service.
- Please provide an electronic reprint of your independent work to the SPRUCE Project so that all publications resulting from these data may be tracked, recorded, and referenced by the Carbon Dioxide Information Analysis Center (CDIAC).

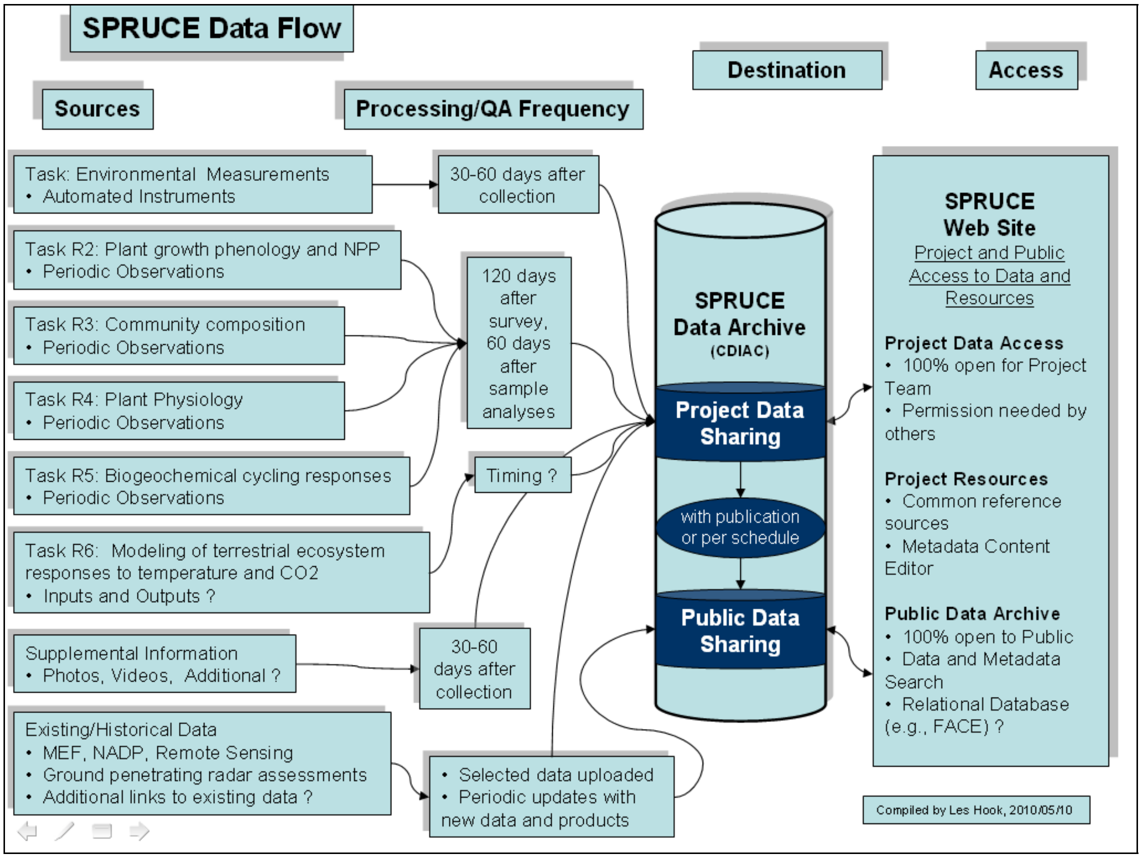
Disclaimer of Liability

Data and documents available from the SPRUCE web site (http://mnspruce.ornl.gov/) were prepared as an account of work sponsored by an agency of the U.S. Government. Neither the U.S. Government nor any agency thereof, or any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Further, Oak Ridge National Laboratory is not responsible for the contents of any off-site pages referenced.

The complete ORNL disclaimer can be viewed at http://www.ornl.gov/ornlhome/disclaimers.shtml.

Data Flow

The data flow diagram is a representation of the policy to inform investigators and potential data users of the general flow of data and information before, during, and after the field experiment. Certain data and metadata reporting standards are necessary to facilitate efficient data reporting, processing and analysis. Data will be archived at the CDIAC Data Archive (http://cdiac.ornl.gov/).



Description of Data and Information Management Features

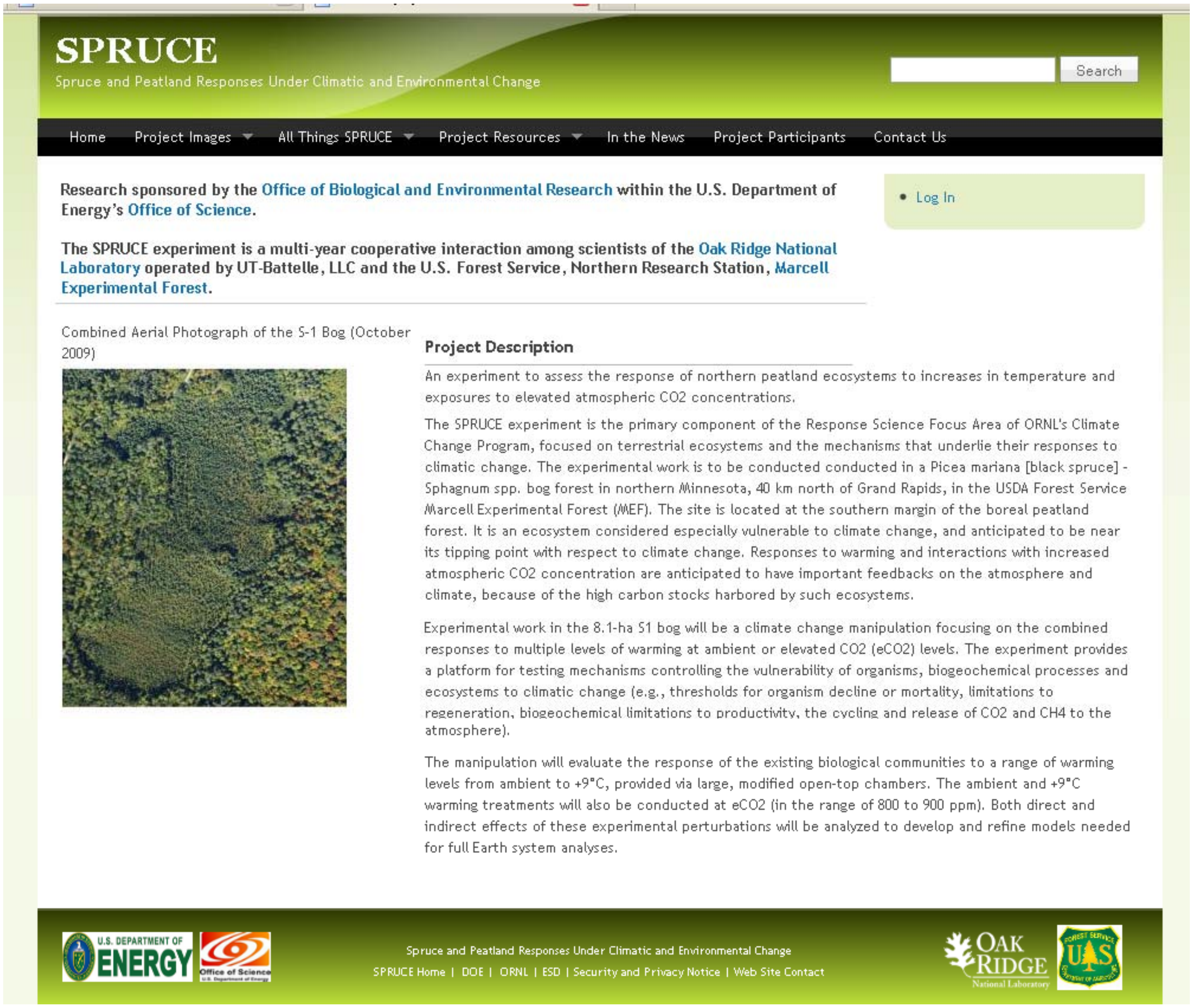
There are three data and information management features identified in the flow diagram; the Project Data Sharing archive, the Public Data Sharing archive, and the SPRUCE Web Site.

The data archives and data access capabilities have not been implemented at the time this Plan was drafted. The following descriptions of the data archives are based on what other similar projects have used and the typical approach of the CDIAC data archive for public access to data.

The **Project Data Sharing archive** will be a straightforward catalog of data files likely organized by Task. The catalog will be accessible through the SPRUCE Web Site to project participants with the appropriate access permission. Data collected in FY2009 (e.g., vegetation survey) and existing MEF and NADP data will be used to initiate development of the catalog.

The **Public Data Sharing archive** will likely follow the CDIAC data set model with compiling of data set metadata and documentation. The exact content and formats to be determined.

The **SPRUCE Web Site** has been implemented (http://mnspruce.ornl.gov/), first and currently as an HTML site with project documents, images, and presentations available to participants and the public. Another set of resources including survey data, meeting materials, and preliminary model outputs are only available only to project participants with permission. The site has recently been implemented using the Drupal web site application and incorporating the same content and access controls. The Drupal site will become the active site in the near future at the same URL.



Project Name Information

It is important to the long-term success of a project that investigators consistently identify the project, the participants, the places, and the sponsors. This section provides the standard names to be used during the project and the guidelines for constructing names for sampling sites and experimental chambers.

A proposed table of site information needed for compiling an all-inclusive list of project locations is included.

SPRUCE Names

***STUDY ACRONYM**
SPRUCE

***STUDY NAME**
Spruce and Peatland Responses Under Climatic and Environmental Change

Funding Organization Acronym	***ORGANIZATION NAME:**
U.S. DOE	U.S. Department of Energy, Office of Science, Program for Terrestrial Ecosystem Science
U.S.D.A. Forest Service	U.S. Department of Agriculture, Forest Service, Northern Research Station, Marcell Experimental Forest
Others?	

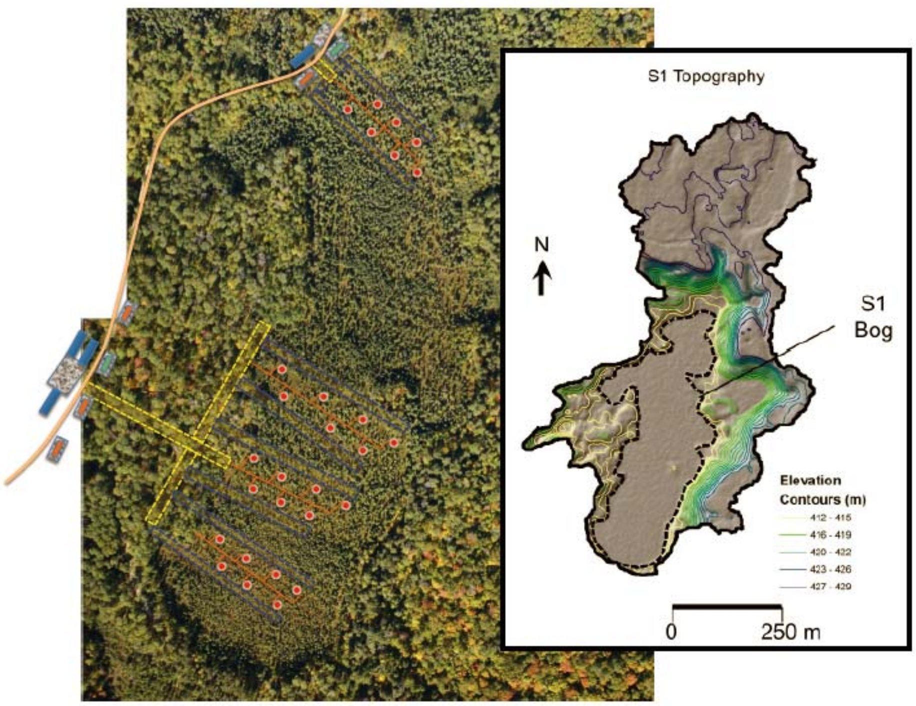
Cooperating Organization Acronym	***ORGANIZATION NAME:**
MEF	U.S. Forest Service, Northern Research Station, Marcell Experimental Forest
Others?	

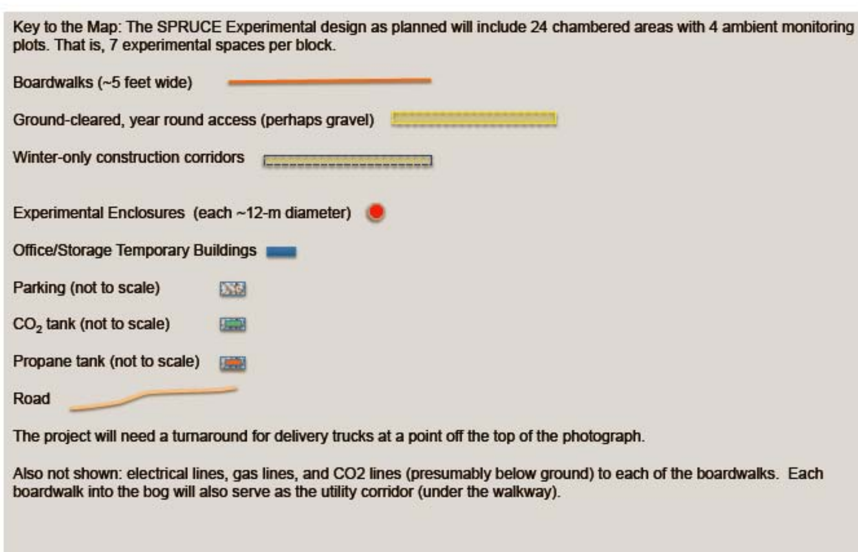
Identifying Measurement and Sampling Sites

A standard for identifying and characterizing fixed measurement locations used by the project as measurement and sampling sites. Specifications, resources, examples are shown in the following template.

TBD

Identifying Measurement and Sampling Sites and Naming experimental blocks and chambers within the S1 Bog. (images from 2010/02/10)





Project Master List of Site Information

The project will maintain a master list of site identifiers, characteristics, and other available information. For the SPRUCE experiment and the S1 Bog, several of the fields would have the same values. This list should include existing historical MEF sites with data relevant to SPRUCE research and modeling, and any new S1 Bog sites established by the SPRUCE project. Picklists of standard values exist for some fields and could be compiled for others as needed.

SPRUCE Site Information Template, Version 1 (2010/04/23)

Data dictionary of fields in the Site information Template for capturing SPRUCE measurement site and treatment plot characteristics.

SHORT NAME	DESCRIPTION	Required/ optional field	LONG NAME	UNITS / FORMAT	FORMAT TYPE	FORMAT FOR DISPLAY (max)	MISSING CODE
Site_ID_level_1	Level 1 identifier	This is a required field.	Site_ID_level_1	None	Char	12	None
Site_ID_level_2	Level 2 identifier	This is a required field.	Site_ID_level_2	None	Char	12	None
Description	This usually would be used to provide the full name of the site and other identifying information for the location.	This is an optional field.	Description	None	Char	50	None
Lat_dd	South latitude is negative.	This is a required field.	Latitude_decimal_degrees	decimal degree	Decimal	10.5	-999.99999
Lon_dd	West longitude is negative.	This is a required field.	Longitude_decimal_degrees	decimal degree	Decimal	10.5	-999.99999
Lat_lon_ref_datum	The reference datum is needed for accurate use of the coordinates. More information is available at: http://www.ngs.noaa.gov/faq.shtml Use the drop-down list to enter the code.	This is a required field.	Lat_lon_ref_datum	None	Char	120	None
Lat_lon_accuracy	Please provide an indication of how accurate the spatial coordinates are. Try to specify a value that would give a circle, centered on the coordinates you are providing, that you believe has a 95% probability of including the site's true coordinate.	This is an optional field.	Lat_lon_accuracy	m (meter)	Decimal	7.1	-999.9

SHORT NAME	DESCRIPTION	Required/ optional field	LONG NAME	UNITS / FORMAT	FORMAT TYPE	FORMAT FOR DISPLAY (max)	MISSING CODE
Lat_lon_method	Please select from the drop-down list the method used to determine the latitude and longitude values. More information is available at: http://www.ngs.noaa.gov/faq.shtml	This is an optional field.	Lat_lon_method	None	Char	50	None
Ground_elev_amsl	Altitude of the ground at the sampling site, in meters above mean sea level.	This is a required field.	Ground_elevation_above_mean_sea_level	m (meter)	Decimal	6.1	-99.9
Meas_start_date	The date when the types of measurements started to be collected at this site.	This is a required field.	Measurement_start_date_at_site	yyyy/mm/dd	Date	10	9999/12/31
Meas_end_date	For ongoing measurements use 9999/12/31.	This is a required field.	Measurement_end_date_at_site	yyyy/mm/dd	Date	10	9999/12/31
Coincident_meas_site	A narrative list, using generic terms separated by commas, of co-incident measurements made at the site for the program. e.g. Temperature, pressure, windspeed, wind direction, etc.	This is an optional field.	Co-incident_measurements_at_site	None	Char	300	None
Comment	If you have no comments, you may leave it blank.	This is an optional field.	Comment	None	Char	120	None
Site_type	Optional. TBD.	This is an optional field.	Site_location_type	None	Char	20	None
Site_start_date	The earliest date this site was used for sampling, regardless of network or study. If missing (future site or unknown), use the missing value code shown.	This is an optional field.	Site_start_date	yyyy/mm/dd	Date	10	9999/12/31

SHORT NAME	DESCRIPTION	Required/ optional field	LONG NAME	UNITS / FORMAT	FORMAT TYPE	FORMAT FOR DISPLAY (max)	MISSING CODE
Site_end_date	The date sampling permanently ceased at this site. If missing (ongoing or unknown), use the missing value code shown (9999/12/31).	This is an optional field.	Site_end_date	yyyy/mm/dd	Date	10	9999/12/31
Site_study_start_date	The earliest sampling at the site that is considered to be in connection with the study.	This is an optional field.	Site_study_start_date	yyyy/mm/dd	Date	10	9999/12/31
Site_study_end_date	The latest sampling at the site in connection with the study. If missing (ongoing or unknown), use the missing value code shown (9999/12/31).	This is an optional field.	Site_study_end_date	yyyy/mm/dd	Date	10	9999/12/31
Site_monitoring_duration	A picklist is provided when you click in the data cell. Characterization of the actual or planned use of the site.	This is an optional field.	**Site_monitoring_duration**	None	Char	50	None
Site_info_source	Site_info_source	This is a required field.	Site_info_source	None	Char	50	None
Date_added_to_template	Date_added_to_template	This is a required field.	Date_added_to_template	yyyy/mm/dd	Date	10	9999/12/31
Date_last_updated_record	Date_last_updated_record	This is a required field.	Date_last_updated_record	yyyy/mm/dd	Date	10	9999/12/31
Added_by	Added_by	This is a required field.	Added_by	None	Char	50	None

<u>Example data records in the Site information Template</u> for capturing SPRUCE measurement site and treatment plot characteristics.

SHORT NAME	DESCRIPTION	DATA BEGINS	Example 1	Example 2	Example 3	Example 4
Site_ID_level_1	Level 1 identifier		**S1**	**NTN**	**1**	**1**
Site_ID_level_2	Level 2 identifier		**Bogwell S1**	**MN16**	**1**	**2**
Description	This usually would be used to provide the full name of the site and other identifying information for the location.		**Bogwell S1**	**NADP/NTN Monitoring Location MN16**	**S1 Survey Transect Plot**	**S1 Survey Transect Plot**
Lat_dd	South latitude is negative.		47.50785	47.5311	47.50454	47.50448
Lon_dd	West longitude is negative.		-93.45242	-93.4686	-93.45419	-93.45394
Lat_lon_ref_datum	The reference datum is needed for accurate use of the coordinates. More information is available at: http://www.ngs.noaa.gov/faq.shtml Use the drop-down list to enter the code.		NAD83 (North American Datum 1983)			
Lat_lon_accuracy	Please provide an indication of how accurate the spatial coordinates are. Try to specify a value that would give a circle, centered on the coordinates you are providing, that you believe has a 95% probability of including the site's true coordinate.					
Lat_lon_method	Please select from the drop-down list the method used to determine the latitude and longitude values. More information is available at: http://www.ngs.noaa.gov/faq.shtml		GPS-Unspecified			
Ground_elev_amsl	Altitude of the ground at the sampling site, in meters above mean sea level.			431		

SHORT NAME	DESCRIPTION	DATA BEGINS	Example 1	Example 2	Example 3	Example 4
Meas_start_date	The date when the types of measurements started to be collected at this site.			1978/07/06	2009/09/21	2009/09/21
Meas_end_date	For ongoing measurements use 9999/12/31.			9999/12/31	9999/12/31	9999/12/31
Coincident_meas_site	A narrative list, using generic terms separated by commas, of co-incident measurements made at the site for the program. e.g. Temperature, pressure, windspeed, wind direction, etc.		ground water level	http://nadp.sws.uiuc.edu/ads/2008/MN16.pdf	vegetation survey, species composition and % cover, DBH, height, peat depth	vegetation survey, species composition and % cover, DBH, height, peat depth
Comment	If you have no comments, you may leave it blank.		Watershed: S-1 Cover type: Lowland black spruce Stand origin: 1968 Soil name: Greenwood peat Slope: 0%	Marcell Experimental Forest (MN16), USGS 1:24000 Map Name Balsam Lake	Site_ID_level_1 is Transect. Site_ID_level_2 is Plot	Site_ID_level_1 is Transect. Site_ID_level_2 is Plot
Site_type	Optional. TBD.		MEF S1 permanent monitoring site	NADP/NTN Monitoring Location		
Site_start_date	The earliest date this site was used for sampling, regardless of network or study. If missing (future site or unknown), use the missing value code shown.				2009/09/21	2009/09/21
Site_end_date	The date sampling permanently ceased at this site. If missing (ongoing or unknown), use the missing value code shown (9999/12/31).				9999/12/31	9999/12/31
Site_study_start_date	The earliest sampling at the site that is considered to be in connection with the study.				2009/09/21	2009/09/21

SHORT NAME	DESCRIPTION	DATA BEGINS	Example 1	Example 2	Example 3	Example 4
Site_study_end_date	The latest sampling at the site in connection with the study. If missing (ongoing or unknown), use the missing value code shown (9999/12/31).				9999/12/31	9999/12/31
Site_monitoring_duration	A picklist is provided when you click in the data cell. Characterization of the actual or planned use of the site.		Long term (more than 5 years)	Long term (more than 5 years)	Short term (from 1 to 5 years)	Short term (from 1 to 5 years)
Site_info_source	Site_info_source		http://www.nrs.fs.fed.us/EF/Marcell/sites/S1/	http://nadp.sws.uiuc.edu/sites/siteinfo.asp?id=MN16&net=NTN	**Copy of BogSurveySep09.xls**	**Copy of BogSurveySep09.xls**
Date_added_to_template	Date_added_to_template		2010/04/28	2010/04/28	2010/04/28	2010/04/28
Date_last_updated_record	Date_last_updated_record		2010/04/28	2010/04/28	2010/04/28	2010/04/28
Added_by	Added_by		hookla@ornl.gov	hookla@ornl.gov	hookla@ornl.gov	hookla@ornl.gov

Data and Metadata Reporting

- Reporting Sampling and Measurement Dates and Times
- Identifying Descriptive Field Variables, Biological Measurements, Chemical and Physical Variables
- Reporting Units for Chemical, Physical, and Descriptive Variables
- Reporting Values below Detection Limits
- Reporting Missing Data
- Reporting Uncertainty Estimates
- Reporting Conventions for Meteorological Data, and Temperature and Pressure Conditions
- Assigning Project-Specific Data Quality Flags

Data Collection Guides

This section is supported by a series of Data Collection Guides that were created to be useful step-by-step guidance for implementing these data collection and reporting activities. These guides were developed based on a review of typical SPRUCE data collection and processing steps and can be updated and expanded as appropriate to meet the needs of the project. The Data Collection Guides are available on the SPRUCE web site (http://mnspruce.ornl.gov).

Reporting Sampling and Measurement Dates and Times

This section provides a standard for reporting sampling and measurement dates and times for SPRUCE investigations and the warming experiment.

Because reporting dates and time are so important to the success of a project, we have tried to be as explicit as possible in this specification and have designed some redundancy into the reporting fields for date and time to prevent many of the reporting problems encountered by similar intensive monitoring projects.

Date and Time Characteristics

Time basis	**Investigators will report data on a Central Standard Time (CST) basis. (Equivalent to Coordinated Universal Time (UTC) - 6 hours)**
Reporting start date and time	**Start date and time must be reported as time at the beginning of the sampling/measurement/averaging period.**

Reporting interval	**When SPRUCE protocols call for a continuous time series of data. The reporting interval will be 30 minutes for parameters with sampling/measurement/recording intervals that can be reliably averaged or summed to 30 minutes.**
Reporting end date and time (optional)	For continuously monitored processes, the end date and time of the preceding period may be the start date and time of the next period
Midnight convention	**There is no 24:00 time. 23:59, then 00:00 the next day.**
Valid values	**The number of valid observations within a sampling/measurement/recording interval that are required to yield a reliable 30 minute average value will be defined by the Task Leader.**
Missing values	**Reporting intervals with unreliable, invalid, or missing data will be set to a missing value code. Evaluation criteria for missing values will be defined by each Task Leader for each variable.**

Reporting Formats for Dates and Times:

Local Time Zone. Specify CST.

Date Formats: 2003-02-28 or 2003/02/28

Time Formats: 07:00:30. (Note leading zero.)

Date Time Stamp: 2010-03-17 05:30:00

CST lags UTC time by 6 hours. If the Universal Time is 14:30 UTC, Central Standard Time would be 08:30 CST.

Dates and Times to Report for Project Data Sharing and Archiving

Date and time parameters to be reported with sampling and measurement parameters.

Columns	Units / formats	Notes
Year	yyyy (year)	Values can be derived from Date and Time CST
Month	mm (month)	
Day	dd (day)	
DOY	DOY (day of year)	
Hour	hh (hour)	
Minute	min (minute)	
Second (optional)	s (second)	
Date_start	yyyy/mm/dd	Date and Time CST can be derived from component values.
Time_start	hh:mm (:ss)	
Date_start_UTC	yyyy/mm/dd	Date and Time UTC can be derived from Date and Time CST. Optional.
Time_start_UTC	hh:mm (:ss)	

Justification for the explicit and redundant data and time format specifications:

- Ease of analysis.
- Minimize issues with varying time basis on computing platforms.
- Values formatted as date and time are needed for temporal searching applications in the data archive.
- Consider including UTC dates and times to support integration with data from external monitoring networks and remote sensing sources. These fields can be added easily in a secondary data product.

Time Resources:

Discussion of Coordinated Universal Time (UTC) [http://www.usno.navy.mil/USNO/time/master-clock/systems-of-time].

http://en.wikipedia.org/wiki/Coordinated_Universal_Time

U.S. Naval Observatory [http://www.usno.navy.mil/USNO]

To set your PC to the correct U.S. time [http://nist.time.gov/]

Identifying Descriptive Field Variables, Biological Measurements, Chemical and Physical Variables

This section provides the approach for identifying descriptive field variables, biological and ecological measurements, chemical and physical variables, and various descriptive metadata elements for sampling, measurement, and modeling tasks.

Being as consistent as reasonably possible with the preferred naming standards of the CDIAC Archive has advantages for future data integration and search applications. Names should be constructed for easy importing by various data systems and should contain only numbers, letters, and underscores -- no spaces or special characters.

For example, CDIAC archives data from the AmeriFlux Program and there is a well established parameter list for meteorological, carbon cycle, and energy flux measurements at flux tower sites. The Ameriflux investigators also collect ecological measurements (e.g., biomass) to characterize their sites and report those per an established set of parameter names. CDIAC is also the host for the ORNL FACE data archive (other FACE sites also) and is currently developing a relational data base for searching and accessing FACE data.

Task Leader will implement a standard for SPRUCE Variables for Modeling and Measurements.

Existing Naming Standards for Reference:

AmeriFlux meteorological and micrometeorological data:
http://public.ornl.gov/ameriflux/AmeriFlux_Data_Submission_Guidelines_WithGaps.pdf

AmeriFlux biological data:
http://public.ornl.gov/ameriflux/AmeriFlux_Biological_Data_Submission_Guidelines.doc or
http://public.ornl.gov/ameriflux/AmeriFlux_Biological_Data_Templates_2009.xls

Reporting Units for Chemical, Physical, and Descriptive Variables

This section provides the approach for identifying the **units to be used to report** the descriptive field variables, biological and ecological measurements, chemical and physical variables, and various descriptive metadata elements for sampling, measurement, and modeling tasks. We will use SI units for the most part

As a reference for units commonly used by the carbon cycle and climate change community, as for names, CDIAC archives data from the AmeriFlux Program and there is a well established parameter list for meteorological, carbon cycle, and energy flux measurements at flux tower sites. The Ameriflux investigators also collect ecological measurements (e.g., biomass) to characterize their sites and report those per an established set of parameter names and units. CDIAC is also the host for the ORNL FACE data archive (other FACE sites also) and is currently developing a relational data base for searching and accessing FACE data. Being as consistent as reasonably possible with the preferred naming and units standards of the CDIAC Archive has advantages for future data integration and search applications.

Task Leaders will implement a standard for SPRUCE Variables for Modeling and Measurements that includes the preferred reporting units.

Reporting Values below Detection Limits

Provides guidance for reporting detection limits and data values below the limit of detection. A similar approach can be taken for values above an upper limit of detection.

> Task Leaders should give this a lot of thought. Would we prefer to assume 0, insert a missing value, or report the **actual measured value** even if the value is below the detection limit (including zero and negative values).

As general guidance:

Report detection limits in the data file or in the data documentation.

- If the detection limit may change with each value or across samples reported in the same file then the detection limit should be reported with each value.
- If the detection limit applies to method, the values may be provided in data archive documentation.

Report values as the actual measured value even if the value is below the detection limit (including zero and negative values). It is not recommended to substitute zero or the detection limit for below detection limit data.

If SPRUCE implements a quality flagging system then data values can be flagged appropriately.

Value **flagged "V1"** (Valid value but comprised wholly or partially of below detection limit data), or

Value **flagged "V7"** (Valid value but set equal to the detection limit (DL) because the measured value was below the DL, or

Value **flagged "V2"** ("Valid estimated value"), if a measured value is below what is considered to be the normal "Detection Limit" but is nonetheless considered meaningful, suggesting where between zero and the DL the value lies.

Reporting Missing Data

This section provides guidance for reporting missing data. All data fields must have a value present, either the measured value, substituted detection limit, or a missing value representation.

There may not be blank data fields. Dates and time should not be missing.

Character fields: Use 'None' as the missing code for character values.

Integer and Decimal fields: For integer and decimal format variables, in general, be negative and large enough to be impossible as actual data value.

1) Use a consistent missing value code (i.e., -9999) for all integer and decimal variables.

2) Alternatively, match the format of the column and use repeated 9's (e.g., Decimal: -999.99; Int: -999).

For Scientific format, match the format of the column and use repeated 9's. Use +02 or a similar appropriate value as the exponent in Scientific notation (e.g., -9.99E+02, NOT -9.99E+99).

Reporting Uncertainty Estimates

This section provides guidance for reporting uncertainty estimates with measured values and providing a brief explanation of the meaning/interpretation of the uncertainty values, or for characterizing the uncertainty, associated with the measurements.

In addition to providing the measured value, most sampling and measurement strategies can be designed to provide one or more estimates of uncertainty associated with the measured value.

Task Leaders should give this a lot of thought. Are your sampling and measurement strategies designed to provide measures of uncertainty? What measure of uncertainty does modeling need?

Reporting Uncertainty

Uncertainty values that can be expressed as a constant for all of the data in the data file can be reported in the data documentation. Similarly, Sample-level uncertainty can be reported in the documentation.

Uncertainty values can be provided for each measurement value by adding one or more associated columns and named using the same variable name but appending an uncertainty code (e.g., _unc).

The reported uncertainty estimate, whether in the documentation or included in the data file must be clearly defined, for example, accuracy, precision, confidence interval (+/-), % lower confidence bound, % upper confidence bound, combined standard, or expanded. A method description could include more specifics.

Reporting Conventions for Meteorological Data, Temperature and Pressure Conditions, and Blank Corrections

This section provides guidance for reporting data for several sample media and data types.

Volumetric and Mass/volume Measurements

- Gas flow, volumetric measurements, and mass/volume conversions, must be reported at known (ambient or standardized) temperature and pressure conditions.
- If measurements are standardized, use standard temperature of 0 °C (273.15 K, 32 °F) and an absolute pressure 101.325 kPa (14.696 psi, 1 atm).

Documenting Temperature and Pressure Conditions

Whatever temperature and pressure conditions are used must be documented, for example, 0 deg. C; 1 atmosphere.

Blank corrections? Is the field or laboratory blank statistics subtracted from the analytical result?

Assigning Project-Specific Data Quality Flags

This section provides information for Task Leaders to consider as they decide about assigning quality flags to data values and what the best approach might be.

Consider that by asking the data originator to assign flags to reported data values, any subsequent data user will be able to easily know whether the data are valid without qualification, valid but qualified/suspect, or were invalidated due to serious sampling or analysis problems.

> Task Leaders should give this a lot of thought. Is it important to document the quality of each measured value? Consider the effort involved vs. the benefit to users.

Examples of quality flags:

V0 Valid value
V1 Valid value but comprised wholly or partially of below detection limit data
V2 Valid estimated value
V3 Valid interpolated value
V4 Valid value despite failing to meet some QC or statistical criteria
V5 Valid value but qualified because of possible contamination (e.g., pollution source, laboratory contamination source)
V6 Valid value but qualified due to non-standard sampling conditions (e.g., instrument malfunction, sample handling)
V7 Valid value but set equal to the detection limit (DL) because the measured value was below the DL
M1 Missing value because no value is available
M2 Missing value because invalidated by data originator
H1 Historical data that have not been assessed or validated

This set of flags for below detection values was suggested above:

Value **flagged "V1"** (Valid value but comprised wholly or partially of below detection limit data), or

Value **flagged "V7"** (Valid value but set equal to the detection limit (DL) because the measured value was below the DL, or

Value **flagged "V2”** ("Valid estimated value”), if a measured value is below what is considered to be the normal “Detection Limit” but is nonetheless considered meaningful, suggesting where between zero and the DL the value lies.

Flagging Missing Values

Missing values may be flagged with either the “M1” or “M2” flag as appropriate. The “M1” flag indicates that no data were collected. The “M2” flag indicates that the value was set to missing by the investigator due to a quality problem.

Data Process Planning

- Data Entry, Transfer, and Transformation
- Managing Hardcopy Format Project Records
- Managing Electronic Format Project Records
- Names and Reporting Formats for Data Files
- Scripted Programs for Processing and Analysis
- Quality Level of Data

Task Leaders have significant responsibilities for defining the collection of data and for moving the data to the Project and finally to the Public Data Archive. This section outlines the steps that will make this happen. Our goal is to establish a series of steps that will define a controlled process leading to data products that are of good quality and valuable to the Project Team for answering the science questions.

This will lead to a more detailed SPRUCE Task-level data flow chart (below) and perhaps to Task-specific data flow charts that identify data collection, processing, and archiving steps.

This section and the data collection guides will provide data processing guidance and best practices, but it's up to the Task Leaders and investigators who understand the specific data collection, quality assurance, and analysis steps that are needed, to indentify the processing details, and to ask for additional assistance from the data management team if needed.

Data Collection Guides

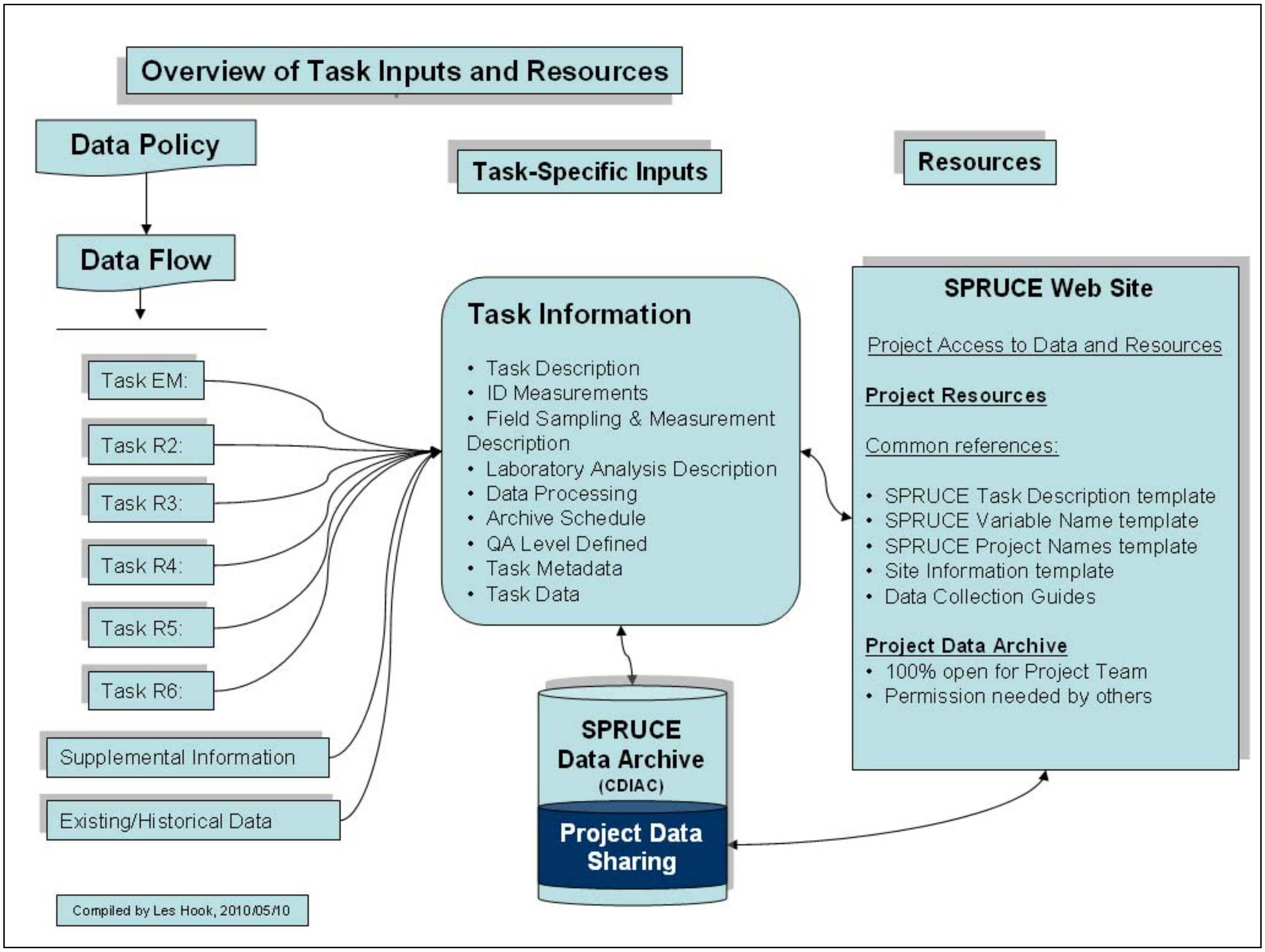
This section is supported by a series of Data Collection Guides that were created to be useful step-by-step guidance for implementing these data processing activities. These guides were developed based on a review of typical SPRUCE data collection and processing steps and can be updated and expanded as appropriate to meet the needs of the project. The Data Collection Guides are available on the SPRUCE web site (http://mnspruce.ornl.gov).

Tabular and Image Data Note

Note that these discussions are focused primarily on measurement and observation results that are stored in tabular ASCII files. Similar guidance is available for processing GIS and image data products and can be added as needed.

Environmental Monitoring -- Sensors

A brief description of the planning for the installation of sensors is included at the end of this section.



SPRUCE Task-level data flow diagram

Task Leaders should keep in mind these data processing issues as they define data flow.

Data provenance? Is your data traceable from field measurements and samples, through instrument output files, data processing, deriving additional parameters, to the final data product file?

Version control? How do you handle changing values and version control of final data products?

Data Entry, Transfer, and Transformation

Data entry, transfer, and transformation activities should be verified to ensure that data integrity is maintained. This includes movement/copying of data from one storage medium to another and transformation from one format to another. All data, including analytical data produced and reported by a laboratory should be verified.

This verification encompasses all data recording media, handwritten or hard copy produced via electronic means, as well as electronically stored, such as in a database. It also includes all data collection methods (e.g., electronic collection through real-time monitoring instrumentation, bar coding equipment, and handwritten log entries).

If a data transformation or transfer activity has occurred before receipt of the data by project personnel (i.e., between creation and final reporting), the verification may be performed by the reporting party but only if sufficient evidence to support the validity of the process can be provided by the reporting party. For example, if a laboratory technician captures data from a laboratory instrument and records it in a logbook, enters the data from the logbook into an electronic data deliverable format, and then transfers the data to the project, the verification process may be performed by the laboratory. The mechanism for a project's data entry, transfer, and transformation verification processes should be documented.

Managing Hardcopy Format Project Records

To ensure that hardcopy records of data collection and generation activities are protected, hardcopy records should be specified, prepared, reviewed, and maintained to document the quality for the work completed. Records are completed documents that provide objective evidence of the quality of an item or process. Project task plans should identify the responsibilities for record retention, protection, preservation, traceability, disposition, and retrievability.

Examples of hardcopy records include field data collection forms, forms for capturing details of field photographs, forms to record sample collection information, and forms for recording laboratory measurements

Managing Electronic Format Project Records

Processed Measurement and Observation Data Products

The data flow diagram (see Organization section) is a representation of the general flow of processed electronic format data and information during the field experiment (Quality Levels 1 and 2) as they move from investigators to the CDIAC Data Archive. And most of this Data Process Planning Section discusses best practices for the creation and quality of processed data to be defined in Task Plans. In this short section, the management (security) of processed data as related to the data system is identified and should be considered in the Task Plan. The following **Data Systems Management** Section includes guidance on Data System Backup, Access Control, and Configuration Control.

Unprocessed/Raw Measurement

Tasks should identify the raw or minimally processed (Quality Level 0) measurement and observation data that are recorded by their instruments or entered from field forms and plan to store this data for a period of time defined by the project – usually a minimum of 5 years after completion of the project. A raw data/Level 0 data archive capability will be provided as needed.

Digital Media

Tasks should identify the digital media they will collect. Digital media is used here to describe digital photographs, video, audio, movies of model outputs, etc. These digital products may be collected in the field, laboratory, or generated by a computer model. As such they may be considered, on the one hand, unprocessed raw data from which measurements will be derived, or on the other, a final product of a modeling task. Plans should reflect the appropriate level of data sharing and archiving. The native format should be defined and any planned transformation identified. The preferred format for long-term archiving should be noted and any special media curation needs identified.

Computer codes, models, input and output data sets

Tasks should identify the specific software codes used to process measurement data and specific versions of model codes plus their input and output data sets that were used to generate a specific product or publication and plan to store this information for a period of time defined by the project – usually a minimum of 5 years after completion of the project. Consider archiving model codes along with input and output data files.

Names and Reporting Formats for Data Files

Data File Names

A consistent syntax for data file names first, simplifies data collection, processing, and version control for investigators and second, simplifies data submission to the Data Archive, archive organization, and retrieval by users. The preferred syntax will be decided by the SPRUCE Team and added as a Data collection Guide. Below are examples.

Data File Formats

Until more measurement details and data and metadata reporting standards are defined, the best guidance is from ORNL DAAC Best Practices. This information is included as Appendix XW.

Data File Naming	**Limits**
Data File: [STUDY ACRONYM]_[unique data file descriptors]_S1.csv [STUDY ACRONYM]_[unique data file descriptors]_V1.csv	64 chars max, upper or lower case (except .xls, .csv, and other extents which should be lower case)

Task Leaders should define a standard syntax for the [unique data file descriptors] portion of the data file name.

The unique data file descriptor may be Task_event_content.

For example, R2_S1_bog_leaf_CNP_S1.csv

Example: spruce_what_where_when_S1.csv

spruce_taskcode_what_where_when_S1.csv

spruce_DataType_

spruce_phenology_

Version:

Where S1<n> indicates version of data file that may be shared among SPRUCE participants.
Where V1<n> indicates version of data file that is publicly available.

Raw or unprocessed file name syntax:

“Task_event_content_date.csv” or “Task_event_content_date.xls”

For example, SPRUCE_R2_S1_leaf_C_N_P_20100413.csv or R2_S1_plot_leaf_CNP_20100413.txt

Resource: http://daac.ornl.gov/PI/bestprac.html

File names should be constructed for easy management by various data systems. **Names should contain only numbers, letters, dashes, and underscores -- no spaces or special characters.** Also, in general, lower-case names are less software and platform dependent and are preferred. When choosing a file name, check for any database management limitations on the use of special characters and file name length. For practical reasons of legibility and usability, file names should not be more than 64 characters in length and if well constructed could be considerably less.

Scripted Programs for Processing and Analysis

Using a scripted program for processing and analysis will save you countless headaches in the future. Analysis scripts are written records of the various steps involved in processing and analyzing data, and provide a form of analytical “metadata.” Such scripts can be easily revised and re-executed any time you need to modify an analysis. Examples of applications that can be used to create a scripted program are the R statistical package (which is free, flexible, and *very* good), MATLAB, or SAS.

This scripted approach is in contrast to a “GUI-driven” analysis, in which various changes to data are made by selecting and altering values in place, and by choosing various processing steps from drop-down menus, followed by a “run” or “execute” or “ok” button. Such GUI-based analyses often seem convenient when working through your data and analysis, but rarely leave a clear accounting of exactly what you have done. In contrast, when you use a scripted program for your analysis, you will always have a record of what you did with your data from the time you collected it to the time you publish it, so it's easy to recollect your decisions, even after a few years have passed.

Quality Level of Data

As described in the Data Policy, data should be quality assured before archiving for sharing and those checks identified. Each SPRUCE Task Leader will define the quality assurance checks to be performed prior to data sharing among SPRUCE participants (Quality Level 1) and then prior to public access (Quality Level 2).

- **Level 1:** Indicates an internally consistent data product that has been subjected to quality checks and data management procedures.
- **Level 2:** Indicates a complete, externally consistent data product that has undergone interpretative and diagnostic analysis by the SPRUCE participants.
- For completeness, **Quality Level 0** data are products of unspecified quality that have been subjected to minimal processing in the field and/or in the laboratory (e.g., raw data, photos, hardcopy data sheets, scanned data sheets, notebooks, etc.). This may, for example, be data from an instrument logger expressed in engineering units or using nominal calibrations, or high resolution data before aggregating to a selected interval. These products should be submitted to the Data Archive for long-term storage but will not be shared.

Environmental Monitoring -- Sensors

Measurement Planning

The design of the sensor array for the experimental chambers is in the preliminary planning stages.

The final implementation will be developed based on the results of preliminary measurements made in and around the full-size experimental prototype being constructed at ORNL (Spring - Summer 2010) and from the deployment of meteorological and other sensors at the MEF S1 Bog (Summer 2010 and continuing).

Data files from these measurement activities will also serve as prototypes for defining the data processing steps needed to move the data from collection, to analysis, and to archive. Particular attention will be given defining a clear automated linkage between measurements, calibration history, and the instruments environment. The sensor and measurement metadata needed for archive documentation will be defined.

Operations and Maintenance

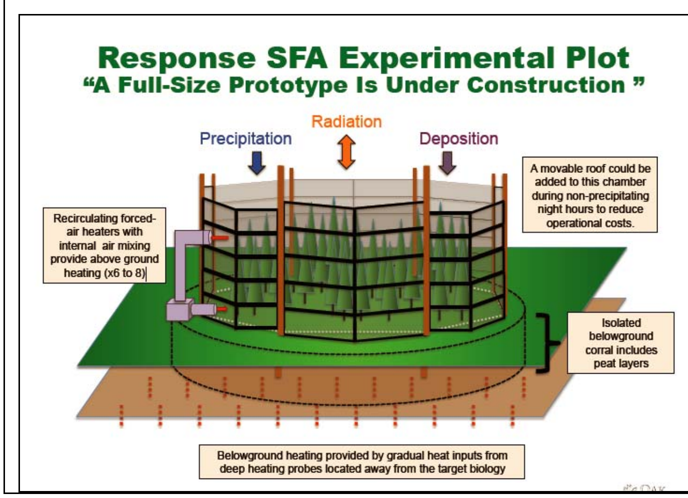
The ORNL prototype monitoring and the MEF S1 Bog monitoring will also serve as test platforms for developing the operations and maintenance procedures needed to support the sizeable installation of sensors in the experimental enclosures. This includes sensors for environmental measurements the sensors that monitor the heating and air handling infrastructure systems.

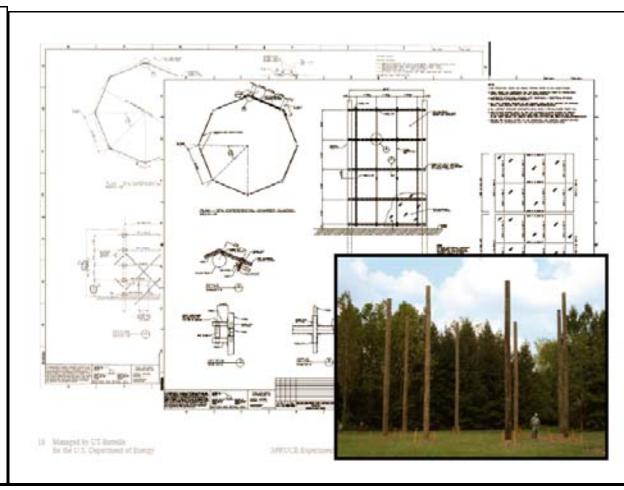
A valuable tool for closely tracking the performance of sensors and the conditions in the enclosures will be a real-time visualization application for display of data as it is being collected. Campbell Scientific has an application that interfaces with their data loggers and may be applicable for task.

Under consideration is the implementation of a bar-code based tracking system for purposes of inventory control, installation tracking, and tracking of sensor calibration, maintenance, and replacement of sensors. This system would have links to the measurement data. The sensors to be tracked include the experimental environmental sensors and the sensors that monitor the heating and air handling infrastructure systems.

Data Systems

The data systems needed to support the experiment data storage (downloaded from data loggers), real-time visualization of sensors performance, backup of data, and transfers of data will be evaluated during the prototype and pre-experiment monitoring. A major consideration is the possible need for local (S1 Bog) data systems and remote (ORNL) data systems and a reliable link between the two.





Data Documentation and Archiving

- Planning to Archive Data for Public Release
- Creating Archive Documentation
- Providing Metadata to Searchable Indexes and Clearinghouses
- Assigning Descriptive Data Set Titles

The SPRUCE Task Description spreadsheet is referenced in the second subsection and is included as Appendix XY for your information and included as a Data Collection Guide on the SPRUCE web site (https://mnspruce.ornl.gov/content/spruce-data-policies).

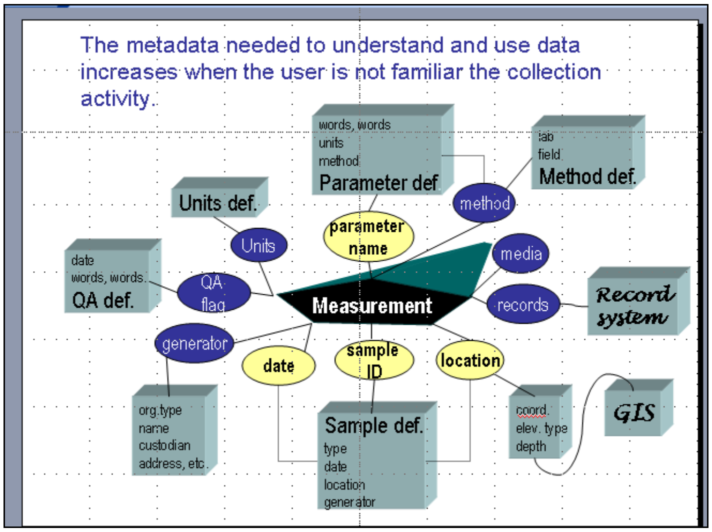
Planning to Archive Data for Public Release

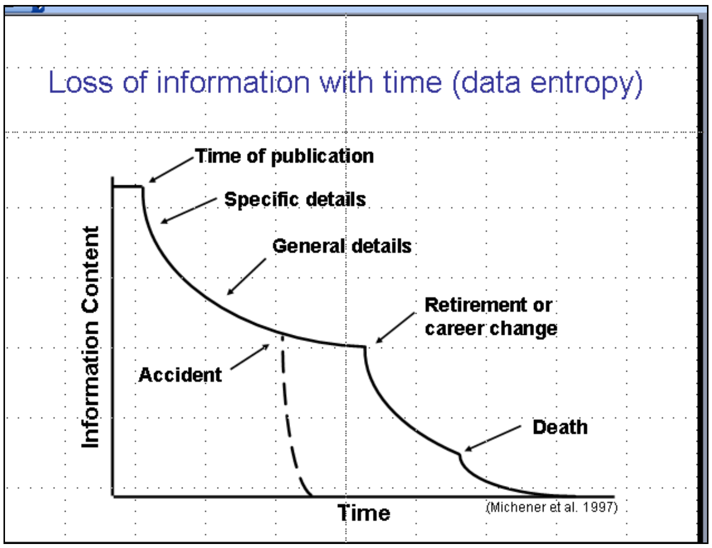
One of the main goals of the data policy is to expeditiously share experimental data among researchers, the broader scientific community, and the public. Advanced planning furthers efforts to identify, collect, and report consistent data and metadata and to facilitate timely data analysis, sharing, integration, and synthesis by all users.

To reiterate, as noted in the Data Policy, the SPRUCE Data Archive will be hosted by the Carbon Dioxide Information Analysis Center (CDIAC) at ORNL and will maintain two levels of data accessibility. The first is for sharing recently collected, derived, and processed data products among SPRUCE participants and the second is for access by the public.

The level of quality assurance needed for sharing newly collected, derived, and processed data among SPRUCE participants for information and confirmation purposes is typically not as great as that for publication and public access.

Similarly, the amount and scope of metadata and data documentation needed to accompany data increases when users are not familiar with the data collection activities, i.e., the public, and perhaps even over the lifetime of this long-term experimental activity. Thus additional companion metadata and documentation are needed and will be added when data products are promoted for public accessibility.





The compilation of the companion metadata records and their inclusion in searchable metadata databases and clearinghouses of larger collections of climate change data facilitate the subsequent discovery of the data sets and products by the scientific community and public for uses beyond those of the original project.

The metadata required (i.e., standard) for the CDIAC searchable metadata database are not extensive (reference below) but must be considered when planning data collection and modeling activities.

Creating Archive Metadata and Documentation

The metadata accompanying a data set should be written for a user 20 years into the future--what does that investigator need to know to use the data? Write the document for a user who is unfamiliar with your project, methods, and observations.

Scientists are encouraged to document their data at a level sufficient to satisfy the well-known “20-year test”. That is, someone 20 years from now, not familiar with the data or how they were obtained, should be able to find data of interest and then fully understand and use the data solely with the aid of the documentation archived with the data. (National

Research Council, Committee on Geophysical Data, Solving the Global Change Puzzle, A U.S. Strategy for Managing Data and Information, National Academy Press, Washington, D.C., 1991.)

The extent of metadata and the content and format of SPRUCE archive documentation has not been finalized. If we use the documentation of similar data collections as a model, the documentation will resemble a typical scientific publication in content. There will be more emphasis on description of the data files. Other large projects have used a web application to enter metadata and documentation and then export the fields in the archive documentation format.

The **SPRUCE Task Description spreadsheet** is an introduction to some of the metadata elements that will be compiled in the CDIAC metadata database to create the searchable index of SPRUCE data set. Other fields are valuable for identifying measured parameters, methods, and units that might be candidates for a picklist of standard values and other aspects of data and metadata reporting. A draft spreadsheet with some draft example values is included in Appendix XY.

Providing Metadata to Searchable Indexes and Clearinghouses

As noted above, SPRUCE data product metadata will be entered into the CDIAC searchable metadata database to permit discovery and use of the data by the by the public. The searchable database uses the “Mercury” search application (http://mercury.ornl.gov/cdiac/). An example search metadata report is included in Appendix XZ.

As noted, the CDIAC metadata and documentation requirements are not extensive, but there are other clearinghouses, web services, and data file formats (e.g., NetCDF) that may be applicable for submitting SPRUCE data product metadata, that have some additional metadata requirements. Several metadata standards including FGDC, NetCDF Attribute Convention for Dataset Discovery, EML, and ISO-19115 need be considered when making decisions about the metadata elements to collect for SPRUCE.

Assigning Descriptive Data Set Titles

Data set titles should be as descriptive as possible. When giving titles to your data sets and associated documentation, please be aware that these data sets may be accessed many years in the future by people who will be unaware of the details of the project.

Data set titles should contain the type of data and other information such as the date range, the location, and the instruments used. These data sets will be part of a larger SPRUCE field project and we will add that name, too (e.g., SPRUCE). In addition, we recommend that the length of the title be restricted to 80 characters (spaces included) to be compatible with other clearinghouses of ecological and global change data collections

A given data set might contain only one data file or many thousands of data files (granules).

Data Set Title	**Notes**
Possible data set title syntax:	80 chars max, title

SPRUCE [Data Description] SPRUCE [taskcode or phase or data type] [Data Description] Names should contain only numbers, letters, dashes, underscores and spaces -- no special characters. The data set title should be similar to the name(s) of the data file(s) in the data set. Some bad titles: • **"The Aerostar 100 Data Set",** • **"Respiration Data"** Some great titles: • **"SAFARI 2000 Upper Air Meteorological Profiles, Skukuza, Dry Seasons 1999-2000"** • **"LBA-ECO CD-07 GOES-8 L3 Gridded Surface Radiation and Rain Rate for Amazonia: 1999"** • **"Global Fire Emissions Database, Version 2 (GFEDv2)"**	case

Data Systems Management

- Day-to-Day Operation of Data Management Systems
- Data Management System and Software Configuration Control Guidelines

This is the "ounce of prevention" section. We all know of data and effort that have been lost due to hardware, software, or operator shortcomings. I would anticipate that we will need to pay special attention to establishing data systems at locations remote to ORNL and to collecting data in an environment that would not be kind to a laptop that slipped off a walkway. More specific guidance can be obtained as needed.

Day-to-Day Operation of Data Management Systems

This section provides guidance to investigators, technicians, instrument operators, and project data managers responsible for the day to day operation of data collection and data management systems including: backups; access and security; data entry, transfer, transformation; and data control.

In the current computing environment, there may not always be a clear distinction between a personal computer and a data collection or data management system computer. Thus, the attention to "data security" needs to be high on both types of systems in use for SPRUCE. We certainly don't want to loose any data or have to recover and to reprocess raw data.

These routine data management protocols can be facilitated through checklists or worksheets (electronic or hardcopy) that aid completing project documentation. An ounce of prevention…

<u>General Computer and Server Security</u>

For ORNL Participants, these precautions should be routinely implemented for computing systems utilizing ORNL networks.

- Operating system and software updates and patches current.
- Antivirus and malware/spyware detection software applications running and up to date.
- Back up services implemented for personal computers and work stations as appropriate (e.g., Connected Backup and Amanda Backup Services).

For computing systems that are not on the ORNL network that may be controlling instruments and data acquisition.

- Operating system and software updates and patches current.
- Antivirus and malware/spyware detection software applications running and up to date.
- Back up services implemented for personal computers and work stations as appropriate (e.g., consider a commercial backup service).

<u>System Backups</u>

Project data should be protected from loss through preventative data system backup and recovery mechanisms. Data system backups should be performed on a periodic basis at a frequency to be defined by each Task. This frequency should be selected to minimize the extent of consequences of data loss and time required for data recovery. Recovery procedures should be developed and documented in preparation for the event of hardware or software failure.

Data System and Database Access

Tasks should protect systems and data from unauthorized access by implementation of administrative and procedural controls. Access controls should be managed based upon specific data user roles that are defined by the types of data and functionality required (e.g., a data management specialist needs the capability to create and update data while a program manager may need read-only access to perform on-line queries). The mechanism for implementing access control should be documented in project data management plans. Maintaining up to date computer security, including operating system patches, and as applicable antivirus and antispyware software on project data systems is essential.

Data Product Content Configuration Control

A Task should establish configuration control requirements for the contents of the Task data products. The requirements should ensure traceability of field and laboratory data from the original reported values, through authorized data changes, to current values stored in a data file or database. The configuration control should define the approval process required for making changes to data products and the documentation required for each change.

The minimum information maintained for each data change should include

- a description of the change;
- the reason for the change;
- the name of the individual making the change;
- the date of the change; and
- a copy of the data before the change took place.

Control of Erroneous Data

Practices are needed for controlling data that are erroneous, rejected, superseded, or otherwise unsuited for their intended use. These practices should provide for the identification, versioning, flagging, and/or segregation of inadequate data to avoid their inadvertent use. Task plans should describe the practices for controlling invalid data in your data systems.

Identification of Data Products

Practices should be established to assure that all data and data products are clearly identifiable and traceable to the Task from which they were produced. It is very important that this identification and traceability be maintained (protected) throughout the needed lifetime of the data. A description of practices to be used on your project should be included in Task plans.

Data Management System and Software Configuration Control Guidelines

This section provides guidance to project data managers responsible for the documentation, quality assurance, and configuration control of project software and data systems.

This software and computer system implementation guidance is applicable to projects using project specific software and an electronic database. The need for project-specific data systems, databases, and software will vary depending on the scale of the Task. This section discusses minimum documentation, QA, and configuration control guidelines for project specific implementations.

Project Database Documentation

Project specific databases include spreadsheets, data sets, and databases (e.g., Excel, ORACLE) defined by investigators and the project data management group to manage project data. The project specific databases should be described in the permanent project record. The description should identify the commercial database product used, the database name, structure, and locations.

The minimum database documentation will consist of

- name and version of commercial software used;
- names of project databases created;
- database structure definitions, including field names and descriptions; and
- storage location and media.

Project Software Documentation

Project specific software includes programs written by investigators, technicians, and the project data management group for data management tasks, and applications written for the production of data products. Data management tasks could include instrument data acquisition and processing, data conversions and derivations, and data quality control checks. Data products are defined as any extraction, summary, or analysis of data that results in a data summary or a hard copy product such as tables, graphs, statistics, or maps.

Software documentation should include the software program name, description, special requirements, author, revision, completion date, and documentation of the QA review. Data products documentation should also include all information to uniquely describe how the data product was produced, including the sources used, the manipulations made, and the tools used to produce the data product. Software documentation can be maintained in electronic or hard copy format or it may be included as comment blocks embedded within the project software program.

The minimum software documentation should consist of

- name and version of the commercial software used;
- name and version of the software program written by the project;
- author;
- date;
- revision;
- system requirements; and
- storage location.

Project Software Quality Assurance

The project should define the QA requirements for project specific software. At a minimum, software programs for data acquisition and processing, data conversions and derivations, data loading, data quality control checks, calculating statistics reported in project deliverables, and producing data products should be reviewed to ensure they meet the desired objectives. The reviewer should be someone other than the person who wrote the software program.

Project Specific Software Configuration Control

Project specific software should be protected from unauthorized modification or deletion. This can be accomplished by administrative controls or file security options provided by many computer operating systems. Changes to project software should be documented and a history of revisions which impact the results or data products should be included in the project file. Commercial products are available to maintain a record of software revisions [e.g., Revision Control Software (RCS)]. Another way to do this is to keep the initial or baseline software in a storage area separate from the working software. Then, when the software changes, the new software can be moved to this separate area also, there maintaining copies of all revisions. Project specific configuration and revision control should be documented in project plans.

The project software configuration control documentation should include

- commercial software used;
- program names;
- approvals
- revisions (including dates of revision); and
- storage locations.

Before the development of software applications, a requirements analysis should be conducted. Developed software applications should be tested and validated to ensure compliance with all user requirements and to provide confidence that the software will perform satisfactorily in service. The technical adequacy of results generated by these applications should also be reviewed by another person, tested and validated. Configuration management of the developed software application programs shall be conducted.

Appendices

Appendix XW: Tabular data file format guidance

Tabular Data

Use Consistent and Stable File Formats For Tabular Data (from DAAC Best Practices)

In choosing a file format, data collectors should select a consistent format that can be read well into the future and is independent of changes in applications.

Tabular Data:

Using ASCII file formats is the best way to ensure that measurement data are readable in the future.

- Use the same format throughout the file - don't have a different number of columns or re-arrange the columns within the file.
- Use a consistent format across all data files prepared for a study or project.
- Figures and analyses should be reported in companion documents - don't place figures or summary statistics in the data file.

At the top of the file, include several header rows.

- The first row should contain descriptors that link the data file to the data set, for example, the data file name, data set title, author, today's date, date the data within the file were last modified, and companion file names.
- Other header rows (column headings) should describe the content of each column, including one row for parameter names and one for parameter units.
- Column headings should be constructed for easy importing by various data systems. Headings should contain only numbers, letters, and underscores -- no spaces or special characters.

Within the ASCII file, follow these guidelines.

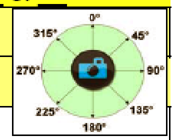
- Delimit the column headings and parameter fields using commas, tabs, or semicolons; these are listed in order of preference.
- Avoid delimiters that also occur in the data fields. If this cannot be avoided, enclose data fields that also contain a delimiter in single or double quotes.
- If the data fields us the comma as the decimal separator (rather than the period) the semicolon would be the preferred column delimiter.
- Don't include rows with summary statistics; it is best to put summary statistics, figures, and other comments in a separate companion file or in the data set documentation.

In the data set documentation, specifically add the following data file information.

- Description of the data file names, particularly if the file names are composed of multiple acronyms, site abbreviations, or other project specific designations.
- Expanded descriptions of the parameters (column headings) and their units of measure from the data file.
- Missing value codes.
- Example data file records.
- Other data file documentation, as listed in Section 7, that would be helpful to a secondary data user.

Appendix XX: Reporting Photos -- Forms

These draft forms are available as Excel spreadsheets.

SPRUCE Photo Log: "Inside looking out"						Page __ of __	
Event (Gen subject)					Date (yyyy/mm/dd)		
Level 1 (eg, Transect)				Level 2 (eg, Plot)			
Coordinates	Method:		All photos at same:		Notes:		
Decimal degrees to 5 decimal places.	Precision:		Latitude:				
	Datum:		Longitude:				
Level_1	Level_2	Time (hh:mm)	Photo_ID*	Subject	Heading_Code**	Latitude	Longitude
* Photo_ID syntax: (Level_2)_(Heading_code).jpg **Approximate degrees from true North or N, E, S, W.							

SPRUCE Photo Log: Landscape / object or "outside looking in" Page __ of __

Event (Gen subject)		Date (yyyy/mm/dd)		
Level 1 (eg, Transect)		Level 2 (eg, Plot)		
Coordinates	Method:		All photos at same:	Notes:
Decimal degrees to 5 decimal places.	Precision:		Latitude:	
	Datum:		Longitude:	

Level_1	Level_2	Time (hh:mm)	Photo_ID*	Subject	Heading_Code**	Latitude	Longitude

* Photo_ID syntax: (Level_2)_(Heading_code).jpg **Approximate degrees from true North or N, E, S, W.

Appendix XY: SPRUCE Task Data Descriptions -- Metadata and Documentation

The SPRUCE Task Description spreadsheet is an introduction to some of the metadata elements that will be compiled in the CDIAC metadata database to create the searchable index of SPRUCE data set. Other fields are valuable for identifying measured parameters, methods, and units that might be candidates for a picklist of standard values and other aspects of data and metadata reporting

Metadata reference:
Other metadata standards to consider are FGDC, NetCDF Attribute Convention for Dataset Discovery, EML, and ISO-19115.

https://www.nosc.noaa.gov/dmc/swg/wiki/index.php?title=Category:ISO_19115

https://www.nosc.noaa.gov/dmc/swg/wiki/index.php?title=NetCDF_Attribute_Convention_for_Dataset_Discovery

	SPRUCE Task		**Example DRAFT** **R2: Plant Growth, Phenology, and NPP Biogeochemical Cycling, Peat characterization**
Task Information	**Data collection description (brief)**		Collect foliar samples for baseline nutrient levels across S1
	Data Type (general)		foliar N and C content
	Start Date (tent) (mmm-yy)		**2009/09/22**
	Sampling / measurement locations identified? Coordinates?		approximately 20 locations across the bog, locations coincide with veg survey plots
	Likely outcome(s)		Identify patterns of N and C content
Field Sampling and Measurement Information	**Sampling / Measurement medium?**		Foliage, no stem
	Sampling / Measurement Frequency?		once
	Will samples or measurements be collected at multiple heights/depths above ground or below ground? Specify.		Spruce 1-2m, Ledum at canopy height (~50 cm)
	[add additional rows for Parameter/Sampling/Measurement/Prep/ Analysis/Method combinations as needed] **Parameters Measured**		**Carbon and Nitrogen**
	Units identified?		
	Field sampling method?		clippers
	Field measurement method?		

	Field sampling or measurement QC samples to be collected? (added 2010/03/03)		
	Field sampling / measurement comments: [reporting frequency, units, more?]		% foliar N or C on mass basis, composite samples of 3-4 individuals, for spruce only sampled current year foliage, samples kept cool for transport back to TN
Laboratory Analysis Information	**[If no laboratory activities, then all fields 'Not applicable']** **Laboratory sample preparation?**		dried at 70 deg C, ground. ~ 10 mg packed into capsules
	Laboratory analytical method?		combustion - C:N analyzer
	Field Sampling and Laboratory QC samples to be analyzed and reported? (added 2010/03/03)		
	Laboratory analysis comments: [reporting frequency, units, more?]		
Task Coordination	**Comments: [Coordination with other Tasks? More?]**		
Data Management Information	**Owner / steward of data**		Rich Norby
	Data provenance? Is your data traceable from field samples, through instrument output files, data processing, deriving additional parameters, to the final data product file?		
	Version control? How do you handle changing values and version control of final data products?		
	Archive Schedule for Task: **SPRUCE sharing**		
	Archive Schedule for Task: **Public sharing**		
	Define QA Checks: **Level 1**		
	Define QA Checks: **Level 2**		
	Comments:		

Appendix XZ: Example CDIAC "Mercury" Search Metadata Report

CDIAC

Carbon Dioxide Information Analysis Center • cdiac@ornl.gov

Metadata Report

Search Criteria:

Email

Search Again

Project Number:	**ORNL FACE Weather Data**
Product Title:	**ORNL FACE Weather Data**
Author(s):	
CDIAC Contact(s):	Lisa Olsen; email: olsenlm@ornl.gov
Data Set Location:	
Data Center URL:	http://cdiac.esd.ornl.gov/
Thematic Area:	Vegetation response to CO2 and Climate
Notes:	Weather data are available for 1999-2007. There is an hourly (*wh*) and a daily (*wd*) file for each of the six plots (rings) on the FACE site, giving a total of 12 files for each year. The 2007 files will be updated as data become available. Daily statistics were calculated only for those variables with at least 12 good hourly values, and the calculation of daily statistics for the variables, PAR2m, PAR22m, and PYRAN, required 12 good daylight hours. Otherwise, the daily values were set to missing.
Keywords:	carbon dioxide CO2 enrichment FACE meteorological data weather data ORNL FACE sweetgum Liquidambar Styraciflua vegetation response
Abstract:	ORNL FACE Weather Data
Site Location:	North: 35.9, West: -84.33, South: 35.9, East: -84.33 (Locate)
Site Information:	Oak Ridge National Laboratory FACE Experiment http://public.ornl.gov/face/ORNL/ornl_data_co2weather.shtml
Variable:	YEAR DOY HOUR ATMEAN2m (C) ATMIN2m (C) ATMAX2m (C) RH2m (%) PAR2m (umol/m2/s) STMEAN (C) STMIN (C) STMAX (C) ATMEAN22m (C) ATMIN22m (C)

	ATMAX22m (C) RH22m (%) PAR22m (umol/m2/s) PYRAN (W/m2) RAIN (mm) WIND (m/s)
Descriptive File:	http://cdiac.ornl.gov/ftp/FACE/ornldata/weather/weather.txt
Content Time Range:	Begin: 19990101 - End: 20070709
Metadata Author:	Name: Email: Phone:
Additional Documentation:	http://public.ornl.gov/face/ORNL/ornl_home.shtml
	View Metadata Source

CDIAC | ORNL DAAC | Mercury | Rate Us | Contact Us | Privacy, Security, Notices