Research Data Alliance

Fall 2013 CNI Member Meeting

Chris Greer
NIST
Larry Lannom
CNRI
Research Data Alliance

Vision

• Researchers and innovators openly share data across technologies, disciplines, and countries to address the grand challenges of society.

Mission

• RDA builds the social and technical bridges that enable open sharing of data.
The RDA Community today:
Over 1000 members from 55 countries

Map courtesy traveltip.org
RDA Plenaries: Venue for community building and WG / IG progress

- **RDA Plenary 1 / Launch**
  - March 2013 in Gothenburg, Sweden
  - 240 participants
  - 3 WG, 9 IG

- **RDA Plenary 2**
  - September 2013 in Washington, DC
  - 380 participants
  - 6 WG, 17 IG, 5 BOF
  - Data Citation Summit co-located in RDA “neutral space”
  - First Organizational Assembly meet-up
RDA Organizational Structure

RDA Council
Responsible for overarching mission, vision, impact of RDA

Technical Advisory Board
Responsible for Technical roadmap and interactions

Secretary-General and Secretariat
Responsible for administration and operations

Organizational Advisory Board and Organizational Assembly
Responsible for organizational and strategic advice

Working Groups
Responsible for impactful, outcome-oriented efforts

Interest Groups
Responsible for defining and refining common issues

RDA Colloquium
Operational and community sponsorship
Organizational Evolution Over the Last Year

RDA Council
7 out of 9 Council members now appointed, all appointed by Plenary 3

Technical Advisory Board
11 out of 12 TAB members now chosen, all 12 chosen by Plenary 3

Secretary-General and Secretariat
Secretary General recruitment ongoing, new Secretary General to be in place by Plenary 3

Organizational Advisory Board and Organizational Assembly
25 organizations interested in Membership; 7 organizations interested in Affiliate status

Working Groups
3 WGs at Plenary 1, 6 WGs at Plenary 2, potentially 12+ WGs at Plenary 3

Interest Groups
9 IGs at Plenary 1, 17 IGs at Plenary 2 + 5 BOFs, 29+ IGs at Plenary 3 + ?? Birds-of-a Feather

RDA Colloquium
Operational and community sponsorship
RDA Organizational Partners

**Member Applicants**
- Barcelona Supercomputing Center
- European Data Infrastructure (EUDAT)
- International Association of STM Publishers
- New Zealand eScience Infrastructure
- Washington University Libraries
- Purdue University Libraries
- Research Data Canada
- eResearch Services and Scholarly Application Development Division of Information Services
- American University Library

**Other interested Organizations**
- Australian Antarctic Data Centre
- Australian National Data Service
- CERN
- CJSD Consulting
- Columbia University Libraries/Information Services
- CSC - IT Center for Science Ltd.
- Digital Curation Centre
- IBM

**Interested Affiliates**
- Institute for Quantitative Social Science at Harvard
- Intersect Australia Limited
- Microsoft
- Oracle
- STFC - Science & Technology Facilities Council
- Corporation for National Research Initiatives (CNRI)
- Terrestrial Ecosystems Research Network
- University of Michigan Libraries
- Committee on Data for Science and Technology (CODATA)
- Connecting Research and Researchers (ORCID)
- DataCite
- International Oceanographic Data and Information Exchange (IODE)
- Scholarly Publishing and Academic Resources Coalition (SPARC)
- World Data System (WDS)
- W3C
RDA Community-Driven Groups

- **Birds-of-a-Feather (met at Plenary 2)**
  - Linked Data
  - Chemical Safety Data
  - Education and Skills Development in Data Intensive Science
  - Libraries and Research Data
  - Cloud Computing and Data Analysis Training for the Developing World

- **Working Groups**
  - Data Type Registries
  - Persistent Identifier Types
  - Data Foundations and Terminology
  - Metadata Standards
  - Practical Policy
  - Data Categories and Codes
  - **WG Case statements being prepared**: Citing Dynamic Data, Publishing Data Workflows, Publishing Data Services, Data Bibliometrics, Cost Recovery Models for Repositories, Data Descriptions Registry

- **Descriptions Registry Interoperability, DSA-WDS Partnership Working Group on Certification**

- **Interest Groups**
  - Agricultural Data Interoperability
  - Certification of Trusted Repositories (joint with ICSU-WDS)
  - Data Citation
  - Metadata
  - Marine Data Harmonization
  - Community Capability Model
  - Engagement
  - Preservation e-Infrastructure
  - Legal Interoperability (joint with CODATA)
  - Defining Urban Data Exchange for Science
  - Marine Data Harmonization
  - Structural Biology
  - Big Data Analytics
  - Data Brokering
  - Publishing Data (joint with WDS)

- Toxicogenomics Interoperability
- Research Data Provenance
- Materials Data Management
- Global Registry of Trusted Data Repositories and Services
- Digital Practices in History and Ethnography
- Biodiversity Data Integration
- Long tail of Research Data
- Development of cloud computing capacity and education in developing world
- Service Management IG (pending)
- Domain Repositories Interest Group (pending)
- Federated Identity Management (pending)
- Persistent Identifier Interest Group – PID-IG (pending)

*Blue* = new between Plenary 1 and Plenary 2
*Green* = new since Plenary 2
## Community-Driven RDA Groups by Focus

### Domain Science - focused
- Toxicogenomics Interoperability IG
- Structural Biology IG
- Biodiversity Data Integration IG
- Agricultural Data Interoperability IG
- Digital History and Ethnography IG
- Defining Urban Data Exchange for Science IG
- Marine Data Harmonization IG
- Materials Data Management IG

### Community Needs - focused
- Community Capability Model IG
- Engagement IG
- Clouds in Developing Countries IG

### Reference and Sharing - focused
- Data Citation IG
- Data Categories and Codes WG
- Legal Interoperability IG

### Data Stewardship - focused
- Research Data Provenance IG
- Certification of Digital Repositories IG

### Data Stewardship - focused
- Preservation e-infrastructure
- Long-tail of Research Data IG
- Publishing Data IG
- Domain Repositories IG
- Global Registry of Trusted Data Repositories and Services IG

### Base Infrastructure - focused
- Data Foundations and Terminology WG
- Metadata Standards WG
- Practical Policy WG
- PID Information Types WG
- Data Type Registries WG
- Metadata IG
- Big Data Analytics IG
- Data Brokering IG
First RDA Infrastructure Deliverables in 2014 (1)

- **Data Type Registries WG**
  - Defining a system of data type registries
  - Defining a formal model for describing types and building a working model of a registry.
  - To be adopted by CNRI, International DOI Foundation, and used by the Deep Carbon Observatory and others
  - (working in conjunction with PID group)
  - Scheduled to complete Summer, 2014

- **Persistent Identifier Information Types**
  - Defining a minimal set of types that must be associated with a PID (e.g. checksum, author).
  - Specifying an API for interaction with PID types
  - Adopted and used by Data Conservancy and DKRZ
  - (working in conjunction with DTR group)
  - Scheduled to complete Summer, 2014

- **Metadata Standards**
  - Creating use cases and prototype directory of current metadata standards from starting point of DCC directory and stakeholder contributions.
  - To be hosted and used by JISC, DataOne and others
  - Scheduled to complete Fall, 2014
First RDA Infrastructure Deliverables in 2014 (2)

- **Data Categories and Codes**
  - Operationalization of ISO language categories for repositories
  - Adopted and used by the Language Archive, PARADISEC
  - Proposal of data categories associated with the CMDI schema as ISO standards.
  - Scheduled to complete Fall, 2014

- **Data Foundations and Terminology**
  - Defining a common vocabulary for data terms based on existing models.
  - Creating formal definitions in a structured vocabulary too which also provides an open registry for data terms.
  - (active input from all RDA WGs)
  - Tested and adopted by EUDAT, DKRZ, Deep Carbon Observatory, CLARIN, EPOS, and others
  - Scheduled to complete Summer, 2014

- **Practical Policy**
  - Survey of policies in production use across data management centers. Test bed of machine-actionable policies (IRODS, DataVerse, dCache) at RENCI, DataNet Federation Consortium, CESNET, Odum Institute.
  - Deployment of 5 policy sets (integrity, access control, replication, provenance / event tracking, publication) on test beds. Publication of standard policies for use as starter kits.
  - Scheduled to complete Summer, 2014
Coming in 2014

- **RDA Plenary 3**
  - March 26-28, 2014 in Dublin, Ireland
  - Hosted by Australia and Ireland
  - Theme: “The Data Sharing community - Playing Your Part”

- **RDA Plenary 4**
  - September 2014 in The Netherlands
  - Being planned now …
Regional RDA groups: Collaborate Globally, Contribute Locally

Regional RDA groups:

– Contribute to RDA Secretariat and operations

– Support plenaries and promote attendance from RDA members in the region

– Implement RDA deliverables regionally to amplify impact

– Bring regional issues to broader RDA community

– Collaborate on outreach, student / early career, and other programs
Working with the Research Data Alliance
What does the Digital Government Strategy mean for digital access?

“I want us to ask ourselves every day, how are we using technology to make a real difference in people’s lives.”

– President Barack Obama

“Unlock the power of government data to spur innovation across our Nation and improve the quality of services for the American people.”

- President’s Digital Government Strategy
What guidance has NIST received?

Increase Access to the Results of Federally Funded Scientific Research

The Office of Science and Technology Policy hereby directs each Federal agency ... to develop a plan to support increased public access to the results of research funded by the Federal Government.

- OSTP Memorandum, February 2013

Manage Information as an Asset

This Memorandum requires agencies to collect or create information in a way that supports downstream information processing and dissemination activities.

- OMB Memorandum M-13-13, May 2013
Conceptual Model: Problem Space

Data
- Reference
- Resource
- Research

Software
- Production
- Tools
- Test, Development, ...

Publications
- Peer Reviewed
- Gray Literature
- White Papers, Talk Slides, ...

NIST Public Servers
NIST Internal Servers
Other NIST Storage
Publishers
Community Repositories
Other Fed Agency Repositories
Cloud
Conceptual Model: Solution Properties

- Must work with 3\textsuperscript{rd} Party repositories
- Must enable search engines to access metadata regardless of location
- Must support the metadata mandated by OMB
- Must support multiple types of metadata
- Must support persistent identifiers for data, metadata, and services
Research Data Alliance

Data Type Registries Working Group

- Larry Lannom, Daan Broeder
- Design a functional specification for type registries and propose a federation strategy

PID Information Types Working Group

- Tim DiLauro, Tobias Weigel
- Give every data object an identity that enables referring to the data resources and metadata
NIST Big Data Public Working Group & Standardization Activities

Wo Chang, NIST, wchang@nist.gov
Robert Marcus, ET-Strategies
Chaitanya Baru, UC San Diego

http://bigdatawg.nist.gov
SUBGROUPS

- Definitions and Taxonomies
- Requirements and Use Cases
- Technology Roadmap
- Reference Architecture
- Security and Privacy
Definitions and Taxonomies Subgroup

• **Data Science** is the extraction of actionable knowledge directly from data through a process of discovery, hypothesis, and analytical hypothesis analysis.
Requirements and Use Case Subgroup

51 Use Cases Received http://bigdatawg.nist.gov/usecases.php

1. **Government Operations (4):** National Archives & Records Administration, Census Bureau
2. **Commercial (8):** Finance in Cloud, Cloud Backup, Mendeley (Citations), Netflix, Web Search, Digital Materials, Cargo shipping (e.g. UPS)
3. **Defense (3):** Sensors, Image Surveillance, Situation Assessment
4. **Healthcare & Life Sciences (10):** Medical Records, Graph & Probabilistic Analysis, Pathology, Bio-imaging, Genomics, Epidemiology, People Activity Models, Biodiversity
5. **Deep Learning & Social Media (6):** Driving Car, Geolocate Images, Twitter, Crowd Sourcing, Network Science, NIST Benchmark Datasets
6. **The Ecosystem for Research (4):** Metadata, Collaboration, Language Translation, Light Source Experiments
7. **Astronomy & Physics (5):** Sky Surveys, Large Hadron Collider at CERN, Belle Accelerator II (Japan)
9. **Energy (10):** Smart Grid

Reference Architecture Subgroup

M0039 | Data Processing Flow

M0017 | Data Transformation Flow

M0047 | IT Stack

Research Data Alliance

Big Data Analytics Interest Group

• Rahul Ramachandran
• Develop community based recommendations on feasible data analytics approaches to address scientific community needs of utilizing large quantities of data.

www.rd-alliance.org/internal-groups/big-data-analytics-ig.html
WWW.RD-ALLIANCE.ORG

RDA is recruiting a Secretary General - Find out more!

Marine Data Harmonization IG
Status: Pending Action
The objective of this working group is to promote the development of a common global framework for the management of marine data to address the impact of natural and man-made phenomena. Multi-disciplinary collaboration and large volumes of good quality interoperable data which can be easily located and accessed is required for this to support the development of this ecosystem based approach to...