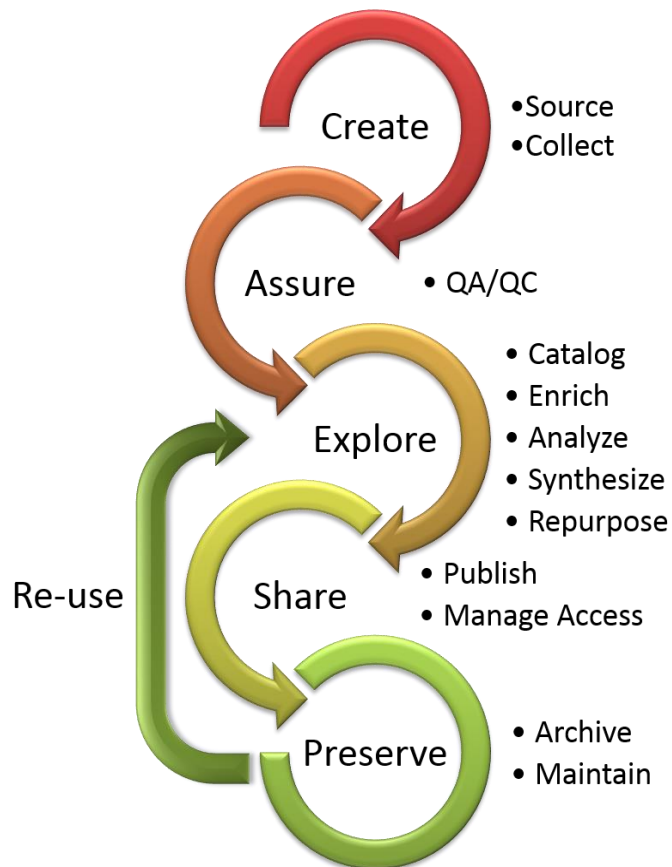# COLLABORATIVE STRATEGIES FOR SUSTAINED ENVIRONMENTAL DATA MANAGEMENT

11/12/2015	Workshop Briefing

This briefing document sets the scene for the upcoming workshop. We hope you will find this information useful.

INTRODUCTION

This Initiative

Many environmental data repositories were initiated to fulfill specific needs or objectives, i.e. archiving and disseminating data from a project, network of research sites, institution, funding source, to accompany paper publications, or more recently, as data papers. This initiative was funded with the goal of exploring how we might develop this network of repositories in a way that will produce new collaboration and curation strategies that also cater to the currently underserved single investigators and move environmental data from 'available' to 'usable', in order to accelerate scientific inquiry.

With this goal in mind we are bringing together data curators from a range of environmental research fields, data aggregators, tool developers, computer scientists and environmental scientists (both data providers and users) for an informed dialog which draws on our collective experience managing data and repositories.

Workshop Organizers

Corinna Gries	University of Wisconsin
Margaret O'Brien	University of California, Santa Barbara
Philip Tarrant	Arizona State University
Ann Chiu (research and support)	University of Wisconsin
Glyn Thomas (facilitator)	Avatar Inc.

Workshop Attendees

Attendee	Organization
Kevin Browne	UC Natural Reserve System
Ken Casey	NCEI (merged NODC + NGDC + NCDC)
Erin Clary	DRYAD MRC
Cindy Chandler	BCO DMO, Woods Hole Oceanographic Institute
Steven Daley-Laursen	Northwest Climate Science Center, Northwest Knowledge Network
Robert Downs	CIESIN
Ruth Duerr	ELOKA, NSIDC
Paul Gessler	LTAR- UIdaho
Paul Hanson	GLEON, LTER
Margaret Hedstrom	SEAD

Nancy Hoebelheinrich	RPI, tetherless world
Rick Hooper	CUAHSI
Greg Janee	CDL/ UCSB
Matt Jones	KNB, DataONE
Inna Kouper	Data2Insight, SEAD
Sherry Lake	University of Virginia Library
Kerstin Lehnert	IEDA
Peter McCartney	National Science Foundation
Carolina Murcia	Org. of Tropical Studies, University of Florida
Cynthia Parr	USDA LTAR
Erin Robinson	ESIP Federation
Dave Rugg	Research Data Archive, USFS
Mark Servilla	LTER Network Information Systems
Wade Sheldon	LTER
Dave Vieglais	DataONE
Anne Wilson	Laboratory of Atmospheric and Space Physics (LASP)
Yaxing Wei	Oak Ridge National Laboratory DAAC

Workshop Goals

To build support for sustained collaboration and coordination in the area of environmental data management that will benefit both research networks and also individual investigators.

The key outcomes are:

- A drafted vision for increased integration in environmental research data management
- A prototype process defining the format of collaboration (e.g. ESIP working groups, RCN, formal consortium) efforts
- Prototyped funding concepts for sustainable data management, i.e. how do we move beyond the grant funded model
- An actionable project to examine how environmental information management can expand to include smaller projects, which are currently not well served
- A coalition willing to address these areas

Workshop Agenda

Agenda	
11/17/15	
8:30 AM	Workshop commences, Introductions
9:00 AM	Defining the Issues & Opportunities
10:00 AM	Break
10:20 AM	Defining the Issues & Opportunities – cont.
11:15 AM	Reflecting on the Issues & Opportunities
12:15 PM	Lunch
1:30 PM	The Future of Environmental Data Repositories
3:30 PM	Break
3:45 PM	Participant Presentations
5:00 PM	Review
5:15 PM	Adjourn
11/18/15	
8:30 AM	Presentations
9:15 AM	Refined Presentations and resource prioritization exercise
10:00 AM	Break
10:20 AM	Refined Presentations and resource prioritization – cont.
11:30 AM	Lunch
1:00 PM	Participant Presentations, version II
3:00 PM	Break
3:15 PM	Where Do We Start?
5:00 PM	Review
5:15 PM	Adjourn
11/19/15	
8:30 AM	Collaboration Options
10:00 AM	Break
10:20 AM	Summarizing Outcomes & Potential next steps
12:00 Noon	Adjourn

CURRENT SITUATION

Recent Developments

In 2011 new guidelines for data management plans were issued by federal funders. These new requirements were intended to ensure that data and other products (models, software, etc.) produced by federally funded projects were effectively preserved and made publicly available.

The National Science Foundation now asks panelists to comment on the inadequacies of data management plans, specifically in the areas of data access and metadata. DEB no longer considers some repositories to be appropriate because of low standards for submission. Panelists also comment on whether or not ALL data will be available. Proposers are expected to correct any noted deficiencies.

Many journal publishers are now making data availability, either through their journal or be other means, a pre-requisite for publication. Consequently, investigator behavior is changing. Whereas it was a challenge to get investigators to supply their data in a timely manner, it is now the investigator who is keen to get the data on-line in order to support publication.

Technology Trends

In a December 2014 commentary, Information Week highlighted six "IT trends to watch in 2015":

- Continuing cloud adoption with an increase in hybrid models
- Increased automation
- More complex supplier environments
- Mobile pervasiveness
- Need for "big data" people
- Unicorn chasing "why can't we match Google?"

All of these areas have the potential to impact (either positively or negatively) environmental data management operations.

Data Trends

In their latest Digital Universe report, EMC[2] highlight that storage volume is increasing by 40% year on year. They estimated that the worldwide storage volume in 2013 was **4.4 zettabytes** (trillion gigabytes). They also estimate that this will increase to **44 zettabytes** by 2020. Current estimates suggest that of the "useful data", less than 1% of these data are actually analyzed and that this proportion will only increase to ~3% by 2020. We should hope that some of these "actually analyzed" data will be ours!

Storage Costs

David Floyer, Wikibon contributor, in an excellent commentary on the emergence of flash storage arrays, predicted that flash storage costs (per TB) would drop below traditional disk storage by early 2016. He also predicts that by 2020 the price ratio (HDD/Flash) will be 732%, or $9 versus $74, 4 year cost/TB, see Figure 1.
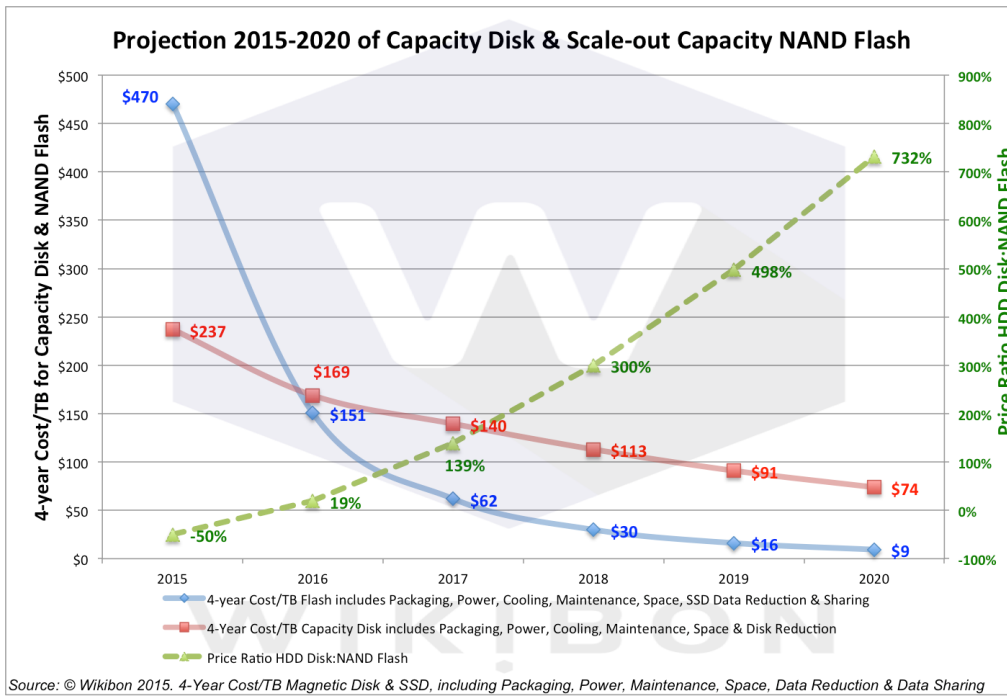
Projection 2015-2020 of Capacity Disk & Scale-out Capacity NAND Flash

Source: © Wikibon 2015. 4-Year Cost/TB Magnetic Disk & SSD, including Packaging, Power, Maintenance, Space, Data Reduction & Data Sharing

FIGURE 1 – PROJECTION 2015-2020 OF CAPACITY DISK AND SCALE-OUT CAPACITY NAND FLASH

SUMMARY OF SURVEY RESULTS

Common Technologies

Within the wide range of technologies used in data management, a possible trend is that repositories are using more open source technologies over commercial technologies, particularly in the area of database technologies, see Figure 2. Other technologies with increasing adoption are Python and R. These products are useful for data processing and analysis. Workflow automation is also expected to increase.
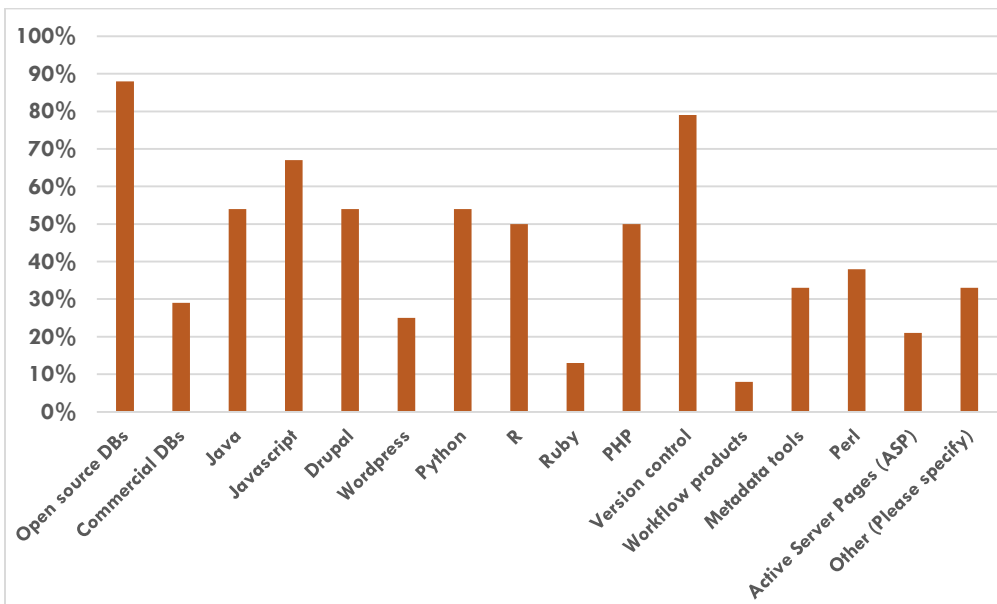

FIGURE 2 – TECHNOLOGIES IN USE AT WORKSHOP PARTICIPANT FACILITIES

Common Metadata Standards

Our survey results indicate that there are a large variety of metadata standards in play, perhaps indicating that specific data types benefit from customized metadata. More than half of respondents indicate that they accepted International Standards Organization (ISO) metadata, while a large number of repositories (71%) use Content Standard for Digital Geospatial Metadata (CSDSM) for spatial data. The wide variety of standards accepted included:

- International Standards Organization (ISO)
- Dublin Core
- Darwin Core
- Ecological Metadata Language (EML)
- Content Standard for Digital Geospatial Metadata (CSDSM) aka FGDC
- Directory Interchange Format (DIF)
- DataCite
- Biological Data Profile (for FGDC)
- OGC standards, WaterML
- NASA Unified Metadata Model (UMM). Bridges to newly adopted ISO standards.
- PRISM. Researchers/authors are prompted to describe their data/publication using mainly Dublin Core elements.
- Project Open Data 1.1
- CF, COARDS

Where a preference was expressed ISO, EML and CSDSM were noted. Some respondents highlighted that their processes were flexible enough to accommodate most standards although some existing standards were considered as inadequate.
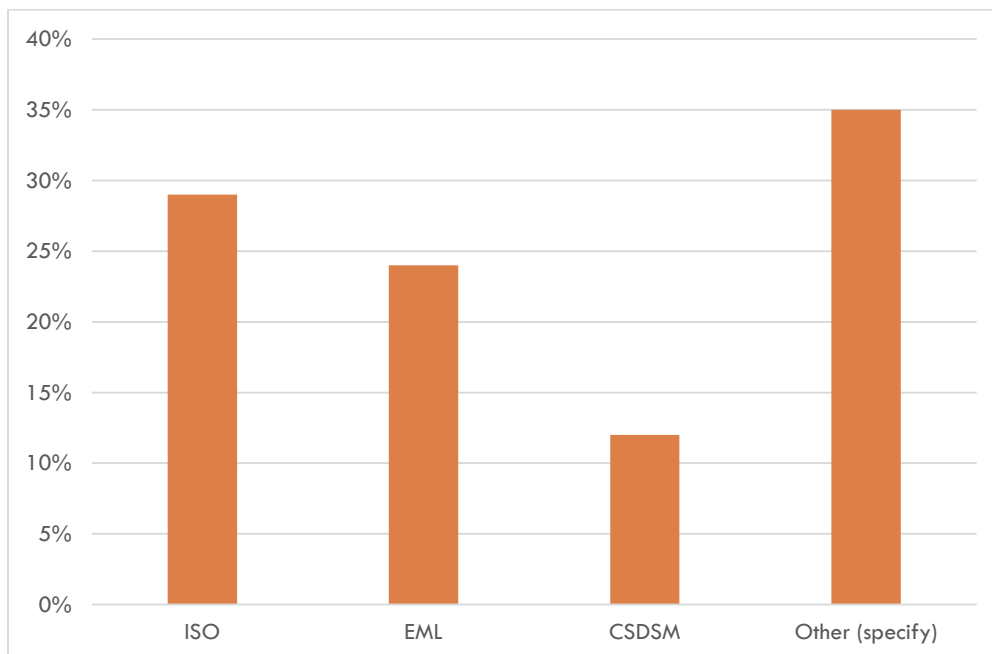

FIGURE 3 – PREFERRED METADATA STANDARDS (WHEN A PREFERENCE WAS EXPRESSED) AT WORKSHOP PARTICIPANT FACILITIES

Operational Considerations

Responses suggest that both budgets and workforce numbers are stable with neither expected to increase or decrease significantly. However, a clear concern expressed by respondents was that there is currently (48%) and will be (57%) insufficient funding to support all the activities considered necessary to support the operation.
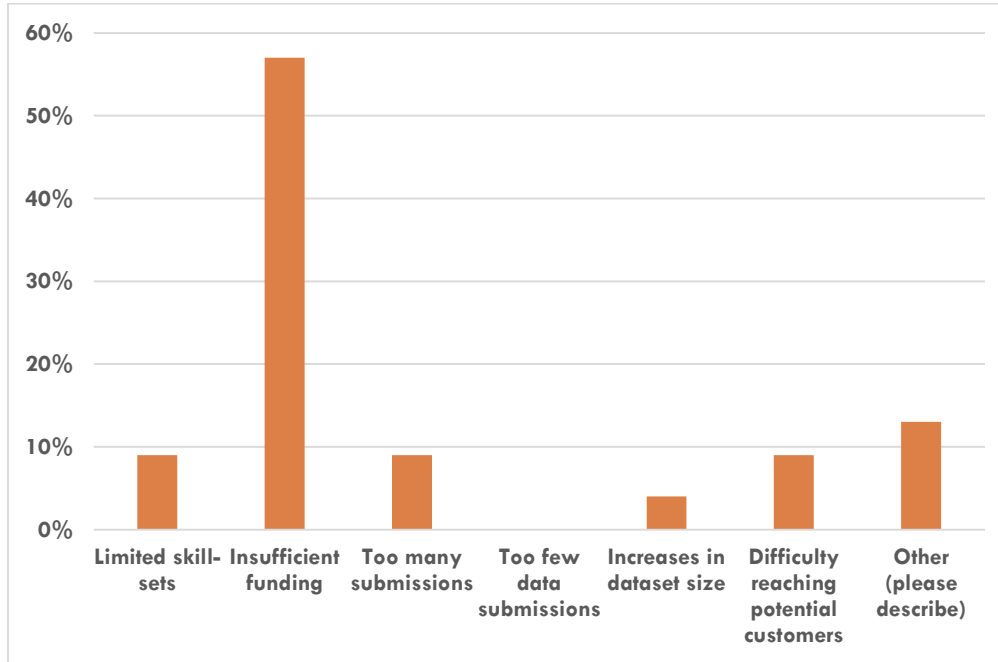

FIGURE 4 – OPERATIONAL CONCERNS EXPRESSED BY WORKSHOP PARTICIPANTS

BACKGROUND INFORMATION

Arizona State University

Arizona State University is a large, public research university located on five campuses in Arizona and on-line. This workshop will be held in the Cochise room on the 2nd floor of the Memorial Union on the Tempe campus. The campus is just five miles from Phoenix Sky Harbor Airport. The meeting starts at 8:30 on day one (17th) with registration and introductions.

Accommodation

Accommodation is provided at the Hotel 1333 close by the main Tempe campus. The address is:

1333 Rural Rd., Tempe, AZ, 85281

Tel: 1-480-968-3451

There is a complimentary airport shuttle between Sky Harbor International Airport and the hotel for workshop participants. http://www.marriott.com/hotels/maps/travel/phxox-hotel-1333 . Shuttle drivers are not allowed to wait at the airport so the best way to get a shuttle to the hotel from the airport: pick up luggage from baggage claim and then call the hotel for pick up. 480-968-3451. You can schedule your departure for the shuttle at check-in, or call down to the front desk at least 4 hours prior to departure when needing to be taken back to the airport. The shuttle is 24/7.
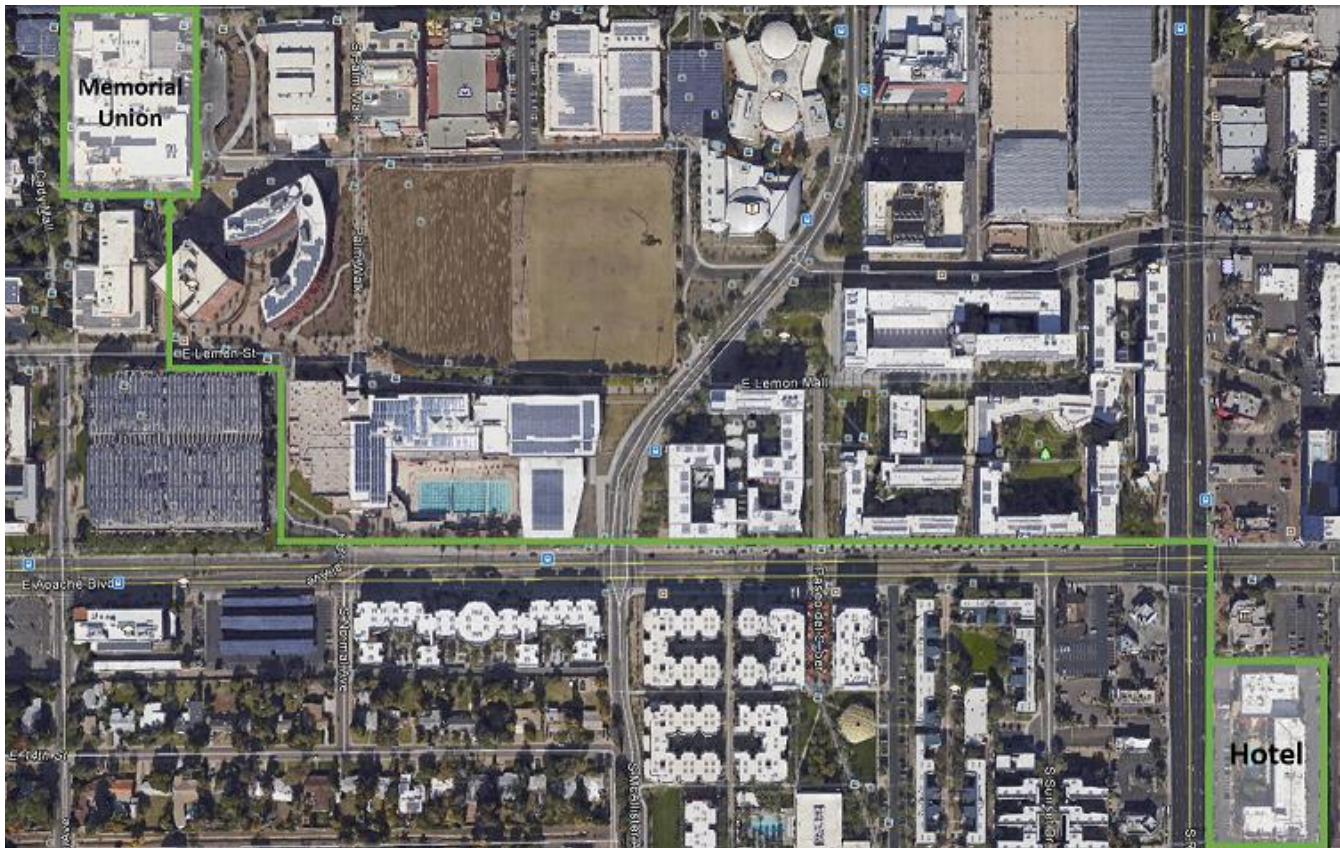

FIGURE 5 – RELATIVE LOCATIONS OF THE WORKSHOP HOTEL AND THE MEETING VENUE

Directions to the Memorial Union from the hotel:

- North to Apache Blvd, turn left
- Turn right on Normal Ave
- Pass through the Business School complex
- Meeting rooms are on 2nd floor of the Memorial Union

Emergency Contact

If you have any travel issues, or difficulties finding the hotel or the workshop venue, please call or email Philip Tarrant (1-623-229-9830, Philip.tarrant@asu.edu) and he will do his best to assist you.

Workshop Expenses

Many of the travel and accommodation arrangements are being paid centrally and will not require reimbursement. Meal expenses will be covered by per diems, which can be claimed using forms available via e-mail or in hard copy at the workshop. For anyone who booked their own travel, the University of Wisconsin will reimburse predetermined amounts based on calculated flight costs.

Organizer Biographies

Corinna Gries is the Information Manager for the North Temperate Lakes Long-Term Ecological Research (NTL LTER) site. She is the Lead Information Manager at the Center for Limnology at the University of Wisconsin-Madison. Prior to this, Gries was the Information Manager at the Center Arizona Phoenix LTER site at Arizona State University. She serves as co-chair for the Long Term Ecological Research Network Information Management Committee and has also been selected to lead the development of a proposal for the next generation LTER Network Information Management Organization (NIMO). Gries was born and raised in Hamburg, Germany and received a PhD in Botany in Kiel, Germany.

Margaret O'Brien is currently the information manager for the Santa Barbara Coastal Long Term Ecological Research project (SBC LTER) at UC Santa Barbara. She coordinates and implements most aspects of data exchange and publication within and outside this group. This includes managing user accounts and services, designing databases, websites and applications, and working with scientists and their staff to plan and publish data products. O'Brien also collaborates with the Ecoinformatics Program at NCEAS, and with the other LTER sites on projects related to data integration and use. Prior to coming to the Marine Science Institute, she worked with the Bermuda Bio-Optics project at the Earth Research Institute, also at UCSB.

Ann Matsushima Chiu is a research librarian. Ann received her MLIS from the University of Illinois, Urbana-Champaign, where her graduate research focused on the zine culture of women of color. Previously she has coordinated the archive-reading room at the Asian American Studies Center Library at UCLA, interned for the Multnomah County Public Library as a metadata librarian, worked at the Oregon Nikkei Endowment as a digital archivist, and at Barnard College Library as a zine librarian. She is currently on the Archives Committee for the Asian Pacific American Librarians Association. She lives in Portland, Oregon with her artist husband and young daughter.

Philip Tarrant is Director of Informatics and Technology and Senior Sustainability Scientist at the Julie Ann Wrigley Global Institute of Sustainability, Arizona State University. In this role he is responsible for

managing the technology needs of the Institute including: research data, the Institute's web presence, application development, and the underlying technology infrastructure. Tarrant serves on ASU's Digital Strategy Board, and is co-chair of the Long Term Ecological Research Network Information Management Committee. Tarrant is originally from England, where he had a successful career in information technology and wireless communications. He moved to Arizona in 2001, and seeking a fresh challenge, he returned to full-time education at Arizona State University in 2002 where he earned a B.S. in Conservation Biology and a M.N.S. His graduate research focused on raptors (birds of prey) in the Phoenix metropolitan area and their response to rapid urbanization. Tarrant actively advocates for high quality information management in the research setting, collaborating with colleagues across Arizona and the rest of the country in pursuit of this goal.

Glyn Thomas is co-founder of AVATAR Inc. an Organization Development Consulting firm that specializes in leadership development, strategic planning, organizational change, and cross-sector partnership building. He has worked across the U.S., the U.K and Egypt. He has advised over 100 senior executives in businesses and governmental agencies in the U.S. Clients include Nike, Texaco, PacifiCorp, Tektronix, LSI Logic, National Park Service, and the U.S. Forest Service. Glyn has a bachelor's degree in education. He works with his wife in their firm, AVATAR, Inc. based in Portland, Oregon.

Attendee Biographies

Kevin Browne is the Information Manager of the UC Natural Reserve System since 2001. He represents the OBFS James Reserve. Previously, he was Data Manager at UC Riverside's James Mountain Reserve. His areas of specialty combine environmental management with computer science, as he consults with field site managers and wildlife scientists regarding effective methods for compiling, managing, and processing data for inclusion in cross-site data management facilities.

Ken Casey is the Deputy Director of the Data Stewardship Division at NOAA's National Center of Environmental Information (NCEI). NCEI is the merger of the three former NOAA National Data Centers: the National Oceanographic Data Center, the National Climatic Data Center, and the National Geophysical Data Center. In this role, he leads the NCEI data stewardship technical direction and advises the Director on technical issues and strategies that match NOAA and Department of Commerce strategic plan objectives. Ken is responsible for service delivery improvements, fostering technical innovation, and overseeing operational mission goals and plans using available data center resources and through national and international data stewardship partnerships. Currently, he is actively involved in NOAA's Big Data Project with Amazon Web Services, Google, IBM, Microsoft Azure, and the Open Cloud Consortium, and leads a high profile project within NOAA known as OneStop, which is working to improve NOAA"s data discovery and access framework. Previously, Ken held positions at NOAA's NODC, as Technical Director from 2008 to 2014 and as Satellite Oceanography Team Lead from 2003 to 2008. He earned his PhD in 1997 from the University of Rhode Island's Graduate School of Oceanography.

Erin Clary is the senior curator at DYRAD Digital Repository and at the Metadata Research Center with the College of Computing and Information at Drexel University since 2014. Previously, she was a metadata specialist in digital publishing at the Carolina Digital Library and Archives. Her areas of specialty are data management, digital curation, data reuse, metadata standards, and data citation.

Cynthia "Cyndy" Chandler is an Information Systems Specialist in the Department of Marine Chemistry and Geochemistry at the Woods Hole Oceanographic Institution. During her career she has participated

in 18 research cruises, predominantly as a CTD technician. Her current research interests are in the field of Ocean Informatics, research that lies at the intersection of computer and information science and oceanography. As a member of several research teams, she uses information systems to facilitate oceanographic research with specific focus on: data management, curation and publication; database design; data system architecture; development of marine ontologies; marine metadata standards; and the use of Semantic Web technologies to facilitate the process by which ocean science data and information become knowledge. She is a co-PI for the US NSF funded Biological and Chemical Oceanography Data Management Office; Lead PI at WHOI for the Rolling Deck to Repository (R2R) project, Lead PI at WHOI for the NSF EarthCube funded GeoLink project; one of the US partners in the Ocean Data Interoperability Platform (ODIP) project; and co-chair of the Marine Data Harmonization Interest Group of the Research Data Alliance. She is currently serving a two-year term as co-chair of the International Oceanographic Data and Information Exchange of the UNESCO Intergovernmental Oceanographic Commission.

Steven Daley-Laursen is Senior Executive in the Office of the Vice President for Research and Economic Development, and Professor in the Department of Natural Resources and Society at the University of Idaho. Dr. Daley-Laursen provides leadership for the development of multi-disciplinary, regional research and outreach initiatives, leads federal government and agency relations, and is PI for the Department of Interior Northwest Climate Science Center. Daley-Laursen was the Founding Director of the university's data management and cyberinfrastructure initiative, the Northwest Knowledge Network, and served as Dean of the College of Natural Resources and as Interim University President. He teaches courses NR Policy, Public Participation, Technology Transfer and Higher Education Leadership.

Robert Downs is the Senior Digital Archivist and Acting Head of Cyberinfrastructure and Informatics research and development at CIESIN, a research and data center of the Earth Institute of Columbia University. Dr. Downs holds the PhD in Information Management from the Stevens Institute of Technology and conducts research on the development and management of information systems, currently focusing on scientific data management and stewardship, data policy, software reuse, digital preservation, and system design and evaluation. He currently serves on the Board of Director of the Foundation for Earth Science, on the Editorial Board for the CODATA Data Science Journal, and as the vice-chair of the Columbia University Morningside Institutional Review Board (IRB). He is a member of the American Geophysical Union (AGU), the Association for Information Science and Technology (ASIS&T), the Association for Computing Machinery (ACM), and the International Association for Social Science Information Services and Technology (IASSIST).

Ruth Duerr is a Research Scholar at the Ronin Institute for Independent Scholarship as well as an adjunct professor, teaching the Foundations of Data Curation at the Graduate School of Library and Information Science at the University of Illinois - Urbana/Champaign. Prior to that, she was the Informatics team lead and manager of the data stewardship program at the National Snow and Ice Data Center (NSIDC). Ruth is a PI/Project Manager for several ongoing data management and cyber infrastructure projects. Her research interests involve nearly all aspects of data stewardship. Active within many of the committees and clusters within the ESIP Federation as well as several Interest Groups within the Research Data Alliance, Ruth is also the primary liaison with the Data Conservancy for NSIDC's ELOKA program; President-elect of the Earth and Space Science Focus Group of the American Geophysical Union and on the board of the Foundation for Earth Science. As a by-product of her teaching duties, several data sets at a variety of repositories have been rescued or made publicly available.

Paul Gessler is Professor of Remote Sensing and Geospatial Ecology in the College of Natural Resources, and Director of the Northwest Knowledge Network (NKN) at the University of Idaho. NKN is a data management and research computing support unit in the Office of Research and Economic Development. He received degrees in Soil Science (B.Sc., 1986) and Environmental Remote Sensing (M.Sc., 1989) from the University of Wisconsin, and a Ph.D. in environmental modeling (Ph.D., 1996) from the Australian National University. Paul's research has involved soil-landscape modeling, remote sensing, and the geospatial characterization of soil patterns and processes via projects around the world. He is currently Co-PI of the $20M NIFA Coordinated Agricultural Project on Regional Approaches to Climate Change for Pacific Northwest Agriculture; Co-PI on a NOAA Climate Impacts Research Consortium involving pacific northwest research universities; and PI on a NSF cyberinfrastructure grant enabling improved network management of research data. NKN is a research data management, high performance computing, and cyberinfrastructure support unit. Paul is active in national and international scientific organizations evolving cyberinfrastructure for research.

Paul C. Hanson is a distinguished research professor at University of Wisconsin Madison's Center of Limnology, which interests in carbon cycling, ecosystem variability, microbes, algae and fish. Hanson is the GLEON Steering Committee Co-Chair, and involved in North Temperate Lakes Long Term Ecological Research Program.

Margaret Hedstrom is the Robert M. Warner Collegiate Professor of Information at the University of Michigan's iSchool. She joined the Michigan faculty in 1995. She is PI for SEAD (Sustainable Environment: Actionable Data), an $8 million project funded by the US National Science Foundation that is building cyberinfrastructure and developing new practices for data sharing, preservation, access and reuse. She is also PI of the Open Data IGERT, a NSF-sponsored traineeship ($2.7 million) in collaboration with Michigan faculty and doctoral students in bioinformatics, computer science, information science, materials science, and chemical engineering that is investigating tools and policies for data sharing and data management. Dr. Hedstrom earned a M.A. in Library and Information Science and a Ph.D. in History at the University of Wisconsin-Madison. She is a Fellow of the Society of American Archivists and has held numerous leadership positions in the area of digital archives and reuse of scientific data.

Nancy Hoebelheinrich is an information analyst, principal and founder of Knowledge Motifs LLC. She has a unique combination of knowledge and skills in working with a variety of data formats and types throughout the lifecycle of the digital resource ranging from creation, retrieval and re-use through to intermediary and archival storage. She has over 15 years of experience working as a liaison between data creators and data managers defining, designing, developing and documenting user needs, in project management, and in working with metadata input / management tools, and archiving requirements. Most recently, she has been working with the University of California Curation Center at the California Digital Library on the Dash data curation portal's software enhancement project, and with the Federation of Earth Science to create a prototype tool / service using Semantic Web technologies to match tool developers with data collection seekers, to manage and support the development and maintenance of ESIP's online data management training resources for scientists, and to participate in the establishment and support of products and services testbeds.

Rick Hooper has been president and executive director of the Consortium of Universities for the Advancement of Hydrologic Science, Inc. (CUAHSI) since 2003. CUAHSI operates that Water Data Center, which is the first data facility funded by NSF for surface earth processes. Prior to this, Hooper was a research hydrologist at the US Geological Survey and worked both at the Panola Mountain Research

Watershed, a small catchment study outside of Atlanta, Georgia, and on the redesign of the National Stream Quality Accounting Network, a continental-scale water-quality monitoring program. He got his PhD in Environmental Systems Engineering from Cornell University and his A.B. in Applied Mathematics from Harvard College.

Greg Janée holds appointments at both the University of California, Santa Barbara (UCSB) and the California Digital Library (CDL). At UCSB he is leading the Data Curation @ UCSB project, an effort to launch new campus-wide data curation services offered by the campus library. At CDL he is the principal developer of the EZID persistent identifier service, and a contributor to CDL's micro-services-based digital preservation architecture. He previously worked on the Alexandria Digital Library and National Geospatial Digital Archive projects.

Matt Jones is the Director of Informatics Research and Development at the National Center for Ecological Analysis and Synthesis, and a Co-PI on the LTER National Communications Office, both at UC Santa Barbara. Jones co-founded the KNB Data Repository, a long-term data archive of environmental data sets, as well as DataONE, a federation of repositories aimed at improving interoperability at global scales. He led the creation and ratification of the Ecological Metadata Language, and helped author the FGDC Biological Data Profile metadata standard. Recent projects focus on metadata improvement and guidance, efforts to achieve environmental data interoperability through semantic modeling of scientific observations, linked data for earth and environmental science, and support for reproducible science via interoperable provenance tracking in commonly used tools like R and Matlab. Jones' research focuses on environmental informatics, including the management, integration, analysis, and modeling of heterogeneous environmental data.

Inna Kouper is a research scientist and assistant director of the Data to Insight Center at Indiana University Bloomington. Her research interests focus broadly on the history and sociology of knowledge production and dissemination, with a particular emphasis on the sociotechnical, or science and technology studies (STS) approaches to research data practices, data science, and visualization. Dr. Kouper has a PhD in information science from the School of Library and Information Science at Indiana University Bloomington and a PhD in sociology from the Institute of Sociology, Russian Academy of Sciences, Moscow, Russia. She was a CLIR data curation postdoctoral fellow and contributed to several publications that focus on metadata and data management and curation. Kouper is involved in the NSF-funded project Sustainable Environment Actionable Data (SEAD) that creates tools and services to facilitate active curation and publication of research data. She is a member of the Research Data Alliance (RDA) Advisory Committee and a Co-PI of the RDA Data Share fellowship program, funded by the Alfred P. Sloan Foundation.

Sherry Lake is the Scholarly Repository Librarian at the University of Virginia Library where she is the public services manager for Libra, UVa's institutional repository. She is currently leading a team to evaluate a locally installed instance of Dataverse. Previously, Sherry worked as the Senior Scientific Data Consultant in the Library's Research Data Services where, for five years, she provided expertise, support and training in the areas of data management, metadata production, data organization and preservation. Sherry was involved in the creation of the DMPTool and continues to support the DMPTool by providing content and expertise on research funder requirements. Sherry also serves as an editor for DataQ and is a partner on the IMLS Grant, "Bridging the Gap between Researcher and Repository. Sherry received her ALA-accredited MS in Information Sciences from the University of Tennessee,

Knoxville and has a BS in Computer Science from the School of Engineering and Applied Science at the University of Virginia.

Kerstin Lehnert is Senior Research Scientist at the Lamont-Doherty Earth Observatory of Columbia University and Director of the NSF-funded data facility IEDA (Interdisciplinary Earth Data Alliance). Kerstin holds a PhD in Petrology from the University of Freiburg in Germany. Over the past 15 years, her work has mostly centered on the development of community-driven data infrastructures for the solid Earth sciences and, in particular, on integrating Earth and space science samples into cyberinfrastructure. Among her primary achievements are the development of the PetDB database, the International Geo Sample Number IGSN, the foundation of an international implementation organization for the IGSN (IGSN e.V.), and most recently the creation of the Coalition for Publishing Data in the Earth & Space Sciences (COPDESS) that brings together publishers and data facilities to implement common policies for data in publications. Kerstin is currently member of the NSF Advisory Committee for Cyberinfrastructure, President of the Earth and Space Science Informatics Focus Group of the American Geophysical Union, member of the AGU Council Leadership Team, and President of the IGSN e.V. She is the newly elected Chair of the EarthCube Leadership Council.

Peter McCartney is a NSF program director in the Division of Biological Infrastructure. He oversees research funded under several programs including Advances in Biological Informatics, National Ecological Observatory Network and Improvements to Field Stations and Marine Laboratories. He also participates in management of several cross directorate programs in software development, data infrastructure, and long term ecological research. Prior to NSF he was a Research Professor in the Global Institute of Sustainability at Arizona State University where he directed projects related to information systems for environmental and archaeological research; use of metadata for designing automated internet access to data and applications; and workflow processing tools for incorporating multiple models into comprehensive analyses.

Carolina Murcia is a Colombian tropical conservation ecologist, with a Ph.D. from University of Florida. She generates and coordinates interdisciplinary and inter-institutional projects in science-based conservation and restoration, focusing mostly in northern South America and Costa Rica. She is currently the Science Director of the Organization for Tropical Studies (OTS), and in such capacity is overseeing the creation of a data management platform to archive and manage data collected by researchers at OTS research stations in the last 50 years. She is also adjunct faculty in the Department of Biology, University of Florida, lecturer at Universidad Javeriana - Seccional Cali in Colombia, and the South American Regional Editor for the journal, Conservation Biology. She also does consulting work in Colombia in conservation and restoration, particularly related to policy and assessments. Her publications span the continuum between science, science transfer and policy

Cynthia Parr is a Technical Information Specialist in the Knowledge Services Division at the National Agricultural Library of the U.S. Department of Agriculture. She leads the Ag Data Commons research data repository project. Previously she served as the Chief Scientist and Director for Species Pages for the Encyclopedia of Life (EOL), based at the National Museum of Natural History of the Smithsonian Institution. She currently serves as the Chair for the Biodiversity Information Standards organization (Taxonomic Database Working Group, TDWG). She serves on the USDA Enterprise Data Management working group, which is drafting departmental policies and manuals, is a member of the BioEco interdepartmental working group, and served on the 2016 Earth Observation Assessment Biodiversity

Societal Benefit Area team. She has conducted research in evolutionary ecology, ornithology, behavior, molecular systematics, community ecology, information visualization, semantic web, and social networks.

Erin Robinson is the Executive Director for the Foundation for Earth Science (FES). FES provides management support to the ESIP Federation as it moved from an operational prototype to an independent organization. FES has since expanded its scope to provide strategic planning and services to a number of Earth science communities. Prior to this, Erin was the Information and Virtual Community Director at FES. During the last 10 years she has created communities and programs with lasting impact around science data and technology. Erin facilitates virtual and in-person collaboration across economic sectors to expedite progress toward data interoperability, envisioning and realizing new programs, and articulating their value to funders to ensure continued success of the programs. She is a natural connector, passionate about fostering innovation through collaboration.

Dave Rugg designed and currently manages the Forest Service's Research Data Archive, established in 2011. He is the author of the Metavist software for editing metadata conforming to the Biological Data Profile of the FGDC Content Standard for Digital Geospatial Metadata. He also organized a special section on data management for the September 2015 issue of the Wildlife Society Bulletin. Mr. Rugg has a BS (Zoology), MS (Ecology), and MS (Statistics). He is a member of the Ecological Society of America, the Research Data Alliance, and the DataONE Users Group. He led the Forest Service proof of concept for data archiving in 2005, and has served as a member of the Federal Government's Interagency Working Group on Digital Data. Most recently, he co-authored USDA's "Implementation Plan to Increase Public Access to Results of USDA-funded Scientific Research", which responds to the OSTP directive on access to scholarly articles and digital scientific data funded by the US Government.

Mark Servilla is the Lead Scientist for the LTER Network Information System (NIS). His primary responsibility is the design and implementation of the NIS — a Service Oriented Architecture framework to support archiving and discovery of LTER data products in a central repository, thus enabling synthetic science at the Network level and beyond. In addition, Mark has been a member of the DataONE Core Cyberinfrastructure Team since the project's inception and currently provides technical support to new and existing DataONE member nodes. Prior to his current position at the LTER Network Office and DataONE, Mark worked in the private sector at Photon Research Associates (PRA), Inc. - he was the architect of a web-based application (GeoServer TM) that provided the discovery, management, and exploitation of geospatial data, including Earth observation imagery and GIS vector objects. In this role, he was responsible for co-management of software development teams that spanned both national and international locations. During his seven years at PRA, Mark participated in numerous software development projects and was involved in the startup of an Internet company that utilized satellite imagery for commercial agriculture. Mark holds a Ph.D. in Earth and Planetary Sciences (Volcanology) and an M.S. in Computer Science, both from the University of New Mexico.

Wade Sheldon is a scientific computing professional in the Department of Marine Sciences at the University of Georgia. He has served as Lead Information Manager for the Georgia Coastal Ecosystems LTER program on Sapelo Island, GA since 2000, and recently took over data management responsibilities for the Savannah River Ecology Laboratory in Aiken, SC. He also provides IT and informatics support for the Coweeta LTER program, UGA Marine Institute, Georgia Coastal Research Council and other small NSF projects at UGA. From 2000-2009 Wade also worked as principal database administrator for the Sapelo Island Microbial Observatory, and provided database design support to the International Census of Marine Microbes, Alpine Lake Microbial Observatory and Mono Lake Microbial Observatory. Outside

of UGA Wade is active in LTER Information Management leadership, where he has served as co-chair of the Network Information System Advisory Committee and currently serves on the Information Management Executive Committee. He has also lead and participated in numerous LTER informatics working groups targeting EML Best Practices, research project metadata management, network web services and personnel database services. In addition he has authored several software packages including the open source GCE Data Toolbox for MATLAB, MATLAB Toolkit for Data Turbine, Metabase Metadata Management System and FLToolbox for MATLAB for analyzing 3D excitation-emission fluorescence spectra.

Dave Vieglais is the Director for Development and Operations at DataONE, overseeing development and implementation of architecture, computer science research, and technological evolution through the activities of the Working Groups and the Cyberinfrastructure CIT, including the staff of full-time developers and post-docs. Dave has extensive experience in developing technical infrastructure for integrating biodiversity information at the global level (i.e. DiGIR, Species Analyst). He also brings significant biodiversity modeling expertise and leadership experience in Global Biodiversity Information Facility (GBIF) and the Natural Science Collections Alliance. He is also a Senior Scientist at the University of Kansas, Lawrence. Dave holds a Ph.D. in Botany from the University of Queensland.

Anne Wilson is a PhD Computer Scientist and has worked in software development and scientific informatics for 25 years. She has been working in scientific labs since 1996, through which she has been exposed to many different kinds of earth science data sets, from geolocated data to spectral observations made in space. Currently she manages LISIRD, the LASP Interactive Solar Irradiance Data Center, which delivers solar irradiance and related data products to the public. In addition to managing software development and operations, Anne works to infuse data management, data stewardship practices, and scientific informatics into organizations in order to improve the practice of science. She organized and coordinates the Boulder Earth and Space Science Informatics Group (BESSIG), which reaches across scientific organizations in Boulder. She is Secretary of AGU's Earth and Space Science Informatics Group, and also participates in the Federation of Earth Science Information Partners (ESIP) and NASA's Earth Science Data Systems Working Group (ESDSWG).

Yaxing Wei is Geospatial Information Scientist at the Oak Ridge National Laboratory Distributed Active Archive Center (ORNL DAAC), one of NASA's 12 data archive centers. Dr. Wei has been heavily involved in geospatial data management, curation, and quality assurance and quality check. He has been leading the development of Web-based geospatial data visualization and distributions tools and services, such as OGC standards-based Spatial Data Access Tool (SDAT) and THREDDS Data Server. Dr. Wei is a co-I of the NASA-funded Multi-scale Synthesis and Terrestrial Model Intercomparison Project (MsTMIP) and providing data management support for 20+ terrestrial ecosystem models participating MsTMIP. He has been previously participated the NSF-funded DataONE project, and is an active member of the former DataONE Scientific Exploration, Visualization, and Analysis Working Group and Scientific Workflows and Provenance Working Group. Dr. Wei is currently member of a number of Working Groups, including Data Quality, Interoperability, and Provenance, of the NASA Earth Science Data System Working Groups (ESDSWG).