

Erich Seamon, MS PMP GISP Environmental Data Manager REACCH College of Agricultural and Life

Sciences University of Idaho

http://reacchpna.org

208.885.1230

erichs@uidaho.edu

REACCH and Data Management: Strategies and Implementation

Paul Gessler

Professor of Remote Sensing and Geospatial Ecology

Department of Forest, Rangeland, and Fire Sciences

College of Natural Resources,

University of Idaho



Broad Overview of Cyberinfrastructure and Data Management Initiatives

- REACCH Physical Network Connections
- REACCH Cell/Smart Phone Coverage
- Northwest Knowledge Network
- Complementary Collaborations
- REACCH Web Site and Data Portal
- Erich

Idaho Regional Optical Network

Presented to: University of Idaho - Moscow

Brad Jordan January 19, 2012

IRON Network 2012



- Core Nodes in Boise, Pullman and IDF
- Aggregation in Spokane, SLC, Clarkston and Lewiston.
- 1GE/10GE Backbone and Access connections
- Connectivity to neighboring RONs (PNWGP, UEN)

Northern Idaho Network



- Redundant 10GE path from Pullman
- Single point of failure to Moscow from Pullman
- Existing fiber along SR-95

Northwest Networks with IRON



Northern Tier Network Consortium



Global Networking







Northwest Knowledge Network (NKN) A Regional Approach for Research Data Lifecycle Management

The NKN mission is to make cross disciplinary research data more accessible, comprehensible, usable, and secure for data providers and users by promoting commonly accepted policies, standards, and protocols and supplying services (e.g. secure data storage, seamless discovery, exploration, transfer of data between and within user groups, etc.) that are adaptive to changing data management and storage needs.

Northwest Knowledge Network (NKN) A Regional Approach for Research Data Lifecycle Management



The 3 Functional Areas of NKN

Northwest Knowledge Network (NKN) The 3 Functional Areas of NKN

To incorporate advancements in cyberinfrastructure research into the NKN architecture and services.

CI

Research

To provide tools and services that enable interdisciplinary research between geographically distributed researchers.

Service Center

To develop policies and procedures to facilitate data sharing and preservation between academic institutions and interested state and federal agencies.

> Data Sharing Cooperative





About

What is DataONE? DataONE Organization Working Groups Partners Communication Videos Contact Us

Find it Fast

Data Management Planning Best Practices Software Tools

What is DataONE?

Home » About » What is DataONE?

Data Observation Network for Earth (DataONE) is the foundation of new innovative environmental science through a distributed framework and sustainable cyberinfrastructure that meets the needs of science and society for open, persistent, robust, and secure access to well-described and easily discovered Earth observational data.

Supported by the U.S. National Science Foundation (Grant #OCI-0830944) as one of the initial DataNets, DataONE will ensure the preservation, access, use and reuse of multi-scale, multi-discipline, and multi-national science data via three principle cyberinfrastucture elements and a broad education and outreach program.

Cyberinfrastructure

Coordinating Nodes:

DataONE currently hosts three Coordinating Nodes that provide network-wide services to enhance interoperability of the Member Nodes and support indexing and replication services. Coordinating Nodes provide a replicated catalog of Member Node holdings and make it easy for scientists to discover data wherever they reside, also enabling data repositories to make their data and services more broadly available to the international community. DataONE Coordinating Nodes are located at the University of New Mexico, the University of California Santa Barbara and at the University of Tennesse (in collaboration with Oak Ridge National Laboratory).

Member Nodes:

DataONE comprises a distributed network of data centers, science networks or organizations. These organizations can expose their data within the DataONE network through the implementation of the DataONE Member Node service interface. In addition to scientific data, Member Nodes can provide computing resources, or services such as data replication, to the DataONE community.

Investigator Toolkit:

-

Library Library <td< th=""><th></th><th>management/muex.ntm</th><th>h</th><th></th><th>DataONE</th><th></th></td<>		management/muex.ntm	h		DataONE		
LIBRARY HOME RESOURCES SERVICES UNIQUE COLLECTIONS COPYRIGHT ABOUT US GIVE TO THE LIBRARY Data Management Home Data Management Planning Data Management protocols, metadata reporting requirements and best practices, and the requirements and protocols, metadata reporting requirements and best practices, and the requirements and recommendations of national funding bodies and/or professional associations. Data Sharing & Preservation Data Management Planning Tools for Data Management Software Tools Catalog Property Data Management Planning Checklist Data Management Planning Checklist Software Tools Catalog Data Managements Data Documentation & Metadata Work is Metadata? Commonly Used Standards Contact Us Data Documentation & Metadata UI Data-Intensive Projects Northwest Knowledge Network (NKN) Data Documentation Data Documentation & Metadata UI Data-Intensive Projects REACCH PNA Data Sharing & Preservation Data Sharing & Preservation UI Data-Intensive Projects	University of Idaho				ю номе		
Data Management Home Data Management Resources and Services Data Management Planning The University of Idaho Library strives to help researchers at the University of Idaho become familiar with data management practices. Library faculty endeavor to communicate information about data storage standards and protocols, metadata reporting requirements and best practices, and the requirements and recommendations of national funding bodies and/or professional associations. Data Management Planning Preservation Data Management Planning Why Manage Your Data? Funding Agency Guidelines Data Planning Checklist Data Management Plans (DMPs) Tools for Data Management Software Tools Catalog From the DataONE project Data Management Plans (DMPs) UI Repositories Northwest Knowledge Network (NKN) INSIDE Idaho Data Documentation & Metadata What is Metadata? Commonly Used Standards Best Practices UI Data-Intensive Projects REACCH PNA Idaho EPSCoR	LIBRARY HOME RESOL	JRCES SERVICES UNIQUE COLLECTIONS C	OPYRIGHT	ABOUT US	GIVE TO THE LIBRARY		
Data Management Planning Data Documentation & Metadata Data Sharing & Preservation Copyright & Intellectual Property Data Citation Contact Us External Links Acknowledgements Data Documentation What is Metadata What is Metadata? Commonly Used Standards Best Practices Data Sharing & Data Sharing & Data Citation Contact Us External Links Data Documentation & Metadata Data Documentation & Metadata Data Documentation & Metadata Data Anangement Planning Checklist Data Management Plans (DMPs) UI Repositories Northwest Knowledge Network (NKN) INSIDE Idaho UI Data-Intensive Projects REACCH PNA Idaho EPSCOR	Data Management Home	Data Management Resource	es and	Services			
Data Documentation about data storage standards and protocols, metadata reporting requirements and best practices, and the requirements and recommendations of national funding bodies and/or professional associations. Data Sharing & Preservation Data Management Planning Tools for Data Management Copyright & Intellectual Property Data Management Planning Fools for Data Management Data Citation Data Planning Checklist Prom the DataONE project Data Management Plans (DMPs) UI Repositories Northwest Knowledge Network (NKN) INSIDE Idaho What is Metadata? Commonly Used Standards UI Data-Intensive Projects Best Practices REACCH PNA Idaho EPSCOR	Data Management Planning	The University of Idaho Library strives to help researchers at the University of Idaho become familiar with data management practices. Library faculty endeavor to communicate information					
Data Sharing & Preservation Copyright & Intellectual Property Data Citation Contact Us External Links Acknowledgements Data Documentation & Metadata What is Metadata? Commonly Used Standards Best Practices Data Sharing & Preservation	Data Documentation & Metadata	about data storage standards and protocols, metadata reporting requirements and best practices, and the requirements and recommendations of national funding bodies and/or professional associations					
Copyright & Data Management Planning Tools for Data Management Property Why Manage Your Data? Funding Agency Guidelines Software Tools Catalog Data Citation Data Planning Checklist Data Management Plans (DMPs) IN Repositories External Links Data Documentation & Metadata What is Metadata? Ommonly Used Standards Best Practices Data Sharing & Preservation UI Data-Intensive Projects	Data Sharing & Preservation						
Data Citation Funding Agency Guidelines From the DataONE project Contact Us Data Planning Checklist UI Repositories External Links Data Documentation & Metadata What is Metadata? What is Metadata? Commonly Used Standards UI Data-Intensive Projects Best Practices REACCH PNA Idaho EPSCoR	Copyright & Intellectual Property	Data Management Planning Why Manage Your Data?		Tools for D Software T	ata Management ools Catalog		
Contact Us External Links Acknowledgements Data Planning Checklist Data Management Plans (DMPs) UI Repositories Northwest Knowledge Network (NKN) INSIDE Idaho UI Data-Intensive Projects Best Practices REACCH PNA Idaho EPSCoR	Data Citation	Funding Agency Guidelines		From the DataONE pro		oject	
External Links Acknowledgements Data Documentation & Metadata What is Metadata? Commonly Used Standards Best Practices Best Practices Data Sharing & Preservation Data Sha	Contact Us	Data Planning Checklist Data Management Plans (DMPs)		UI Repositories			
Acknowledgements Data Documentation & Metadata What is Metadata? Commonly Used Standards Best Practices Best Practices Data Sharing & Preservation Data Sharing & Preserva	External Links			Northwest	Knowledge Network		
Commonly Used Standards Best Practices UI Data-Intensive Projects REACCH PNA Idaho EPSCoR	Acknowledgements	Data Documentation & Metadata What is Metadata?		INSIDE Ida	iho		
Idaho EPSCoR		Commonly Used Standards Best Practices		UI Data-In REACCH PN	tensive Projects		
	Data Sharing & Preservation		CoR				



Home What's New







REACCH's Mission

To enhance the sustainability of Inland Pacific Northwest (IPNW) cereal production systems under ongoing and projected climate change while contributing to climate change mitigation.

Welcome to REACCH!

The Regional Approaches to Climate Change in Pacific Northwest Agriculture (REACCH PNA) gathers 20 scientists from three Land-Grant Universities in Idaho, Washington and Oregon and the USDA Agricultural Research Service.

Connecting with wheat and barley farmers, and using innovative methods, we will strive to ensure agriculture and grain production will endure future climate change.

Project Mission Statement



High Level DM Vision and

Strategy

Systems and Data Implementation

(project plan)

Objective

team data interactions

Introduction

Introductions

- 1. High level REACCH Data Management (DM) overview
 - Strategy, policy and planning and integration with larger REACCH goals and objectives
- 2. DM Project plan overview
 - Year 2
 - Years 3-5
- 3. Drill down into objective team DM interactions



1. REACCH DM Strategies



February 29, 2012



1. REACCH DM Strategies



- Aligning REACCH DM tech with overall project goals: REACCH DM Plan
- Modularity, sustainability, extensibility





DM Technology

- Has evolved into a free standing discipline
- Is recognized as essential for sustainable projects that involve large amounts of data (e.g. most research projects)
- Is a primary area of emphasis for NSF, USDA, NIH, and many other funding agencies



DM Resources

- Experimental Program to Stimulate Competitive Research
- http://www.idahoepscor.org/
- UCAR University Corporation for Atmospheric Research
- DataOne Data Observation Network for Earth
- NSF/USDA
- LTERS Long Term Ecological Research Networks
- CUAHSI HIS Consortium of Universities for the Advancement of Hydrologic Science, Inc.
- Other CAPS

February 29, 2012





DM supports REACCH goals



- Ensure long term viability of cereal based farming in the Inland PNW
- Identify farming practices that can help reduce agricultural greenhouse gas emissions
- Create DM support systems that enable success of the above statements



February 29, 2012



REACCH Modularity, Flexibility, Extensibility

- Dynamic integration between raw data and objective team science questions
- Use modular tech approaches to interconnect data (XML, XLST)
- Go from science questions to computation to data back to science questions.
- Metadata used to describe not just raw data, but literature, and data tools. Allow metadata harvesting (OAI)
- Expose specified data to internet with other datasets accessible only to researchers and/or stakeholders



Altered from the presentation "The Fourth Paradigm, eResearch and Digital Scholarship" Tony Hey, Microsoft (From Jim Gray's last talk)

February 29, 2012



Pendleton, OR







2. REACCH DM Plan Overview



February 29, 2012



REACCH DM Plan focus

- Defines the strategy and operations for REACCH DM over year2-5
- Describes YR2 build out
- Includes DM Project Plan
- Describe data policies
- Build out of systems and support services

December 14, 2011





2. REACCH DM Plan Overview

- Plan focuses
 on Year 2
- Vision and implementation of systems
- Spiral/iterative development thru YR3-5

EXECUTIVE SUMMARY	2
INTRODUCTION	3
WHAT IS DATA MANAGEMENT?	3
REACCH DATA MANAGEMENT VISION AND STRATEGY.	3
REACCH DATA MANAGEMENT GOALS AND OBJECTIVES	3
REACCH OPERATIONAL DATA MANAGEMENT GUIDELINES	4
INTRODUCTION	4
REACCH EXPECTED DATA INCLUSION	4
DATA RETENTION	5
DATA FORMATS AND DISSEMINATION	5
INPUT FORMATS	5
DATA STORAGE AND PRESERVATION OF ACCESS	5
REACCH DATA MANAGEMENT ORGANIZATIONAL FRAMEWORK	6
INTRODUCTION	6
REACCH DM ORGANIZATIONAL MODEL	6
REACCH DM TEAM STRUCTURE	7
Data Management Members	7
DATA MANAGEMENT ROLES AND RESPONSIBILITIES	10
Roles	10
Transferred Responsibilities	11
Data management and ongoing maintenenace	12
Data Access Standards	13
METADATA POLICY STANDARDS	14
What is metadata?	14
Why is metadata important?	14
Steps for tagging data with meta-information	15



REACCH DM Plan Overview (cont'd)

Systems design and build out

 DM Project plan implementation

REACCH DAT A MANAGEMENT PROPOSED DE SIGN AND ARCHITE CTURE	16
DESIGN OVERVIEW	16
CONCEPTUAL/SEMANTIC DESIGN	16
LOGICAL/PHYSICAL DESIGN	16
DATA MANAGEMENT ORGANIZATIONAL DATA FRAMEWORK	17
REACCH DAT A MANAGEMENT PROJECT IMPLEMENT ATION STRUCTURE	21
DATA MANAGEMENT IMPLEMENTATION APPROACH	21
DM IMPLEMENTATION FUNCTIONAL PROJECT OVERVIEW	21
DM Implementation Functional Project Teams	21
DATA MANAGEMENT IMPLEMENTATION PROJECT PLAN	22
DM Implementation Project Summary Statement	22
DM Implementation Team Makeup and Reporting	22
DM Implementation Quality Assurance and Control Management	22
DM Implementation Issue Management	22
DM Implementation Project Plan Milestones and Deliverables	22
DM Implementation Activation	22
DM Implementation Project Completion	22



REACCH DM Plan Overview(cont'd)

DM
 Sustainability
 Plan
 implementation

Build out for YR3-5

REACCH DATA MANAGEMENT PROJECT MAINTENANCE AND SUSTAINABIL	ITY
STRUCTURE	23
DM SUSTAINABILITY APPROACH	23
DM SUSTAINABILITY FUNCTIONAL PROJECT OVERVIEW	23
DM Sustainability Functional Project Teams	23
DM SUSTAINABILITY PROJECT PLAN	23
DM Sustainability Scope Statement	23
DM Sustainability Team Makeup and Reporting	23
DM Sustainability Quality Control Management	23
DM Sustainability Issue Management	23
DM Sustainability Project Plan Milestones and Deliverables	23



REACCH DM Project Plan – Y2

- REACCH Project Plan is a sub-component of the overall REACCH DM Plan
- Project Plan is built out in Open Workbench
- Tasks and resources assigned
- Managed as a project plan – transferred and updated with Central Desktop

8.1	Cl Preparation					
8.2.1	CI - DM Planning and Pre-Initiation					
8.2.2	🗄 CI - DM Project Intiation					
8.2.3	CI - DM Project Execution					
8.2.3.1	DM Project Execution - System Implementation - development	and production				
8.2.3.2	■ DM Project Execution - System Support Services					
8.2.3.3	■ DM Project Execution - Staging Server Organization - Develop	pment				
8.2.3.4	■ DM Project Execution - Staging Server Organization - Product	tion				
8.2.3.5	■ DM Project Execution - DM Application Server Organization -	Development				
8.2.3.6	E DM Project Execution - Application Server Organization - Proceedings	duction				
8.2.3.7	DM Project Execution - Client Application Organization	Essontial				
8.2.3.8	E DM Project Execution - Web Portal Organization	for				
8.2.3-9	E DM Project Execution - DM Policy Dev and Management	Objective				
8.2.3-10	E DM Project Execution - DM Metadata Dev and Management	teams				
8.2.4	CI - DM Project Control					
8.2.4.1	E DM Project Control - Replication and Redundancy					
8.2.4.2	■ DM Project Control - Project Quality Control and Quality Assur	ance				
8.2.4.3	■ DM Project Control - Reporting					
8.2.5	🗉 Cl - DM Project Closing					
8.3	🖃 CI Sustainability					



REACCH DM Timeline YR2

DM project vision and plan yr2 developed. DM plan yr2 implemented. Objective team tasks assigned. System and data build out begins. Development and production systems developed, processes in place for management. Data interaction and management begins. System build out YR2 completed. Data interaction and management continues refinement thru YR4.

February 29, 2012

REACCH Annual Meeting 2012 Pendleton, OR

2012



REACCH DM Timeline YR3-5



Pendleton, OR



REACCH DM Project Plan – Objective Team Interaction



DM and Objective Team Interaction



- Project plan includes tasks for objective teams
- Each team will have a set of standardized DM tasks
- Will work in conjunction with Team 8 – cyberinfrastructure
- DM tasks that are part of the plan will be assigned to objective team leaders and members

February 29, 2012



- Metadata
 - Can my data fit within the proposed REACCH metadata standards?
 - Declare use of a standard tag your data with metadata within your own research processes
 - Work to understand your data, and begin to metadata tag your information as you upload it to REACCH repositories, using REACCH metadata upload tools
- Policy
 - Work to provide input for data access policy February 29, 2012 REACCH Annual Meeting 2012 Pendleton, OR



Data Processes

- Understanding your data relationships
- Use of a database (if applicable to your data)
- Have a defined data schema i.e. a diagram that shows how your data is associated.
- Understand and document how your scientific research step relate to your data – by organizing your data and instantiating those steps in technology (programming).





Metadata Discussion

- Ecological Metadata Language (EML)
- Federal Geographic Data Commission (FGDC) metadata standards workbook
- Dublin Core
 - NETCDF/climatic conventions
 - Hydrology/CUAHSI WaterML standards
 - GEOMS Generic Earth Observation Metadata Standard
 - Ameriflux



Geoportal Server is likely the software we will use for metadata cataloging and tagging http://www.esri.com/softw are/arcgis/geoportal/liveuser-sites.html



View Metadata Examples



REACCH Soft Content Management

- Central Desktop: streamline and integrate
 e.g. databases vs. file structure for CD content
- Development of educational and extension based portals for public and other secure content
- Metadata connects soft content to hard content





REACCH DM Policy

- REACCH data access policy draft completed
- Data Tier levels:
 - Public
 - Stakeholder
 - Researcher
- Data use and warranties



- User agreement vs. free unstipulated use of data
- Liability, appropriate data use



Summarization of DM Tasks and Resources

- Objective team lead responsible for completion/assignment of related DM tasks for your team. Tasks will be assigned in CD, and will be focused in the areas of:
 - Policy
 - Data processes
 - Metadata
 - Application development



 DM plan, DM project plan, and data access policy (in one document) will be posted to CD as a guide.

March 1, 2012



Summarization of DM Tasks and Resources (cont'd)

- Team 8 will lead efforts to complete DM tasks thru bi-monthly meetings and coordination with objective team meetings.
- DM task completion will be tracked and presented to the bi-monthly leadership for review and tracking.





REACCH DM Discussion

February 29, 2012



Thank you!!

Paul Gessler

Professor of Remote Sensing and Geospatial Ecology

Department of Forest, Rangeland, and Fire Sciences College of Natural Resources, University of Idaho



Erich Seamon, MS PMP GISP Environmental Data Manager REACCH College of Agricultural and Life Sciences University of Idaho http://reacchpna.org 208.885.1230

erichs@uidaho.edu



Potential DM Discussions and Interaction Primer

- **Team 1**: how do other datasets relate to modeling efforts?
- **Team 2**: how can monitoring data feed dynamically into REACCH repositories?
- Team 3: Can we run cropsyst scenarios dynamically automatically – using python?
- **Team 4**: Can the TOA model be run using python to twiggle excel spreadsheets (run climate/cropping/econ model as a whole)?
- **Team 5**: Does Biotic data feed into the cropsyst analysis, or into the climatic modeling (see data flow diagram)?
- **Team 6**: What might an data driven application look like that an educator might use in a classroom?
- **Team 7**: What might an data driven application look like that a stakeholder might use in the field?