



Update on the Research Data Alliance

Fran Berman, RPI
Chair, RDA / US

research data sharing without barriers
rd-alliance.org

When we last spoke (January 2013) ...

First RDA organizational telecon: August 2012

Global Data Planning Meeting: October 2012

**T
H
E
N**

RDA was “pre-official” and pre-launch. The Organizing Committee consisted of 8 individuals from 5 countries

**RDA Launch /
First Plenary**

**RDA Second
Plenary**

**RDA Third
Plenary**

March 2013

September 2013

March 2014

Organization of DAITF: 2011-2012

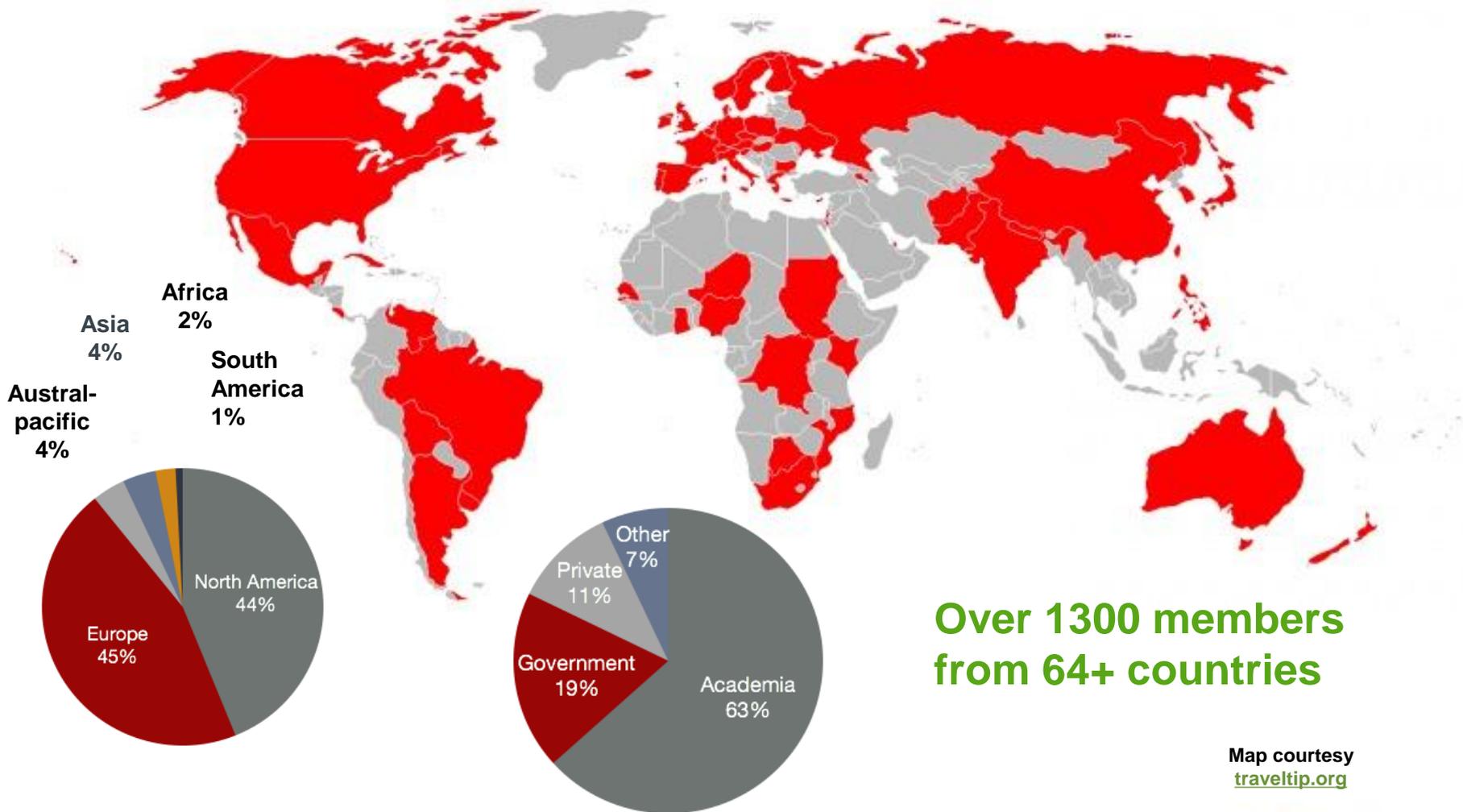
“Data Web Forum” White Paper (Greer and Blatecky)

RDA is moving towards “steady state”. The RDA community has over 1300 members from 64+ countries.

**N
O
W**

The Research Data Alliance Community

(as of 2-14)



Over 1300 members
from 64+ countries

Map courtesy
traveltip.org



Recap: Why RDA? What is RDA?



Motivation: Data Sharing a Critical Accelerant for New Discovery and Innovation



InformationWeek Healthcare Digital Bundle

Software Security Cloud Mobility Social Business Big Data Windows Global CIO Government Healthcare Education Financial SMS More

Electronic Medical Records Mobile & Wireless Clinical Information Systems Security & Privacy CPOE The Patient Leadership More Healthcare

HP: The power of HP Converged Infrastructure is here. Now you can afford them. Watch video

NEWS: Sharing Psychiatry EHR Data Cuts Readmission Rates

Get InformationWeek Daily

Don't miss each day's hottest technology news, sent directly to your inbox, including occasional breaking news alerts.



OFFICE OF JUSTICE PROGRAMS

NATIONAL INSTITUTE OF JUSTICE

Research • Development • Evaluation

HOME | FUNDING & AWARDS | PUBLICATIONS & MULTIMEDIA | EVENTS | TRAINING | TOPICS

NIJ Home Page > NIJ Journal > NIJ Journal No. 267

NIJ JOURNAL NO. 267

In Brief: Expanding Research by Sharing Data

by NIJ staff

Director's Message: NIJ makes data available for future research.

Police Use of Force: The Impact of Less-Lethal Weapons and Tactics

Toward a Better Way to Interview Child Victims of Sexual Abuse



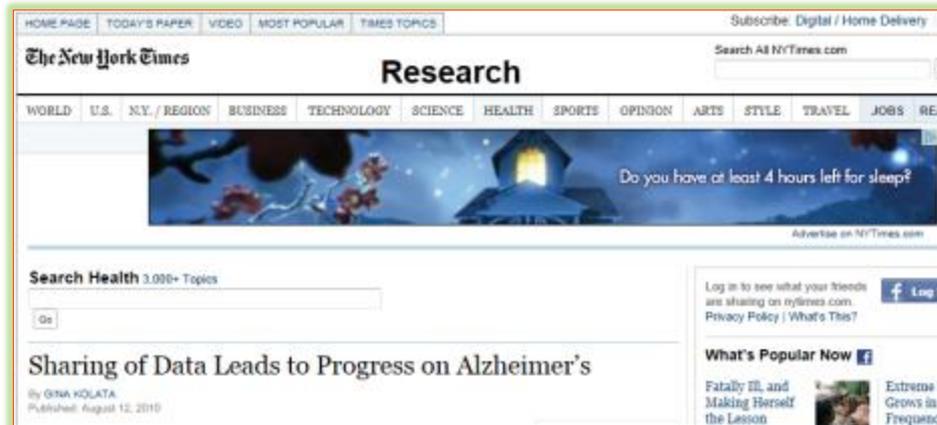
News

ADMISSIONS, VISIT, DISCOVERY, ACCESS

Back to All News

Astronomers Release Unprecedented Data Set on Celestial Objects that Brighten and Dim

PASADENA, Calif.—Astronomers from the California Institute of Technology (Caltech) and the University of Arizona have released the largest data set ever.



HOME PAGE TODAY'S PAPER VIDEO MOST POPULAR TIMES TOPICS

Subscribe: Digital / Home Delivery

The New York Times

Research

WORLD U.S. N.Y. / REGION BUSINESS TECHNOLOGY SCIENCE HEALTH SPORTS OPINION ARTS STYLE TRAVEL JOBS REAL ESTATE

Do you have at least 4 hours left for sleep?

Search Health 3,000+ Topics

Log in to see what your friends are sharing on nytimes.com. Privacy Policy | What's This?

What's Popular Now

Fatally Ill, and Making Herself the Lesson

Extreme Weather Grows in Frequency

Sharing of Data Leads to Progress on Alzheimer's

By GRINA HOLATA

Published August 12, 2010



nature medicine

nature.com • journal home • archive • issue • news • abstract

ARTICLE PREVIEW

view full access options

NATURE MEDICINE | NEWS

BARriers

The delay in sharing research data is costing lives

Josh Sommer

Data infrastructure Fundamental to Support Data Sharing and Data-driven Innovation

Data discoverability **tools**

Data access via **portals**, science gateways, etc.

Data collection, storage and access **systems**

Data **services** to support use and re-use

Data analysis **algorithms**

Data-driven **models and simulations**

Data **visualization** tools

Semantic **frameworks**

Data management frameworks

Expert **assistance** for data-oriented applications, ...

World-wide Need for More Effective Infrastructure to Support Research Data Sharing

A Europe-Japan-United States GNSS data-sharing pilot project for the Geohazard Supersites and Natural Laboratories

Falk Amelung, University of Miami, USA (GEO task lead)
Craig Dobson, NASA and Committee of Earth Observation Satellites (CEOS)
Rui Fernandes, EOS and EUREF <rmanuel@di.uibj.pt>



Science, Humanities, Arts
Communities



Cyberinfrastructure professionals,
data analysts, data center staff, ...



Libraries, Archives,
Repositories, Museums



**National Data Sharing and
Accessibility Policy-2012
(NDSAP-2012)**

Department of Science & Technology
Ministry of science & Technology
Government of India

Data
Scientists

A screenshot of the Australian National Data Service (ANCS) website. The main heading is "Australian National Data Service" with the vision statement: "Our Vision: More Australian researchers reusing research data more often". Below this, it states "ANCS is enabling the transformation of:" followed by a diagram showing the transition from "Data that are:" (Unmanaged, Disconnected, Invisible) to "Structured Collections that are:" (Managed, Connected, Findable). The diagram uses arrows to show the flow from the left to the right. There is also a sidebar with navigation links and a news section.



The Research Data Alliance (RDA)

- Global community-driven organization launched in March 2013 to accelerate data-driven innovation
- RDA focus is on building the **social, organizational and technical infrastructure** to
 - *reduce barriers to data sharing and exchange*
 - *accelerate the development of coordinated global data infrastructure*



- **Research Data Alliance Vision:** *Researchers and innovators openly share data across technologies, disciplines, and countries to address the grand challenges of society.*
- **Research Data Alliance Mission:**
RDA builds the social and technical bridges that enable data sharing.



RDA Approach: CREATE → ADOPT → USE



RDA Members come together as

- **Working Groups** – 12-18 month efforts to build, adopt, and use specific pieces of infrastructure
- **Interest Groups** – longer-lived discussion forums that spawn Working Groups as specific pieces of needed infrastructure are identified.

Working Group efforts focus on the development and use of data sharing infrastructure

- **Code, policy, infrastructure, standards, or best practices that are adopted and used** by communities to enable data sharing
- **“Harvestable” efforts** for which 12-18 months of work can eliminate a roadblock
- **Efforts that have substantive applicability** to groups within the data community, but may not apply to everyone
- **Efforts for which working scientists and researchers can start today**

RDA Evolution



RDA Current Status



RDA Interest (IG) and Working Groups (WG) by

13

Focus (as of 12/13, more by Plenary 3)

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
- 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

Coming in 2014: First RDA Infrastructure Deliverables

Scheduled to Complete Summer 2014

Data Type Registries WG

- **Deliverables:** System of data type registries, formal model for describing types, working model of a registry.
- **Initial Adopters and Users:** CNRI, International DOI Foundation, Deep Carbon Observatory

Practical Code Policies

- **Deliverables:** Survey of policies in production use, testbed of machine actionable policies, deployment of 5 policy sets, policy starter kits
- **Initial Adopters and Users:** RENCI, DataNet Federation Consortium, CESNET, Odum Institute

Persistent Identifier Information Types

- **Deliverables:** Minimal set of PID types, API
- **Initial Adopters and Users:** Data Conservancy, DKRZ

Scheduled to Complete Fall 2014

Language Codes

- **Deliverables:** Operationalization of ISO language categories for repositories.
- **Initial Adopters and Users:** Language Archive, Paradisec

Data Foundations and Terminology

- **Deliverables:** Common vocabulary for data terms, formal definitions and open registry for data terms
- **Initial Adopters and Users:** EUDAT, DKRZ, Deep Carbon Observatory, CLARIN, EPOS

Metadata Standards

- **Deliverables:** Use cases and prototype director of current metadata standards starting from DCC directory
- **Initial Adopters and Users:** JISC, DataOne

RDA Plenaries Emerging as a Data Community “Town Square”

Emerging Plenary Format:

- **Working sessions:** Face-to-face opportunities for global Interest Groups, Working Groups, and BOFs to meet and advance their agendas
- **All-hands sessions:** Place for community networking and exchange of information (funding agencies, data organizations, key stakeholders)
- **Neutral meeting place:** Place for multiple groups to meet and form a common agenda and action plan (e.g. Plenary 2 Data Citation Harmonization Summit)



To be launched at Plenary 3: RDA Organizational Assembly

Organizational Affiliates (MOUs in process)

- CODATA
- World Data system
- International Oceanographic Data and Information Exchange
- DataCite
- ORCID

Organizational Members (Confirmed)

- Barcelona supercomputing Center
- Australian National Data Service
- International Association of STM Publishers
- European Data Infrastructure (EUDAT)
- New Zealand eScience Infrastructure
- Purdue University Libraries
- STFC (Science and Technology Facilities Council)
- Research Data Canada
- eResearch Services and Scholarly Application Development Division of Information Services, Griffith University
- American University Library
- Scholarly Publishing and Academic Resources Coalition
- Washington University in St. Louis Libraries

Organizational Member Interest

- Microsoft
- IBM
- Oracle
- Corporation for National Research Initiatives
- Australian Antarctic Data Centre
- Intersect Australia United
- Terrestrial Ecosystems Research Network
- CSC-IT Center for Science Ltd.
- Corporation for National Research Initiatives (CNRI)
- CERN
- Columbia University Libraries Information Services
- CJSJ Consulting
- Institute for Quantitative Social Science
- Purdue University Libraries
- Digital Curation Center
- Scholarly Publishing and Resources Coalition

Affiliate Interest

- World Wide Web Consortium
- International Oceanographic Data and Information Exchange
- Global Alliance for Genomics and Health

RDA Organizational Structure

Provides Coordinating Framework

RDA Membership

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 (Research Funders)

Operational and community sponsorship

- **Create a pipeline of data sharing infrastructure efforts**
 - that are adopted and used by communities during their development
 - that increase their impact through greater adoption over time
- **Build and expand the research data community for effective impact**
 - globally, regionally, and within constituent groups
- **Evolve as a useful, relevant, and agile organization**
 - that helps the community capitalize on opportunity and respond to challenges within the data community

RDA/US: Collaborate Globally, Contribute Locally

RDA/US Goals:

- ← Contribute to RDA “international” efforts and leadership
- ← Bring US efforts to broader RDA community
- **Build the RDA community within the US**
- **Leverage and implement RDA deliverables in the US to amplify impact**
- ↔ Collaborate closely with other RDA “regions” on key programs and initiatives



NSF-supported RDA/US initiatives:

- Outreach (RDA → RDA/US)
- RDA Deliverables Amplification
- Student / Early Career Engagement

RDA/US Steering Committee

- Fran Berman, RPI
- Larry Lannom, CNRI
- Mark Parsons, RPI
- Beth Plale, IU

- Participant support for Working Groups and Interest Groups to convene outside of RDA Plenaries
- Support for U.S. early career professionals and students to participate in RDA and RDA/US
- Support for further / expanded / amplified implementation, adoption and use of RDA deliverables within the U.S.
- Support for RDA/US Plenaries and co-sponsored events
- Engagement with areas / agencies / sectors for whom RDA can collaborate to expand infrastructure and impact



RDA, RDA/US and CENDI -- How can we better support one another's efforts?





Thank You!

bermaf@rpi.edu
enquiries@rd-alliance.org

research data sharing without barriers
rd-alliance.org