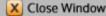
	Control ID	Final ID	Title	Presenting Author
1	1787664		Managing Data and Facilitating Science: A spectrum of activities in the Centre for Environmental Data Archival.	Lawrence, Bryan
2	1801371		Data stewardship – a fundamental part of the scientific method	Foster, Clinton
3	1803399		Roles of Data Stewardship: The people who put the theory into practice.	Parsons, Mark
4	1815458		Enabling Long-Term Earth Science Research: Changing Data Practices	Baker, Karen
5	1819525		Managing Data, Provenance and Chaos through Standardization and Automation at the Georgia Coastal Ecosystems LTER Site	Sheldon, Wade
6	1820346		Metadata Standards in Theory and Practice: The Human in the Loop	Yarmey, Lynn
7	1809122		Supporting Data Stewardship Throughout the Data Life Cycle in the Solid Earth Sciences	Ferrini, Vicki Lynn
8	1818033		Establishing Long Term Data Management Research Priorities via a Data Decadal Survey	Wilson, Anne



Print



Proof

CONTROL ID: 1787664

TITLE: Managing Data and Facilitating Science: A spectrum of activities in the Centre for Environmental Data Archival.

PRESENTATION TYPE: Assigned by Committee (Oral or Poster) [Invited]

CURRENT SECTION/FOCUS GROUP: Earth and Space Science Informatics (IN)

CURRENT SESSION: IN011. Data Stewardship: in Theory and in Practice

AUTHORS (FIRST NAME, LAST NAME): Bryan Lawrence^{1, 2}, Victoria Bennett², Sarah Callaghan², Martin N Juckes², Sam Pepler²

INSTITUTIONS (ALL): 1. NCAS and Meteorology, University of Reading, Reading, United Kingdom.

2. Centre for Environmental Data Archival, STFC, Didcot, United Kingdom.

PRESENTER: Bryan Lawrence

PRESENTER (E-MAIL ONLY): b.n.lawrence@reading.ac.uk

ABSTRACT BODY: The UK Centre for Environmental Data Archival (CEDA) hosts a number of formal data centres, including the British Atmospheric Data Centre (BADC), and is a partner in a range of national and international data federations, including the InfraStructure for the European Network for Earth system Simulation, the Earth System Grid Federation, and the distributed IPCC Data Distribution Centres. The mission of CEDA is to formally curate data from, and facilitate the doing of, environmental science.

The twin aims are symbiotic: data curation helps facilitate science, and facilitating science helps with data curation. Here we cover how CEDA delivers this strategy by established internal processes supplemented by short-term projects, supported by staff with a range of roles. We show how CEDA adds value to data in the curated archive, and how it supports science, and show examples of the aforementioned symbiosis.

We begin by discussing curation: CEDA has the formal responsibility for curating the data products of atmospheric science and earth observation research funded by the UK Natural Environment Research Council (NERC). However, curation is not just about the provider community, the consumer communities matter too, and the consumers of these data cross the boundaries of science, including engineers, medics, as well as the gamut of the environmental sciences. There is a small, and growing cohort of non-science users. For both producers and consumers of data, information about data is crucial, and a range of CEDA staff have long worked on tools and techniques for creating, managing, and delivering metadata (as well as data). CEDA "science support" staff work with scientists to help them prepare and document data for curation.

As one of a spectrum of activities, CEDA has worked on data Publication as a method of both adding value to some data, and rewarding the effort put into the production of quality datasets. As such, we see this activity as both a curation and a facilitation activity.

A range of more focused facilitation activities are carried out, from providing a computing platform suitable for big-data analytics (the Joint Analysis System, JASMIN), to working on distributed data analysis (EXARCH), and the acquisition of third party data to support science and impact (e.g. in the context of the facility for Climate and Environmental Monitoring from Space, CEMS).

We conclude by confronting the view of Parsons and Fox (2013) that metaphors such as Data Publication, Big Iron, Science Support etc are limiting, and suggest the CEDA experience is that these sorts of activities can

and do co-exist, much as they conclude they should. However, we also believe that within co-existing metaphors, production systems need to be limited in their scope, even if they are on a road to a more joined up infrastructure. We shouldn't confuse what we can do now with what we might want to do in the future.

INDEX TERMS: 1900 INFORMATICS.

(No Table Selected) (No Image Selected)

Additional Details

Previously Presented Material:

Scheduling Request:

TITLE OF TEAM: CEDA Team

ScholarOne Abstracts® (patent #7,257,767 and #7,263,655). © ScholarOne, Inc., 2013. All Rights Reserved. ScholarOne Abstracts and ScholarOne are registered trademarks of ScholarOne, Inc.



Follow ScholarOne on Twitter

Terms and Conditions of Use





Proof

CONTROL ID: 1801371

TITLE: Data stewardship – a fundamental part of the scientific method **PRESENTATION TYPE:** Assigned by Committee (Oral or Poster) [Invited]

CURRENT SECTION/FOCUS GROUP: Earth and Space Science Informatics (IN)

CURRENT SESSION: IN011. Data Stewardship: in Theory and in Practice

AUTHORS (FIRST NAME, LAST NAME): Clinton Foster¹, Jonathon Ross¹, Lesley A Wyborn¹

INSTITUTIONS (ALL): 1. Geoscience Australia, Canberra, ACT, Australia.

PRESENTER: Clinton Foster

PRESENTER (E-MAIL ONLY): clinton.foster@ga.gov.au

ABSTRACT BODY: This paper emphasises the importance of data stewardship as a fundamental part of the scientific method, and the need to effect cultural change to ensure engagement by earth scientists. It is differentiated from the science of data stewardship per se.

Earth System science generates vast quantities of data, and in the past, data analysis has been constrained by compute power, such that sub-sampling of data often provided the only way to reach an outcome. This is analogous to Kahneman's System 1 heuristic, with its simplistic and often erroneous outcomes.

The development of HPC has liberated earth sciences such that the complexity and heterogeneity of natural systems can be utilised in modelling at any scale, global, or regional, or local; for example, movement of crustal fluids. Paradoxically, now that compute power is available, it is the stewardship of the data that is presenting the main challenges. There is a wide spectrum of issues: from effectively handling and accessing acquired data volumes [e.g. satellite feeds per day/hour]; through agreed taxonomy to effect machine to machine analyses; to idiosyncratic approaches by individual scientists. Except for the latter, most agree that data stewardship is essential. Indeed it is an essential part of the science workflow.

As science struggles to engage and inform on issues of community importance, such as shale gas and fraccing, all parties must have equal access to data used for decision making; without that, there will be no social licence to operate or indeed access to additional science funding (Heidorn, 2008).

The stewardship of scientific data is an essential part of the science process; but often it is regarded, wrongly, as entirely in the domain of data custodians or stewards.

Geoscience Australia has developed a set of six principles that apply to all science activities within the agency:

Relevance to Government

Collaborative science

Quality science

Transparent science

Communicated science

Sustained science capability

Every principle includes data stewardship: this is to effect cultural change at both collective and individual levels to ensure that our science outcomes and technical advice are effective for the Government and community.

INDEX TERMS: 1912 INFORMATICS Data management, preservation, rescue.

(No Table Selected)

(No Image Selected)

Additional Details

Previously Presented Material: Nil

Scheduling Request:

TITLE OF TEAM:

ScholarOne Abstracts® (patent #7,257,767 and #7,263,655). © ScholarOne, Inc., 2013. All Rights Reserved. ScholarOne Abstracts and ScholarOne are registered trademarks of ScholarOne, Inc.



Follow ScholarOne on Twitter

Terms and Conditions of Use





Proof

CONTROL ID: 1803399

TITLE: Roles of Data Stewardship: The people who put the theory into practice. **PRESENTATION TYPE:** Assigned by Committee (Oral or Poster) [Invited]

CURRENT SECTION/FOCUS GROUP: Earth and Space Science Informatics (IN)

CURRENT SESSION: IN011. Data Stewardship: in Theory and in Practice

AUTHORS (FIRST NAME, LAST NAME): Mark A Parsons¹

INSTITUTIONS (ALL): 1. Rensselaer Polytechnic Institute, Boulder, CO, United States.

PRESENTER: Mark Parsons

PRESENTER (E-MAIL ONLY): parsons.mark@gmail.com

ABSTRACT BODY: To understand something very complex like the Earth system it is helpful to have a model that sketches the major components of the system and their interactions. This helps us to understand the relative importance of those components and their behavior. The same is true, when considering a complex social enterprise such as data stewardship. It is helpful to have a model or a metaphor that allows us to conceive of the entire enterprise, the key players, and their interactions.

In the spirit that "all models are wrong, but some are useful," this presentation explores several models of data stewardship and how different conceptions of the enterprise define different roles for the participants—be they researchers or data practitioners. The goal is to illustrate how consideration of multiple theories or models of data stewardship can help us better define the actual practice necessary to ensure well-preserved and useful data.

INDEX TERMS: 1900 INFORMATICS.

(No Table Selected)
(No Image Selected)

Additional Details

Previously Presented Material: 0

Scheduling Request: TITLE OF TEAM:

ScholarOne Abstracts® (patent #7,257,767 and #7,263,655). © ScholarOne, Inc., 2013. All Rights Reserved. ScholarOne Abstracts and ScholarOne are registered trademarks of ScholarOne, Inc.

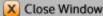


Follow ScholarOne on Twitter

Terms and Conditions of Use







Proof

CONTROL ID: 1815458

TITLE: Enabling Long-Term Earth Science Research: Changing Data Practices PRESENTATION TYPE: Assigned by Committee (Oral or Poster) [Invited]

CURRENT SECTION/FOCUS GROUP: Earth and Space Science Informatics (IN)

CURRENT SESSION: IN011. Data Stewardship: in Theory and in Practice

AUTHORS (FIRST NAME, LAST NAME): Karen S. Baker¹

INSTITUTIONS (ALL): 1. Center for Informatics Research in Science and Scholarship, Graduate School of Library and Information Science, University of Illinois at Urbana Champaign, Champaign, IL, United States.

PRESENTER: Karen Baker

PRESENTER (E-MAIL ONLY): ksbaker2@illinois.edu

ABSTRACT BODY: Data stewardship plans are shaped by our shared experiences. As a result, community engagement and collaborative activities are central to the stewardship of data. Since modes and mechanisms of engagement have changed, we benefit from asking anew: 'Who are the communities?' and 'What are the lessons learned?'. Data stewardship with its long-term care perspective, is enriched by reflection on community experience. This presentation draws on data management issues and strategies originating from within long-term research communities as well as on recent studies informed by library and information science. Ethnographic case studies that capture project activities and histories are presented as resources for comparative analysis.

Agency requirements and funding opportunities are stimulating collaborative endeavors focused on data reuse and archiving. Research groups including earth scientists, information professionals, and data systems designers are recognizing the possibilities for new ways of thinking about data in the digital arena. Together, these groups are re-conceptualizing and reconfiguring for data management and data curation. A differentiation between managing data for local use and production of data for re-use remotely in locations and fields remote from the data origin is just one example of the concepts emerging to facilitate development of data management. While earth scientists as data generators have the responsibility to plan new workflows and documentation practices, data and information specialists have responsibility to promote best practices as well as to facilitate the development of community resources such as controlled vocabularies and data dictionaries. With data-centric activities and changing data practices, the potential for creating dynamic community information environments in conjunction with development of data facilities exists but remains elusive.

INDEX TERMS: 1912 INFORMATICS Data management, preservation, rescue, 0434 BIOGEOSCIENCES Data sets.

(No Table Selected) (No Image Selected)

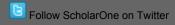
Additional Details

Previously Presented Material:

Scheduling Request:

TITLE OF TEAM:

ScholarOne Abstracts® (patent #7,257,767 and #7,263,655). © <u>ScholarOne</u>, Inc., 2013. All Rights Reserved. ScholarOne Abstracts and ScholarOne are registered trademarks of ScholarOne, Inc.



Terms and Conditions of Use





Proof

CONTROL ID: 1819525

TITLE: Managing Data, Provenance and Chaos through Standardization and Automation at the Georgia

Coastal Ecosystems LTER Site

PRESENTATION TYPE: Assigned by Committee (Oral or Poster)

CURRENT SECTION/FOCUS GROUP: Earth and Space Science Informatics (IN)

CURRENT SESSION: IN011. Data Stewardship: in Theory and in Practice

AUTHORS (FIRST NAME, LAST NAME): Wade Sheldon¹

INSTITUTIONS (ALL): 1. Marine Sciences, University of Georgia, Athens, GA, United States.

PRESENTER: Wade Sheldon

PRESENTER (E-MAIL ONLY): sheldon@uga.edu

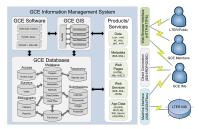
ABSTRACT BODY: Managing data for a large, multidisciplinary research program such as a Long Term Ecological Research (LTER) site is a significant challenge, but also presents unique opportunities for data stewardship. LTER research is conducted within multiple organizational frameworks (i.e. a specific LTER site as well as the broader LTER network), and addresses both specific goals defined in an NSF proposal as well as broader goals of the network; therefore, every LTER data can be linked to rich contextual information to guide interpretation and comparison. The challenge is how to link the data to this wealth of contextual metadata.

At the Georgia Coastal Ecosystems LTER we developed an integrated information management system (GCE-IMS) to manage, archive and distribute data, metadata and other research products as well as manage project logistics, administration and governance (figure 1). This system allows us to store all project information in one place, and provide dynamic links through web applications and services to ensure content is always up to date on the web as well as in data set metadata. The database model supports tracking changes over time in personnel roles, projects and governance decisions, allowing these databases to serve as canonical sources of project history. Storing project information in a central database has also allowed us to standardize both the formatting and content of critical project information, including personnel names, roles, keywords, place names, attribute names, units, and instrumentation, providing consistency and improving data and metadata comparability. Lookup services for these standard terms also simplify data entry in web and database interfaces.

We have also coupled the GCE-IMS to our MATLAB- and Python-based data processing tools (i.e. through database connections) to automate metadata generation and packaging of tabular and GIS data products for distribution. Data processing history is automatically tracked throughout the data lifecycle, from initial import through quality control, revision and integration by our data processing system (GCE Data Toolbox for MATLAB), and included in metadata for versioned data products. This high level of automation and system integration has proven very effective in managing the chaos and scalability of our information management program.

INDEX TERMS: 1912 INFORMATICS Data management, preservation, rescue, 1976 INFORMATICS Software tools and services, 1904 INFORMATICS Community standards, 1948 INFORMATICS Metadata: Provenance.

(No Table Selected)



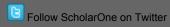
Additional Details

Previously Presented Material:

Scheduling Request:

TITLE OF TEAM:

ScholarOne Abstracts® (patent #7,257,767 and #7,263,655). © ScholarOne, Inc., 2013. All Rights Reserved. ScholarOne Abstracts and ScholarOne are registered trademarks of ScholarOne, Inc.



Terms and Conditions of Use



Proof

CONTROL ID: 1820346

TITLE: Metadata Standards in Theory and Practice: The Human in the Loop

PRESENTATION TYPE: Assigned by Committee (Oral or Poster)

CURRENT SECTION/FOCUS GROUP: Earth and Space Science Informatics (IN)

CURRENT SESSION: IN011. Data Stewardship: in Theory and in Practice

AUTHORS (FIRST NAME, LAST NAME): Lynn Yarmey¹, Sandy Starkweather²

INSTITUTIONS (ALL): 1. National Snow and Ice Data Center, University of Colorado Boulder, Boulder, CO, United States.

2. IASOA/Arctic Observing Implementation, NOAA Earth System Research Laboratory, Boulder, CO, United States.

PRESENTER: Lynn Yarmey

PRESENTER (E-MAIL ONLY): lynn.yarmey@colorado.edu

ABSTRACT BODY: Metadata standards are meant to enable interoperability through common, well-defined structures and are a foundation for broader cyberinfrastructure efforts. Standards are central to emerging technologies such as metadata brokering tools supporting distributed data search. However, metadata standards in practice are often poor indicators of standardized, readily interoperable metadata.

The International Arctic Systems for Observing the Atmosphere (IASOA) data portal provides discovery and access tools for aggregated datasets from ten long-term international Arctic atmospheric observing stations. The Advanced Cooperative Arctic Data and Information Service (ACADIS) Arctic Data Explorer brokers metadata to provide distributed data search across Arctic repositories. Both the IASOA data portal and the Arctic Data Explorer rely on metadata and metadata standards to support value-add services. Challenges have included: translating between different standards despite existing crosswalks, diverging implementation practices of the same standard across communities, changing metadata practices over time and associated backwards compatibility, reconciling metadata created by data providers with standards, lack of community-accepted definitions for key terms (e.g. 'project'), integrating controlled vocabularies, and others. Metadata record 'validity' or compliance with a standard has been insufficient for interoperability. To overcome these challenges, both projects committed significant work to integrate and offer services over already 'standards compliant' metadata. Both efforts have shown that the 'human-in-the-loop' is still required to fulfill the lofty theoretical promises of metadata standards.

In this talk, we 1) summarize the real-world experiences of two data discovery portals working with metadata in standard form, and 2) offer lessons learned for others who work with and rely on metadata and metadata standards.

INDEX TERMS: 1946 INFORMATICS Metadata, 1982 INFORMATICS Standards, 1960 INFORMATICS Portals and user interfaces, 1916 INFORMATICS Data and information discovery.

(No Table Selected)
(No Image Selected)

Additional Details

Previously Presented Material: Scheduling Request: TITLE OF TEAM:

ScholarOne Abstracts® (patent #7,257,767 and #7,263,655). © ScholarOne, Inc., 2013. All Rights Reserved. ScholarOne Abstracts and ScholarOne are registered trademarks of ScholarOne, Inc.

Follow ScholarOne on Twitter

Terms and Conditions of Use



Proof

CONTROL ID: 1809122

TITLE: Supporting Data Stewardship Throughout the Data Life Cycle in the Solid Earth Sciences

PRESENTATION TYPE: Assigned by Committee (Oral or Poster)

CURRENT SECTION/FOCUS GROUP: Earth and Space Science Informatics (IN)

CURRENT SESSION: IN011. Data Stewardship: in Theory and in Practice

AUTHORS (FIRST NAME, LAST NAME): Vicki Lynn Ferrini¹, Kerstin A Lehnert¹, Suzanne M Carbotte¹,

Leslie Hsu¹

INSTITUTIONS (ALL): 1. Lamont-Doherty Earth Observatory, Columbia University, Palisades, NY, United

States.

PRESENTER: Vicki Lynn Ferrini

PRESENTER (E-MAIL ONLY): ferrini@ldeo.columbia.edu

ABSTRACT BODY: Stewardship of scientific data is fundamental to enabling new data-driven research, and ensures preservation, accessibility, and quality of the data, yet researchers, especially in disciplines that typically generate and use small, but complex, heterogeneous, and unstructured datasets are challenged to fulfill increasing demands of properly managing their data. The IEDA Data Facility (www.iedadata.org) provides tools and services that support data stewardship throughout the full life cycle of observational data in the solid earth sciences, with a focus on the data management needs of individual researchers. IEDA builds upon and brings together over a decade of development and experiences of its component data systems, the Marine Geoscience Data System (MGDS, www.marine-geo.org) and EarthChem (www.earthchem.org). IEDA services include domain-focused data curation and synthesis, tools for data discovery, access, visualization and analysis, as well as investigator support services that include tools for data contribution, data publication services, and data compliance support. IEDA data synthesis efforts (e.g. PetDB and Global Multi-Resolution Topography (GMRT) Synthesis) focus on data integration and analysis while emphasizing provenance and attribution. IEDA's domain-focused data catalogs (e.g. MGDS and EarthChem Library) provide access to metadata-rich long-tail data complemented by extensive metadata including attribution information and links to related publications. IEDA's visualization and analysis tools (e.g. GeoMapApp) broaden access to earth science data for domain specialist and non-specialists alike, facilitating both interdisciplinary research and education and outreach efforts.

As a disciplinary data repository, a key role IEDA plays is to coordinate with its user community and to bridge the requirements and standards for data curation with both the evolving needs of its science community and emerging technologies. Development of IEDA tools and services is based first and foremost on the scientific needs of its user community. As data stewardship becomes a more integral component of the scientific workflow, IEDA investigator support services (e.g. Data Management Plan Tool and Data Compliance Reporting Tool) continue to evolve with the goal of lessening the 'burden' of data management for individual investigators by increasing awareness and facilitating the adoption of data management practices. We will highlight a variety of IEDA system components that support investigators throughout the data life cycle, and will discuss lessons learned and future directions.

http://www.iedadata.org

INDEX TERMS: 1912 INFORMATICS Data management, preservation, rescue, 1960 INFORMATICS Portals and user interfaces.

(No Table Selected) (No Image Selected)

Additional Details

Previously Presented Material:

Scheduling Request:

TITLE OF TEAM: IEDA Team

ScholarOne Abstracts® (patent #7,257,767 and #7,263,655). © ScholarOne, Inc., 2013. All Rights Reserved. ScholarOne Abstracts and ScholarOne are registered trademarks of ScholarOne, Inc.



Follow ScholarOne on Twitter

Terms and Conditions of Use





Proof

CONTROL ID: 1818033

TITLE: Establishing Long Term Data Management Research Priorities via a Data Decadal Survey

PRESENTATION TYPE: Assigned by Committee (Oral or Poster)

CURRENT SECTION/FOCUS GROUP: Earth and Space Science Informatics (IN)

CURRENT SESSION: IN011. Data Stewardship: in Theory and in Practice

AUTHORS (FIRST NAME, LAST NAME): Anne Wilson^{1, 2}, Paul Uhlir³, Carol B Meyer², Erin Robinson²

INSTITUTIONS (ALL): 1. CU Boulder, Boulder, CO, United States.

2. Foundation for Earth Science, Raleigh, NC, United States.

3. Board on Research Data and Information, National Research Council, National Academy of Sciences, Washington, DC, DC, United States.

PRESENTER: Anne Wilson

PRESENTER (E-MAIL ONLY): anne.wilson@lasp.colorado.edu

ABSTRACT BODY: We live in a time of unprecedented collection of and access to scientific data. Improvements in sensor technologies and modeling capabilities are constantly producing new data sources. Data sets are being used for unexpected purposes far from their point of origin, as research spans projects, discipline domains, and temporal and geographic boundaries. The nature of science is evolving, with more open science, open publications, and changes to the nature of peer review and data "publication". Data-intensive, or computational science, has been identified as a new research paradigm. There is recognition that the creation of a data set can be a contribution to science deserving of recognition comparable to other scientific publications. Federally funded projects are generally expected to make their data open and accessible to everyone.

In this dynamic environment, scientific progress is ever more dependent on good data management practices and policies. Yet current data management and stewardship practices are insufficient. Data sets created at great, and often public, expense are at risk of being lost for technological or organizational reasons. Insufficient documentation and understanding of data can mean that the data are used incorrectly or not at all. Scientific results are being scrutinized and questioned, and occasionally retracted due to problems in data management. The volume of data is greatly increasing while funding for data management is meager and generally must be found within existing budgets.

Many federal government agencies, including NASA, USGS, NOAA and NSF are already making efforts to address data management issues. Executive memos and directives give substantial impetus to those efforts, such as the May 9 Executive Order directing agencies to implement Open Data Policy requirements and regularly report their progress. However, these distributed efforts risk duplicating effort, lack a unifying, long-term strategic vision, and too often work in competition with other priorities of the research enterprise.

This presentation will introduce the Data Decadal Survey, an initial concept created in collaboration between the Federation of Earth Science Information Partners (ESIP) and of the National Research Council's Board on Research Data and Information (BRDI). Consistent with Executive open data policies, the Survey will provide a coordinating platform to address overarching issues and identify research needs and funding priorities in scientific data management and stewardship for the long term. The Survey would address at the broadest level gaps in data management knowledge and practices that hold back scientific progress, and recommend a strategy to address them. The goal is to provide a long term strategic vision that will

• increase the meaningful availability of higher quality data,

- create or improve tools, processes, and practices that support accreditation, accountability, traceability, and reproducibility.
- redirect resources previously required by scientists for data discovery, acquisition, and reformatting to performing actual science.

thereby ultimately enhancing scientific knowledge.

INDEX TERMS: 1999 INFORMATICS General or miscellaneous, 6349 POLICY SCIENCES General or miscellaneous, 6620 PUBLIC ISSUES Science policy.

(No Table Selected) (No Image Selected)

Additional Details

Previously Presented Material: Portions of this work have been discussed at the Winter 2013 and Summer 2013 ESIP meetings.

Scheduling Request:

TITLE OF TEAM: ESIP Data Decadal Survey Cluster

ScholarOne Abstracts® (patent #7,257,767 and #7,263,655). © <u>ScholarOne</u>, Inc., 2013. All Rights Reserved. ScholarOne Abstracts and ScholarOne are registered trademarks of ScholarOne, Inc.



Follow ScholarOne on Twitter

Terms and Conditions of Use