Comparative Study: Scientific Research Group Data Management Practices & Local Data Repositories

IMLS DCERC Partners: UIUC, NCAR, UTenn

Karen S. Baker, University of Illinois Urbana-Champaign, karensbaker@gmail.com

Background

· An increasing number of scientific research groups (SRGs) in the earth and environmental sciences are considering ways to provide data access. Data repositories in a variety of sizes and configurations have emerged in the last decades. Preparing data for release to a digital repository requires new kinds of data practices particularly when scientific inquiry involves the heterogeneity of data associated with fieldwork in the natural sciences.

Object of Study

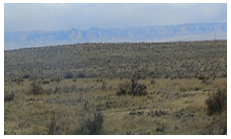
The decisions a research group makes about data management and data sharing may hinder or foster the growth of information infrastructure needed for identification, description, and aggregation of datasets as well as for coordination via systems, standards and partnering. The **object of study** will be decisions made relating to data management including data policy, selection, metadata, and distributions of data related work.

Methods

An ethnographic approach is planned with data for this study collected at the sites, thus taking into account the actual work settings. A participant observation role will be adopted at meetings. Semi-structured interviews and document analysis will be carried out for each case.

The Sites & Their Characteristics

TFSE
Illinois River Ecology
New Site



SGS LTER
Prairie Ecology
Site in Transition

NCAR/EOL
Atmospheric Science
Mature Site

The **unit of analysis** for the study is scientific research groups associated with a data repository. The SRGs represent a convenience sample of differing site-repository configurations and maturity.

Sites	Environmental Field Stations		National Research Center
Related Factors	TFSE Emiquon, Illinois River	SGS LTER Shortgrass Steppe	NCAR/EOL Atmospheric Research
Stage of site development	new	mature site in transition	mature
Scientific focus	ecological and restoration science	ecological science	atmospheric science
Study coordination	research station: site of independent studies in coordination with partners, e.g. INHR and USGS	research station: long-term study with LTER network office as communication hub	data hub and field instrumentation coordination for multiple independent projects
Stage of DM development	new	mature	mature
Stage of data infrastructure	new	in transition	advanced
DM goals	define data management	site-based information system and interoperability	data composites and project-oriented information system and archive
DM evolution	computer scientist and data manager initiating data management	from SGS LTER project-oriented data management team to CSU Libraries IR	data management group supporing local and community projects
Data repository	no	yes	yes
Liaison role	planning ongoing	information manager position embedded in science-driven long-term project and member of Network IM Committee	data management group has full time staff; facility groups have project scientist liaison to DM
Study site	Illinois River and floodplains situated on Nature Conservancy land	prairie biome in Eastern Colorado situated on USDA ARS experimental range land and USFS public land	atmospheric sampling for projects with aircraft & instrument facilities
Configuration	site-based, independent field science projects	site-based collaborative science with embedded IM	project-based partnering for atmospheric field campaigns
Funding	individual researchers have state, university, INHS, USGS, NSF	long-term site-based multi-investigator funding ending	multiple national agencies
Organizational home	University of Illinois Springfield (undergraduate)	Colorado State University	NCAR Earth Observing Laboratory Division
Primary partners	Nature Conservancy, Illinois Natural History Survey	LTER Network; transitioning to Colorado State University Libraries	UCAR, University of Colorado
Date began	2008	1982	1967
Date end	ongoing	2014	ongoing

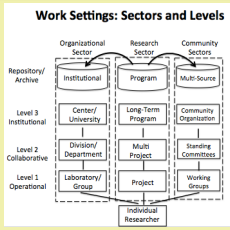
Acknowledgements

This thesis work is supported by the Data Curation Education at Research Centers (DCERC) IMLS Award #RE-02-10-0004-10.

Some Research Questions

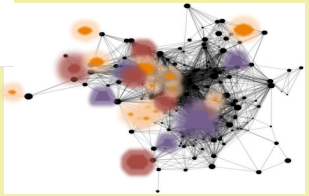
What are the characteristics of scientific data repositories? What is their impact on related SRGs? What are key decisions that scientists make about their data and its management? What are measures of maturity & kinds functionality for data management and repository development?

Repository Relations



Complex mix of organizational levels & sectors at each data repository

Envisioning a web of repositories



Conclusion

The diversity of existing repository configurations is noteworthy. Empirical studies of scientific research groups currently associated with data repositories at various stages of development can contribute to understandings of how to support scientific data practices and can inform development of new repositories.

January 21, 2014