

DATA MANAGEMENT FOR NREL AND BEYOND: A ROADMAP AND RECOMMENDATIONS

By: Nicole Kaplan, MS and Greg Newman, PhD
Version 1.1 – June 2013

Supported by 2012 NREL Program Development Funds

Summary: This report contains a summary of activities lead by Nicole Kaplan, RA, and Greg Newman, Research Scientist, and a strategy for managing data, which were supported by 2012 Program Development Funds at the Natural Resource Ecology Laboratory (NREL). We set out to describe current components, functions and expertise of the NREL cyber-infrastructure and inquire about current work and future needs for managing, archiving and providing access to data. The activities we coordinated included: (1) conducting a workshop, (2) arranging the Spring 2012 seminar series entitled *Data Literacy: Bridging the Gap Between Science and Society*, (3) contributing to over a dozen data management plans for NREL proposals, (4) providing database services to PHACE (a large collaborative project), and (5) attending conferences and workshops to obtain new ideas and build partnerships. The results of what we learned and accomplished include overall data management goals for NREL, a model for a data management system for NREL, requirements for a web-based data access and delivery system, policies for data sharing and attribution, training for undergraduate and graduate students, and scholarly work in Ecological Informatics (e.g. Chu et al. in press, Newman et al. 2011, Vanderbilt et al. 2009).

Our work was presented at the NREL 2012 Annual Retreat and the presentation file is available upon request. Here, we present resources for NREL to plan and conduct data management as part of the research process. We propose a model that establishes a hierarchy of workspaces to support, preserve, and secure the flow of data from the private laboratory space of the Principal Investigator, to a web-based NREL Data Repository to disseminate data to the public and stakeholders. We detail how current cyber-infrastructure capabilities can be leveraged to meet needs within NREL for data management, facilitate research discovery and data re-use, and contribute to NREL excellence. Major benefits to managing, sharing and re-using data include increasing the competitiveness of NREL proposals, influencing expectations of funding agencies for data management, accelerating global change research (Wolkovich et al. 2012), and broadening scientific understanding and services to support decision-making (Dozier and Gail 2009). In addition, federal research sponsors are committed to increasing open access to data and are requiring data management plans from agencies and departments with over 100 million dollars in research and development expenditures (Holdren 2013 and e.g. The Fair Access to Science and Technology Research Act (FASTR)).

Suggested Citation: Kaplan, N.E. and G.J. Newman (2013). Data Management for NREL and Beyond: A Roadmap and Recommendations. Digital Collections of Colorado. Colorado State University, Fort Collins, Colorado. <http://hdl.handle.net/10217/87381>

The work is licensed under a Creative Commons Attribution-ShareAlike 3.0 United States License. To view a copy of this license view: <http://creativecommons.org/licenses/by-sa/3.0/>



PROPOSED DATA MANAGEMENT APPROACHES

Mission Statement: Develop and contribute to a suite of sustainable cyber-infrastructure systems to support data management at the Natural Resource Ecology Laboratory (NREL) and more broadly Colorado State University (CSU) that facilitates long-term data curation, provides appropriate access to high quality data and information, and supports ecosystem and sustainability science for a variety of stakeholders.

Data Management Goals:

- Provide a full suite of data management services to meet mandates from funding agencies
- Complement existing NREL content management and information delivery systems (e.g. International Biological Information System, NREL Central Website, MSP Website)
- Increase effectiveness and efficiency of NREL science

We are in the age of networked science (Nielsen, 2012), or collaborative eScience, that mandates easy access to distributed information (NSF's Cyberinfrastructure Vision 2006) and capabilities to manage big data - meaning not only a deluge of digital data, but also volumes of heterogeneous data typical in the environmental sciences (Borgman 2007 and Maier 2012). We need a framework to meet and navigate new requirements and procedures to support data access, management, integration and synthesis (Killeen 2012). Ecologists rely on access to data and are expected to provide open access to data. We are all consumers of digital information and depend on technology to conduct our research. We need a suite of data management systems and services that incorporates principles of good data management (Table 1) and our values to preserve a private work space, while curating a suite of shared resources for the community with controlled access. We want the public face of NREL data to foster competitive proposals and communicate important research results – coupled with rich project context and actual research data. Such a suite of data management systems and services will enable us to meet our data management goals and streamline data management work – from writing data management plans for proposals to processing data from the field or lab into long-term data repositories to telling the story of our research.

TABLE 1: PRINCIPLES OF GOOD DATA MANAGEMENT	
(1) Start Early	
(2) Have a Plan	
(3) Ensure No Data Loss	Preserve core observational data for re-use by colleagues in the broader research community
(4) Provide Useful Data	Document and preserve metadata and the stories behind the data
(5) Maintain Excellence in Research and Teaching	Contribute to the integrity of research and the reputation of the lab by making data available
(6) Ensure Repeatability of Research Projects and Complex Models	

Enable researchers to continue to make sense of current and future analyses
(7) Provide Access (Accelerate Science and Career Development) Make sure scientists can discover, combine, and/or re-use your data, and get credit for such work
(8) Provide Proper Attribution and Data Use Policies Have a policy for data access and reuse Minimize or eliminate restrictions on use Provide a tag line or citation for attribution Choose a Creative Commons license (http://creativecommons.org/choose/), etc.
(9) Review Plan Frequently Conduct quarterly reviews, before each field season
(10) Identify and Follow Best Practices in your Domain Examples include Ecological Metadata Language (EML), GIS FDGC Standards, Darwin Core, etc.
(11) Always Use Data Management Best Practices, Not Only Per Funder Requirements Share and communicate results and get attribution for all your research products, including data products
(12) Train the Next Generation Offer courses and workshops on data management as part of learning how to conduct research and/or contribute to scientific collaboration
(13) Leverage Partnerships and Existing Repositories Integrate with existing systems (e.g., CSU Libraries , NREL , DCC , UC3 , DataONE).
(14) Embrace Policies Support Data Re-Use Attribution - our social responsibility to credit scientists, funders, and/or organizations Usability - assurance that data can be accessed, re-used and then cited appropriately Persistence - ability to locate and cite data products using persistent identifiers (Handle, ARK, DOI, GUID, ORCHID)

Our model (Figure 1) is comprised of three tiers. In Tier I, the majority of current NREL research is (and will most likely continue to be) conducted privately by individual Principle Investigators (PIs), researchers, or graduate students on personal computing spaces available. In Tier II, collaborative work will be conducted by research teams as directed by PIs and some datasets may be shared for easy access among NREL researchers. This second layer of the proposed model (Tier II) represents the current NREL shared drive system, managed by centralized NREL IT personnel for controlled access to research team data and information. This layer is indicative of a shared drive for any given Principle Investigator (PI) and may also include options for NREL researchers to post datasets for use by other NREL researchers on an opt-in basis. No system currently supports this functionality in a formal, centralized, and inventoried manner, but

discussions have been initiated about possible solutions down the road. In Tier III, multiple opportunities exist for NREL researchers to document, store, curate, disseminate, and generally make accessible, scientific information to the vast sea of ecological data – as data become more finalized they become more primed for public access and use through various available systems and services.

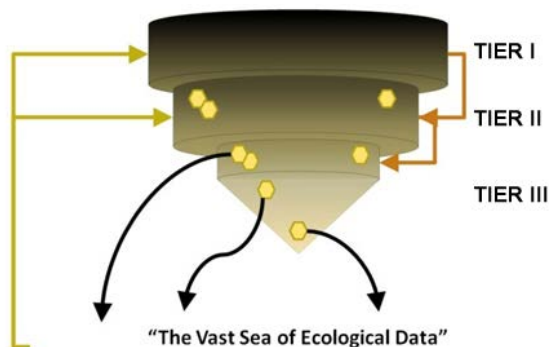


Figure 1 – Proposed Conceptual Model: Navigating the 'Sea' of Ecological Data

We recommend leveraging several current and future initiatives to support key aspects of the third tier (Tier III) of the proposed model. These initiatives include numerous existing and growing local resources and services such as the NREL Central Website, the NREL International Biological Information System (IBIS) cyber-infrastructure and team, creating your own NREL-hosted website, and the CSU Libraries Digital Repository. Various other repositories and services exist, including DataONE data repositories, using the DataONE DataUP tool to upload data, or other domain specific repositories such as the ORNL DAAC for biogeochemical data, ACADIS for data collection efforts

in the Arctic and the GenBank system for genetic data. Each of these alternatives offers various tools and has pros and cons that make them more or less applicable to various NREL research projects. We recommend using a suite of resources and services available that most closely align with your research project goals and objectives (See Table 2). Detailed descriptions of four specific alternatives (e.g. NREL Central Website, IBIS, your own website, and CSU Libraries) are included in the next section of this report. Any combination of these resources or services may be used in concert to meet your needs.

TABLE 2: TIER III ALTERNATIVES				
Service/Resource/Capability*	NREL Central	IBIS	Your Own ***	Library
(1) Archival of informative or historical hard-copy documents, reports, books, prints, and field notes			?	X
(2) Attribute level metadata documentation and standardization		X	?	
(3) Bulk Data Uploads to data management system	X	X	?	
(4) Customizable Data Entry Sheets	X	X	?	
(5) Customizable Feedback Surveys		X	?	
(6) Customizable Lab Website	X		?	
(7) Customizable Personal Profile	X	X	?	
(8) Customizable Project Profile	X	X	?	
(9) Data Backup	X	X	?	X

(10)	Data Management Consultation	X	X	?	X
(11)	Data Management Plan Development	X	X	?	X
(12)	Data Processing Automation		X	?	
(13)	Data Visualization	X	X	?	
(14)	Database Development	X	X	?	
(15)	Downloadable, Curated Data Packages			?	X
(16)	Downloadable Data Tables	X	X	?	X
(17)	File Format Conversion and Upgrade Services			?	X
(18)	Long-Term Storage & File Upkeep	X	X	?	X
(19)	Mapping Capabilities		X	?	
(20)	Persistent URLs			?	X
(21)	Quality Assurance / Quality Control (QA/QC)		X	?	
(22)	Query-able Datasets	X	X	?	
(23)	Search Optimization Meta Tags	X	X	?	X
(24)	Server Administration	X	X	?	X
(25)	Standards Compliant Websites	X	**	?	
(26)	Transfer of large datasets through DMZ				X
(27)	Web Application Development	X	X	?	
(28)	Web Mapping		X	?	

* Note: Service/Resource/Capability categories are listed in alphabetical order and do not represent level of importance.

** Note: IBIS websites will become standards compliant throughout 2013.

*** Consider the tradeoffs in flexibility versus development and maintenance costs in developing your own system.

These are a few ideas from what we've learned the past year, but to navigate the sea of data researchers need many differently equipped vessels – sets of skills, tools, know-how, etc. Thus, our recommendations are to respect the way in which people work, and to leverage the systems and partnerships we already have and can further cultivate.

The NREL Central Website to serve and discover professional profiles and project information

Jonathan Straube and Ty Boyack lead the development, maintenance, testing, and administration of the current NREL Central Website, which runs in part with a web-based Content Management System (CMS). The history of the NREL Central Website is that the first iteration of the system used the freely available Open Source Joomla CMS out of the box CMS software system. The current system uses the freely available Open Source ModX CMS system. Each of these systems allows logged in users with certain permissions to contribute and modify NREL web content. These systems were built to conform to CSU website standards and meet W3C web standards as well. The current system allows NREL PIs to develop their own "Researcher Lab Pages" that contain (1) a customizable "Directory Page" serving as the PIs' customizable professional profile page and (2) a customizable "Home Page" serving as each Principle Investigator's customizable lab website. The Directory Page enables PIs to add/edit College Directory News, Contact Information, Department Affiliations, Bio, Education, Certifications, Awards/Honors/Grants, Publications, Selected Articles, Presentations, memberships, Interests & Studies, and Web Links. The Home Page section allows PIs to create their own lab website, including features such as their own customizable navigation and their own custom web content. This allows PIs to upload images and enter text along with adding hyperlinks to internal and external web content and web pages. The benefits include the ability to write HTML and perhaps dynamic PHP scripts. Using the CMS, PIs can create standards based HTML. To upload images, PIs may use third party software such as GIMP or Adobe Photoshop to process/resize the images properly before uploading. The NREL IT staff offers training sessions on how to use the CMS and are available to assist PIs at any time. The system yields relatively consistent user interface layouts for all Project lab Home Page websites while offering flexibility in content for each website. An example of a researcher lab website is shown in Figure 2 below and illustrates how the extensible CMS can be used to create a website within the current NREL Central Website.

Colorado State University CSU Directory Search CSU Apply to CSU





Natural Resource Ecology Laboratory



Ecosystem Science and Sustainability

HOME STUDENTS CURRENT RESEARCH » SELECTED PUBLICATIONS EXTERNAL LINKS

The Evangelista Lab



Dr. Paul H Evangelista
 Research Scientist
 NREL, Colorado State University

NESB B234
 Natural Resource Ecology Laboratory
 Campus Delivery 1499
 Fort Collins , CO 80523-1499

Phone: (970) 491-2302
Fax: (970) 491-1965
Email: paulevan@nrel.colostate.edu

[Curriculum Vitae](#)

Paul is currently a Research Scientist II at the Natural Resource Ecology Laboratory and holds affiliate faculty/researcher positions with the Department of Ecosystem Science and Sustainability, the Graduate Degree Program in Ecology, and the USDA UV-B Monitoring and Research Program. His research interests span across multiple disciplines, including invasive species, wildlife and habitat, forest ecology, and climate change. He has worked in Ethiopia since 1999 on several high-profile endangered wildlife species, biodiversity inventories, and ecosystem services that are critical to local people. From 1999 to 2005, Paul worked in the Grand Staircase-Escalante National Monument in southern Utah conducting floral inventories funded by the BLM, and later by NASA. During this time, he served as the Systems Ecologist for the Secretary of the US Department of Agriculture's Advisory Board for the Monument. In recent years, his work has centered on modern technologies such as remote sensing, GIS, and spatial models to support management and conservation of natural resources across spatial and temporal scales. This includes mapping distributions of native and non-native plant species, wildlife, insects, land-cover and land-use, and backcountry recreational activities. All of his research incorporates training and outreach elements designed to build geospatial capacities for natural resource managers and other stakeholders through a series of on-line training programs and several internship programs supported by AmericaView, NASA's Develop Program and the North Central Climate Science Center.

Contact NREL Disclaimer Equal Opportunity Privacy Statement

© 2012 Natural Resource Ecology Laboratory
Fort Collins, Colorado 80523 USA

Figure 2 – Example of a Home Page for a Researcher Lab Page created by a PI using the existing NREL ModX CMS.

IBIS: NREL Cyber-infrastructure to serve and visualize searchable data and related information through an NREL Data Repository

Dr. Greg Newman leads and coordinates a local team of experts to evaluate, design, develop and maintain ecological informatics platforms to improve the accessibility of biological data for various outcomes including conservation, environmental literacy, decision support, spatial modeling, visualization and re-use. The International Biological Information System cyber-infrastructure is well-positioned to meet new expectations of National Science Foundation and other funders for data access and interoperability. Dr. Newman’s team and IBIS infrastructure they’ve developed supports 20+ websites, including an NSF-funded citizen science website (www.citsci.org), the Global Invasive Species Information Network (www.gisin.org), the USDA ID Source search engine (<http://idsource.colostate.edu>), the Math and Science Partnership environmental literacy assessment website (MSP; <http://ibis.colostate.edu/msp>), and a newly developed NPS Repeat Photography website (<http://ibis.colostate.edu/ccep>). Dr. Jim Graham, Greg and others created the IBIS platform over eight years ago, a platform that, over time, has proven reliable,

flexible, and sustainable. It currently stores and curates over 308,600 unique species occurrence records, some 500,480 taxonomic records, over 172,100 species profiles, 533,450+ species and site location attributes, and 2,727 pest identification resources / aids / website URLs. The MSP website provides controlled access, meeting Integrity in Research Board (IRB) requirements, to over 1 million student literacy assessments and a data archive to facilitate researchers' analysis and re-use of coded quantitative and qualitative data.

The IBIS cyber-infrastructure supports (1) customizable "User Profile" pages, (2) customizable "Project Profile" pages, and (3) customizable data entry sheets. Customizable feedback forms akin to Survey Monkey, support data entry for species occurrences and associated attribute measurements, support site characteristic data entry and associated attributes, and data entry for single observations as well as bulk data uploading capabilities from tab delimited text files and ESRI shapefiles. IBIS also supports the ability to upload photos that are automatically resized into three separate sizes/resolutions. Query and search functions contain algorithms to discover projects and species alphabetically and based on search criteria, filter data based on species, date, project, and location, and provide web mapping capabilities. The IBIS Project Profile Page enables PIs to add/edit the Project Name, Project Manager, Website URL(s), Start Date, End Date, Project description, Purpose, Study Extent, Status (e.g., starting, active, closed, etc.), Sample Design(s), QA/QC Procedure(s), Sampling Design Descriptions, QA/QC Procedural Descriptions, Funding Organization, Research Question(s), Project Training Materials and Other Resources such as Related Publications, and uploaded Media such as Team Photos or Videos. It also dynamically summarizes the total number of data records (e.g., number of observations) collected and entered to date in the project as well as all observations/data records that are searchable. It allows PIs to create their own data entry sheets to specify what they wish their graduate students or research associates to collect in the field and eventually enter into the database. These data sheets allow PIs to select the species to be observed, the measurements to be made (e.g., stream pH, length of fish, height of plant, DBH of tree, sex of fox, soil texture of soil pit, etc.). The system maintains a growing yet vetted set of standardized attributes/measurements and conforms to standards in each domain of study (e.g., it uses the standardized call index for amphibian monitoring (Nelson and Graves 1967)). In this way, it is a collaboratively built ecological data repository built by its users. The IBIS Species Profile Page enables approved taxonomic experts to add/edit species common names, description, habitat, unique features, invasive history (if invasive), life history, and citations, upload photos of the species to a species profile photo gallery and add external web links.

We propose to leverage the IBIS cyber-infrastructure to develop a portal to serve project data and metadata, data products, and value-added data related tools and services. We envision this portal serving the data access needs of the NREL researcher and broader ecological research community by allowing easy submission of and access to finalized research data. The IBIS cyber-infrastructure also accommodates front-facing stories and metadata about each dataset served. Because the IBIS cyber-infrastructure supports various user levels and project roles, we believe this system is ideally suited to meet the needs identified by NREL researchers. We acknowledge that no data management system will address the data management needs of all research projects; thus we anticipate that IBIS will support NREL research focused on species observations, environmental measurements, and environmental education and we

will share data with other domain specific repositories where appropriate (e.g. sequence data collections). We envision an optional web portal accessible from, and linked to, the NREL Central Website to provide access to research data, tools for managing or working with data, and communicating supplemental information. This web portal would be on an “opt-in” basis for each NREL principle investigator and could be configured to “pull” basic data already being entered by PI’s into the NREL Central Website into the NREL data portal as to eliminate the need to enter the same basic information and two places and avoid redundancy of information. In addition to descriptive project information, attribute metadata, datasets, models, and algorithms, additional media could be served from other online digital repositories (e.g., Digital Collections of Colorado - the CSU Digital Repository, and DataONE member nodes) through the IBIS portal and/or linked from the IBIS portal. The IBIS system would exchange information and data regularly with other data repositories (e.g. DATAONE, LTER, NEON) as required by funding agencies through web services and/or automatic data exchange protocols. Permissions to access content and data for each project on the portal would be managed by project managers (e.g., NREL principle investigators) and NREL system administrators online, thereby eliminating the need for system users to know programming. The portal would follow a service-based, open architecture so that other stakeholders could use data delivered by the NREL digital repository portal in creative ways for decision support, natural resource management, data integration, synthetic research, meta-analyses, conservation, or education.

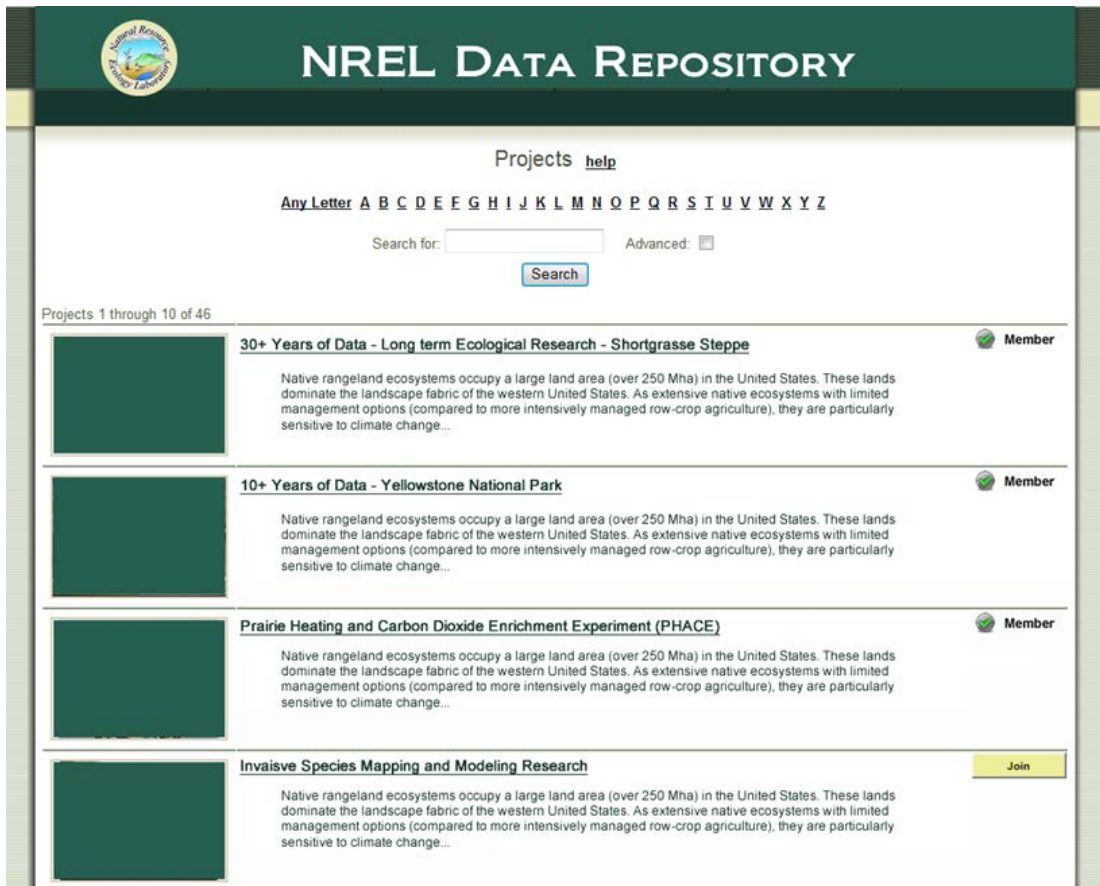


Figure 3 – Mock-up of the primary list page listing current and legacy NREL research projects accessible through the proposed web portal.

The proposed web portal would offer principle investigators customizable “Project Profile” pages that deliver project datasets, summaries and abstracts, data access at the record level; metadata access; graphing, mapping, and data visualization capabilities; media (pictures and video) delivery; and access to the stories behind the data. Projects would be searchable by keyword and be able to be filtered by theme, topic, or principle investigator and would link back to project or professional information contained within the NREL Central Website or other webpages. Data documentation would meet metadata standards, such as the Ecological Metadata Language (EML) or the Federal Geographic Data Commission (FGDC) geospatial metadata standard. Each project will have their own dynamically generated “Project Profile” page where members of the public will be able to access and download data deemed ready to distribute by the PI (project manager or data manager) and where project contributors will be able to edit project metadata, post and document new datasets or related publications, add photos, and contribute stories. The pages will be akin to facebook pages for a given research project. Dynamic links would be able to be created to other research portals such as Research Gate as desired by each project manager / PI.

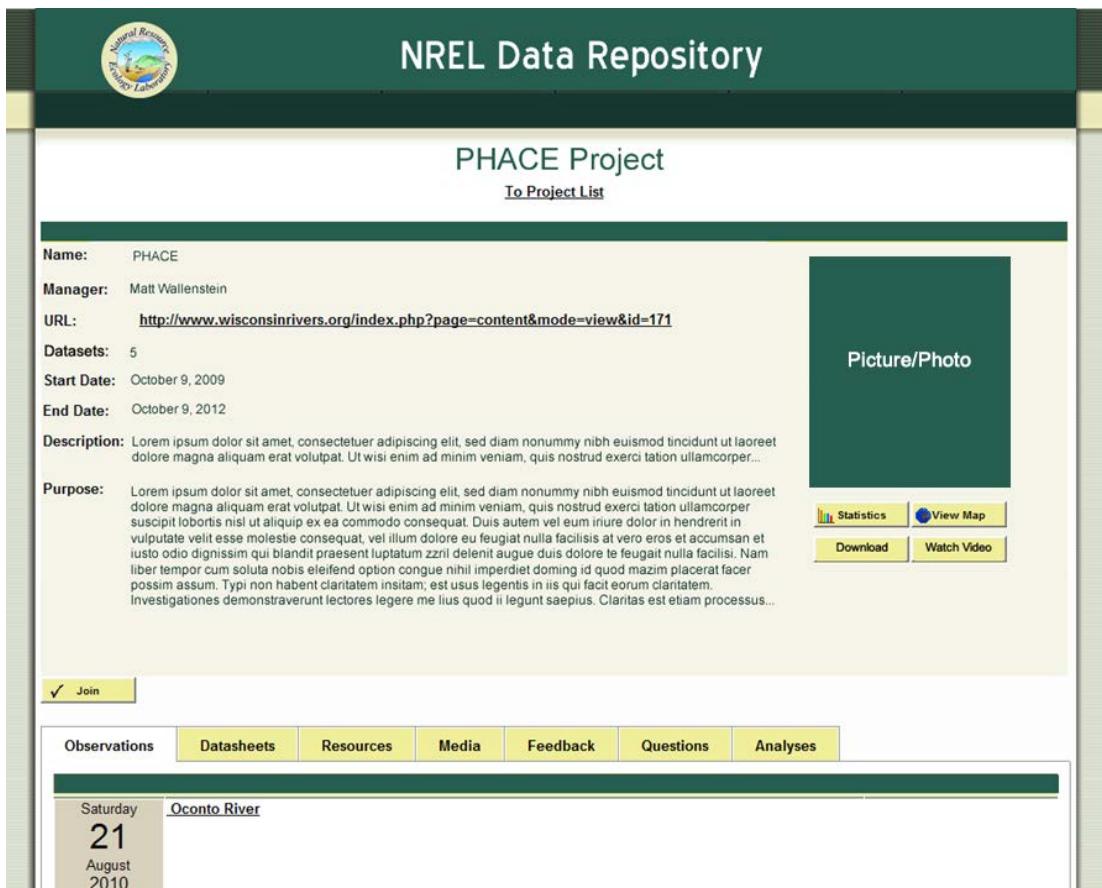


Figure 4 - Mock up of Project profile pages for each project managed and configured by each project manager (principle investigator).

NREL RFD: NREL cyber-infrastructure to share files within the community

The existing NREL file server can be expanded to be used as a platform for sharing datasets and other digital products among scientists at NREL. We recommend that a new Research File Directory be created and managed as a library of re-usable, finalized digital objects (Figure 4). Access and permission will be managed by leveraging WCNR accounts and associated permissions and augmenting these permissions with NREL PI assigned permissions. Objects available will be organized and inventoried with full documentation, including suggestions for appropriate re-use and/or re-analysis and method employed in creation of the data product. Policies and data re-use agreements will ensure attribution. Example objects include PRISM geospatial data, NLCD data, LiDAR, LandSAT, and DEM datasets that are deemed useful to the lab by those contributing.

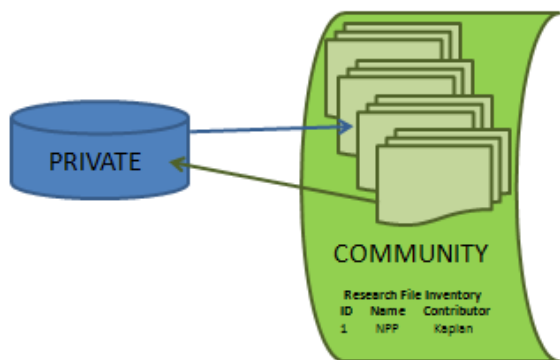


Figure 4. Proposed model for a community resource of finalized digital research files available for re-use by NREL community members with permission. PIs and/or staff work in their private work space in various configurations. Their finalized products, available for re-use, are transferred and stored on a centralized Research File Directory. Items contributed to the RFD can be inventoried for contents, controlled for access, and documented to guide appropriate re-use. The current NREL server supported by CSS may be expanded to support a community RFD.

Next Steps: We propose to apply what we have learned from the data management activities and needs assessments conducted over the past several months and used them to design requirements for a full suite of data management resources, tools, and services that can be available to NREL scientists. We propose to use data from the SGS-LTER program (a long-term series of projects currently being sunsetted that have produced long-term datasets of interest to the broader ecological community) to pre-populate and test a pilot NREL Data Repository. The SGS-LTER has its roots in the NREL, so archiving those data at NREL in the proposed portal seems appropriate as their history is tied to NREL as an organization. In addition, SGS-LTER funds are available to support transfer of data products to the proposed system. As time permits, as proposals are funded, or our cyber-infrastructure is further developed, we will conduct data audits to learn more from NREL PIs about data practices, and inventory potential final research data products available for re-use by the NREL community. We will work closely with Jonathan and Ty on refining and creating the design of the RFD and developing the proposed IBIS portal to complement the NREL Central Website and the CSU Digital Repository at the Library. We will pursue other opportunities to grow these tools and will continue to push these systems to accommodate new and exciting data types, such as those from mobile applications and crowd sourcing as to maintain both reliable data management and delivery services and a competitive edge for the NREL. Lastly, in the following pages we have provided resources for the researcher, including approaches to data management, recommended elements of a

data management plan, a brief description of green and gold levels data management services at NREL, and information on NSF data management requirements.

RECOMMENDED DATA MANAGEMENT APPROACHES	
(1) Start Early	
(2) Identify requirements and/or approach	<p>If a data management plan (DMP) is not required as a section in a proposal, consider if the research plan is complex, federated, distributed, and/if data, education, or if analysis products would be of interest to the broader community and accessible at some point during or after the project. If so, include a DMP in the research or project management description.</p> <p>The scope of the DMP may reflect the size or complexity of the project</p> <p>Leverage CI that you or your colleagues have and grow CI only when necessary</p>
(3) Consult research team members, students, colleagues, CI experts, and repository curators	
(4) Identify and follow Best Practices in your domain	
(5) Keep it simple, doable, concise and clear	<p>Have someone outside the project not familiar with your practices review it</p>
(6) Keep data accessible	<p>Have a policy for access and reuse</p> <p>Minimize or eliminate restrictions on use</p> <p>Provide a tag line or citation for attribution</p> <p>Choose a Creative Commons license (http://creativecommons.org/choose/)</p>
(7) Review frequently with the team	<p>Quarterly reviews, before each field season</p>
(8) Note for International Collaborations:	<p>There may be cases where data management plans are affected by formal data protocols established by large international research consortia or set forth in formal science and technology agreements signed by the United States Government and foreign counterparts. Be sure to discuss this issue with your sponsored projects office (or equivalent) and your international research partner when first planning your collaboration. (From NSF Data Management Plan Frequently Asked Questions: http://www.nsf.gov/bfa/dias/policy/dmpfaqs.jsp)</p>

RECOMMENDED ELEMENTS OF A DATA MANAGEMENT PLAN

- **Describe Cyber-infrastructure (or computing resources and expertise)**

- ✓ describe the arrangement(s) at your institution or across institutions
- ✓ describe how you will share data with your collaborators
- ✓ name a dedicated person to be responsible for data management
- ✓ highlight person or team's data management experience
- ✓ describe off-site backup procedures

- **Describe the data and the methods used to collect or create the data**

- ✓ describe whether data will be observational, experimental, mined, derived etc.
- ✓ describe data themes (e.g. abiotic, sensor data, community diversity data etc.)
- ✓ describe data storage types (e.g. excel spreadsheet, database, images, ASCII)

- **Indicate how data will be documented throughout the research project**

- ✓ describe involvement of data management personnel in research process
- ✓ consider creating a data inventory or tracking system

- **Describe how data will be assured for quality**

- ✓ describe QAQC procedures
- ✓ describe enactment of standards or specifications

- **Detail how data will be made available for collaborators, and/or public re-use**

- ✓ describe your project website, NREL website, IBIS or other community, network or domain online repository, ftp site, etc.

- **Include plans for long-term preservation and data access**

- ✓ provide persistent URL and metadata associated with the data
- ✓ describe secure storage and backup systems (e.g. LOCKSS)
- ✓ describe how format changes and migration will be addressed
- ✓ describe your local NREL-IBIS data delivery system and value-added features
- ✓ describe enhanced data discovery and synthesis opportunities through partnerships (e.g. CitSci.org, DataONE, KNB)

- **Include a data access policy statement**

- ✓ list any exceptional arrangements that might be needed to protect sensitive species information, participant confidentiality or intellectual property

- **Report on progress in managing data in annual and final reports**

- ✓ describe system development, number of data records captured, QAQC procedures, metadata documentation efforts, data re-use scenarios, etc.

NREL DATA MANAGEMENT SERVICES		
GREEN LEVEL OF SERVICE	<p>“Do It Yourself” Plan</p> <p><i>(some professional consultation)</i></p>	<p>Access to emerging standards, tools and templates to support data curation, discovery, and download capabilities for your data packages within a project website, the Digital Collections of Colorado, or other online data repository.</p>
		<p>Access to Centralized NREL Computing Services and current website Content Management System (CMS).</p>
		<p>Specific Services: Customizable WCNR directory page, customizable WCNR/NREL researcher website pages, centralized IT support for desktop PCs, printing, poster printing, copying, and license support for GIS and statistical software packages, server administration and backup for shared LAN storage, HPC computing services.</p>
GOLD LEVEL OF SERVICE	<p>NREL-IBIS Team Service Plan</p> <p><i>(dedicated professional data management, application development, and IT infrastructure support)</i></p>	<p>Access to dedicated informatics professionals for help writing and addressing all components of your data management plans, database(s), and data collection design; data management training for your research team; and data management coordination for large interdisciplinary projects.</p>
		<p>Access to the IBIS and NREL cyber-infrastructure systems to support advanced features including online tools for data entry, bulk data submission, data discovery, data downloads and access queries, data integration, synthesis and mash-ups, as well as innovative web application design and development.</p>
		<p>Specific Services: Data management plan development and editing, metadata documentation, data management consultation, database development, QA/QC services, data processing automation, customizable IBIS project page, customizable data entry pages, web mapping, data visualization, web application development consultation, web mapping consultation, server administration and backup services.</p>

Data Management Plans typically state what types of data will be created and how, outline procedures for analysis, QA/QC and back-up, and describes final products and plans for sharing and long-term preservation. You may note what is appropriate given the nature of the data and any restrictions that may need to be applied. Consider that there are different levels of data management needs that we are addressing in Green and Gold levels of service. We recommend a data manager be dedicated to any project to help with data management work and to ensure your data archive will be created with high integrity, and be sustainable and accessible. The Green level of service may satisfy needs for design of databases, data entry and QA/QC procedures, metadata documentation, and data integration and dissemination. The Gold level of service may include additional functions such as a dedicated data management staff to stay involved in the research process, reporting on data management progress, training for team members, developing additional system features, programming or new application design, and supporting synthesis activities with collaborators.

We are beginning to get a sense of budget costs required at different levels of service and are available to help you with determining your needs at the time of proposal writing. Please see our summary sheet on what Green and Gold levels of service entail.

**IBIS is proposed to be the public web-based portal for NREL Data Repository*

Templates for Data Management Plans and Data Access Policies are available upon request.

NSF Data Management Requirements:

As of January 18, 2011, all proposals to NSF must include a supplementary document of no more than two pages labeled "Data Management Plan." This supplement should describe how the proposal will conform to NSF policy on the dissemination and sharing of research results (see AAG Chapter VI.D.4). The NSF policy includes the sharing of results, primary data, physical samples and collections. This policy also mentions that NSF will enforce this policy through a variety of mechanisms and provide appropriate support and incentives for data cleanup, documentation, dissemination, and storage. NSF suggests that the plan "may" contain:

1. the types of data, samples, physical collections, software, curriculum materials, and other materials to be produced in the course of the project;
2. the standards to be used for data and metadata format and content (where existing standards are absent or deemed inadequate, this should be documented along with any proposed solutions or remedies);
3. policies for access and sharing including provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements;
4. policies and provisions for re-use, re-distribution, and the production of derivatives; and
5. plans for archiving data, samples, and other research products, and for preservation of access to them.

New Requirements for the Directorate of Biological Sciences (As of Feb 20, 2013):

DMPs will be considered during the merit review process, to help reviewers, and as appropriate, please organize the DMP as follows:

1. Describe the data that will be collected, and the data and metadata formats and standards used.
2. Describe what physical and/or cyber resources and facilities (including third party resources) will be used to store and preserve the data after the grant ends.
3. Describe what media and dissemination methods will be used to make the data and metadata available to others after the grant ends.
4. Describe the policies for data sharing and public access (including provisions for protection of privacy, confidentiality, security, intellectual property rights and other rights as appropriate).
5. Describe the roles and responsibilities of all parties with respect to the management of the data (including contingency plans for the departure of key personnel from the project) after the grant ends.

POST-AWARD MANAGEMENT

After an award is made, implementation of the DMP will be monitored through the annual and final report process and during evaluation of subsequent proposals. Data management must be reported in subsequent proposals by the PI and Co-PIs under “Results of prior NSF support”.

Annual project reports required for all NSF multi-year awards must include information about progress made in data management and sharing of research products (e.g., identifier or accession numbers for data sets, citations of relevant publications, conference proceedings, and other types of data sharing and dissemination)

Final project reports required for all NSF awards should describe the implementation of the DMP including any changes from the original DMP and contain the following information:

- The data produced during the award period
- The data that will be retained after the award expires
- How the data will be disseminated and verification that it will be available for sharing
- The format (including community standards) that will be used to make the data – including any metadata – available to others
- Where the data generated by the project has been deposited/is being stored for long-term public access

References:

Borgman, C. 2007. *Scholarship in the Digital Age: Information, Infrastructure, and the Internet*. The MIT Press. Cambridge, MA.

Chu, C., J. Norman, R. Flynn, N. Kaplan, W.K. Lauenroth, P.B. Adler. Cover, density and demographics of shortgrass steppe plants mapped 1997-2010 in permanent grazed and ungrazed quadrats. *Ecology*, in press.

Holdren, J.P. 2013. Memorandum for the heads of executive departments and agencies, Increasing access to the results of federally funded scientific research. Executive office of the President, Office of Science and Technology Policy, Washington, DC. (Accessed 10 June 2013: http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf)

Killeen, T. 2012. Exploring the Life Cycle, Citation and Integration of Geo-Data. Bridging Data Lifecycles: Tracking Data Use via Data Citations Data Workshop, University Corporation for Atmospheric Research (UCAR), Center Green Campus, Boulder, CO, April 5-6, 2012. (http://library.ucar.edu/content/events/data_workshop/presentations/Killeen_BridgingDataLifecyclesWorkshop.pdf, accessed 01 August 2012)

Maier, D, V. M. Megler, A. M. Baptista, A. Jaramillo, C. Seaton and P. J. Turner. 2012. Navigating Oceans of Data. *Scientific and Statistical Database Management: Lecture Notes in Computer Science*, 2012, Volume 7338/2012, 1-19, DOI: 10.1007/978-3-642-31235-9_1

National Science Foundation's Cyber-infrastructure for the 21st Century Discovery. 2006. National Science Foundation Cyber-Infrastructure Council.

Newman, G., J. Graham, A. Crall, and M. Laituri. 2011. The art and science of multi-scale citizen science support. *Ecological Informatics* 6:217-227.

Nielsen, M. 2012. *Reinventing discovery: the new era of networked science*. Princeton University Press, Princeton, New Jersey.

Vanderbilt, K., J. Cushing, J. Gao, N. Kaplan, J. Kruger, C. Leroy, J. Mallett, K. Ramsey and L. Zeman. 2009. Data integration challenges: an example from the International Long-Term Ecological Research Network (ILTER). *Ecological Circuits*. 2: 12 – 13