



“Managing Digital Research Objects in an Expanding Science Ecosystem”

Wednesday, November 15th, 2017

9:00 am - 5:00 pm EST

Lister Hill National Center for Biomedical Communications

U.S. National Library of Medicine/NIH

8600 Rockville Pike, Bethesda, MD 20894

AGENDA

9:05 - 9:15 am

Welcome and Opening Remarks

Michael Huerta, Associate Director of the US National Library of Medicine/NIH and NLM Coordinator of Data and Open Science Initiatives, NIH

9:15 - 10:00 am

Digital Objects – The Core of Our Complex Data Market

Introduction: George Strawn, Director, Board on Research Data and Information, The National Academies of Science, Engineering, and Medicine

A detailed overview of a number of open questions still being debated within communities, such as granularity, versioning, mutability, lifecycle aspects, etc. and argument for concrete steps to develop a concept and a core model to make it the basis of software developments that will construct the global data market. The data community has experience from over 20 years with PID and metadata systems enabling us to build the open data market.

- *Peter Wittenburg, RDA Director Europe, Max Planck Computing and Data Facility*
-

10:00 - 11:00 am

Developing Frameworks

Moderator: George Strawn, Director, Board on Research Data and Information, The National Academies of Science, Engineering, and Medicine

This session will address both technical and institutional frameworks. The RDA Data Fabric Group was formed to bridge across research data management from many different perspectives and identify the minimal set of common components upon which domain-specific workflows may be built. The Digital Object model has been adopted by the group and the scientific data management frameworks that are emerging from this work share a common set of components: identifiers; an identifier resolution system, specifically the handle system; fine-grained data types; a federated set of data type registries; digital object repositories; and digital object metadata registries. In addition to the technical challenges and opportunities, successful leveraging of DRO's will require new social infrastructure. This social fabric may include new policies, organizational arrangements, practices, and communication channels. This

session will present complementary approaches to both technical and social frameworks.

- **Larry Lannom**, Director of Information Services and Vice President at the Corporation for National Research Initiatives (CNRI), RDA Data Fabrics Group
 - **Brooks Hanson**, Senior Vice President, Publications, American Geophysical Union
-

11:00 - 11:15 am **Break and Networking Opportunity**

11:15 am – 12:45 pm **Incorporating Identifiers as Part of the Scholarly Record**

Moderator: **Marcie Granahan**, Executive Director, The National Federation of Advanced Information Services (NFAIS™)

The scholarly record is evolving beyond the published article. It is now common practice to include digital identifiers for funders, researchers, and data sets, and standards are being developed to ensure consistency across content providers and aggregators. Identifier best practices also impact how science gets done, and science can be more impactful and better utilized when databases, journals, and content, in general, can consistently reference an object to its authoritative source. A shared identifier system facilitates text mining for related content, resource tracking and reproducibility, attribution and dataset indexing among and across different yet connected objects. In this session, we'll discuss opportunities for the use of identifiers, challenges, emerging standards, and best practices that enable connectivity between objects in a synthesized, efficient and effective way.

- “Potholes” in Creating and Maintaining Object Identifiers
Julie McMurry, Software Project Manager, Senior Research Associate in the Library and the Department of Medical Informatics and Clinical Epidemiology, Oregon Health and Science University
 - Standards Being Developed
Todd Carpenter, Executive Director, National Information Standards Organization (NISO)
 - Persistent Identifiers for Data Sets
Patricia Cruse, Executive Director, DataCite
 - Industry Incorporation of Identifiers
Mike Yalter, Software Developer, CrossRef
-

12:45 pm – 1:30 pm **Lunch** (to be provided)

1:30 pm – 2:00 pm **Research Objects: More than the Sum of the Many Parts**

Moderator: **Bonnie C. Carroll**, Project Manager, CENDI

The experimental methods, computational codes, data, algorithms, workflows, Standard Operating Procedures, samples and so on are the objects of research that enable reuse and reproduction of scientific experiments. They need to be examined and exchanged as research knowledge. Think of DROs as a broadening out to embrace these assets of research. The next step is to recognize that investigations use multiple, interlinked, evolving artefacts. Multiple datasets and multiple models support a study; each model is associated with datasets for construction, validation and prediction; an analytic pipeline has multiple codes and may be made up of nested sub-pipelines, and so on. Research Objects (<http://researchobject.org/>) is a framework by which the many, nested and contributed components of research can be packaged together in a systematic way, and their context, provenance and relationships richly described.

- **Carole Goble**, Professor, School of Computer Science, The University of Manchester
-

2:00 pm – 3:30 pm

Connecting Users with Digital Research Objects

Moderator: Lynn Yarmey, Community Development Director, Research Data Alliance/US

This session aims to ensure strong DRO drivers and connections to research by highlighting the importance of maintaining a user perspective in development and management. Speakers will describe different approaches to DRO planning and implementation in service to their research community needs, and will share lessons learned from their user-focused work to date. Short presentations will be followed by a panel discussion.

- **Lisa Kempler**, Co-Chair, Use Case Working Group, EarthCube; MATLAB Community Strategist at MathWorks
- **David Vieglais**, Senior Scientist at University of Kansas; Director of Development and Operations, DataONE
- **Danie Kinkade**, Information Systems Associate at the Biological and Chemical Data Management Office (BCO-DMO) at the Woods Hole Oceanographic Institution
- **James Myers**, Co-Principal Investigator, Sustainable Environment/Actionable Data (SEAD) Project with the National Data Service; Associate Research Scientist at University of Michigan

3:30 pm – 3:45 pm

Break and Networking Opportunity

3:45 pm – 4:45 pm

Filling the Gaps and Moving the Agenda Forward

Moderator: Amanda J. Wilson, Head, National Network Coordinating Office, National Library of Medicine/NIH; Chair, CENDI

Group discussion of the challenges, solutions, gaps, and the steps we can take to move the agenda forward.

Responder: Jim Hendler, Director, Institute for Data Exploration and Applications and Tetherless World Chair of Computer, Web and Cognitive Sciences, Computer Science, Rensselaer Polytechnic Institute; Member, Board on Research Data and Information (BRDI), The National Academies of Science, Engineering, and Medicine

4:45- 5:00 pm

Wrap Up and Adjournment

Marcie Granahan, Executive Director, The National Federation of Advanced Information Services (NFAIS™)

Bonnie C. Carroll, Executive Director, CENDI

Contact Information:

Heather M. Parrish
CENDI
E-mail: hparrish@iaweb.com
Phone: 865-298-1245

Nancy Blair-DeLeon
NFAIS
E-mail: nblairdeleon@nfais.org
Phone: 443-221-2980, ext. 102

Lynn Rees Yarmey
RDA/US
E-mail: yarmel@rpi.edu
Phone: 858-722-0127

Ester Sztein
NAS
E-mail: esztein@nas.edu
Phone: 202-334-3049