



## **Building the Research Data Alliance**

**Dr. Beth Plale**

Vice Chair of Technology Programs, RDA/US  
Indiana University

**research data sharing without barriers**  
**[rd-alliance.org](http://rd-alliance.org)**

# Data Driving Solutions to Complex Science and Societal Challenges

2

What is the likely impact of a large-scale earthquake?



Earthquake simulations integrate

- Fault geometry data
- + Subsurface data
- + Structure data
- + Population data
- + Physical infrastructure data

Disease spread models couple

- Medical data
- + Population data
- + Env't. data

How does disease spread in large urban populations?



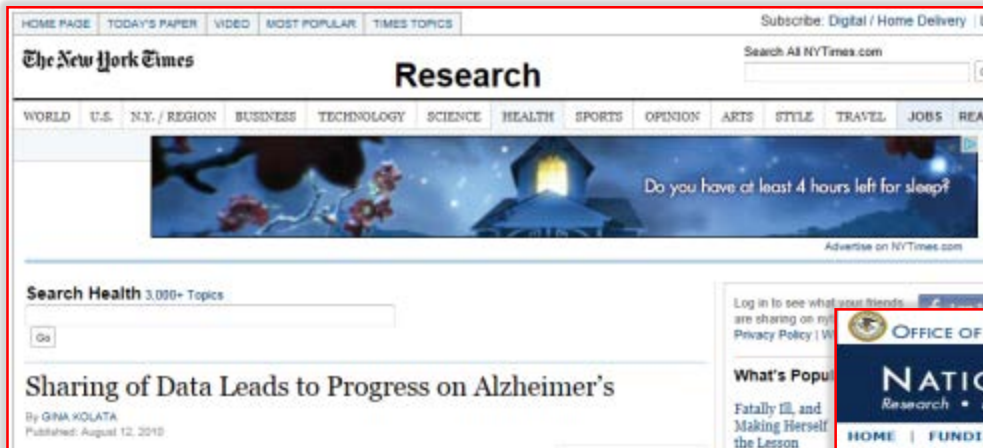
How can we increase wheat yields?



Productivity analysis leverages interoperability of

- Germplasm data
- + Genetic and phenotypic data
- + Statistical data
- + Bibliographic data

# Sharing and Exchange of Data Leading to an Acceleration of Innovation



The New York Times Research page. The header includes navigation links like 'HOME PAGE', 'TODAY'S PAPER', 'VIDEO', 'MOST POPULAR', and 'TIMES TOPICS'. A search bar is present. The main content area features a large image of a house at night with the text 'Do you have at least 4 hours left for sleep?'. Below this is a search bar for 'Health' with 3,000+ topics. A featured article is titled 'Sharing of Data Leads to Progress on Alzheimer's' by GINA KOLATA, published August 12, 2010.



Article titled 'Astronomers Release Unprecedented Data Set on Celestial Objects that Brighten and Dim'. The page features a dark background with a grid of celestial objects and connecting lines, suggesting a data visualization or network graph.



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

NIJ makes data available for future research.

Director's Message

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

Toward a Better Way to Interview Child Victims of Sexual Abuse



VB  
BIG DATA

THE WALL STREET JOURNAL  
Mortgages Fade at Wells Fargo

channels:  
Main  
Big Data  
Business

**Tech needs leverage data to help you avoid the flu**

17 hrs ago: The Moneyball strategy is the future for venture capital firms

2 days ago: Facebook's new regional fan breakdown outs brands that have been buying likes

4 days ago: Discover World's BEST business travel destinations



InformationWeek Healthcare

Software Security Cloud Mobility Social Business Big Data Windows Global CIO Government Healthcare Education Financial SMB 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.**

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.

# Integrating Data Across Cultures, Scales, Technologies Requires Infrastructure

4

- Common metadata standards
- Interoperability / integration framework
- Data access and preservation policy and practice
- Harmonized standards
- Common economic model for sustaining data
- Digital object identifiers
- Tools for data discoverability
- Etc. ...



*Harmonized standards*

*Policy and Practice*



# The Research Data Alliance

- Global community-driven organization launched in March 2013 to accelerate data-driven innovation
- 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
- **CREATE → ADOPT → USE:** RDA Working Groups operate for 12-18 months to build and use targeted pieces of data infrastructure



**RDA Interest Groups** are specific communities mapping out infrastructure that will enable data sharing and exchange, with concrete pieces of infrastructure then proposed as Working Groups

**RDA Working Groups** are “Tiger Teams” focusing on short-term deliverables:

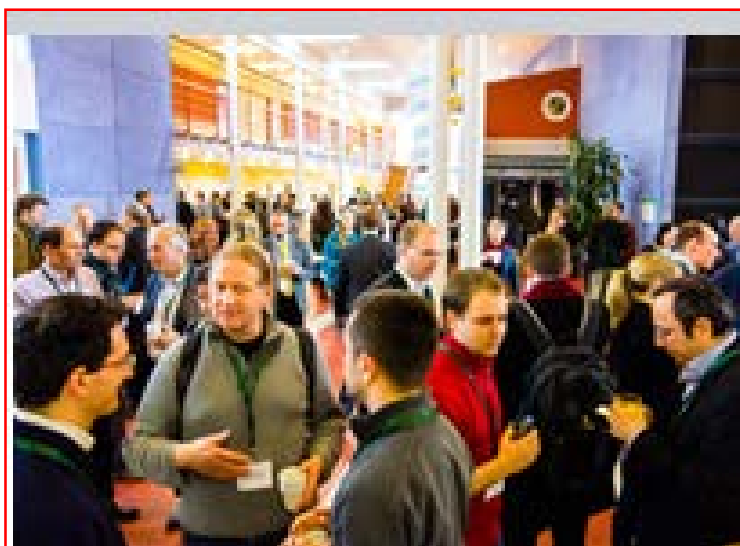
- **Focused pieces of adopted code, policy, infrastructure, standards, or best practices** that enable data to be shared and exchanged
- **“Harvestable” efforts** for which 12-18 months of work can eliminate a roadblock for a substantial community
- **Efforts that have substantive applicability** to “chunks” of the data community, but may not apply to everyone
- **Efforts for which working scientists and researchers can start today** while more long-term or far-reaching solutions are appropriately discussed in other venues

# RDA Launch March 18-20, 2013 Gothenburg, Sweden

- Over 200 participants
- 31 countries
- 5 continents
- > 6,400 tweets
- Public, private, academic sectors
- High-profile Govt. and Science speakers



7



# RDA Interest Groups focusing on infrastructure needed for data sharing and exchange

## RDA Interest Groups

- Legal Interoperability (joint with CODATA)
- Engagement
- UPC Code for Data
- Defining Urban Data Exchange for Science
- Marine Data Harmonization
- Repository Audit and Certification
- Preservation e-infrastructure
- Contextual Metadata
- Community Capability Model
- Big Data Analytics
- Agricultural Data, etc.



## Agricultural Data Interest Group

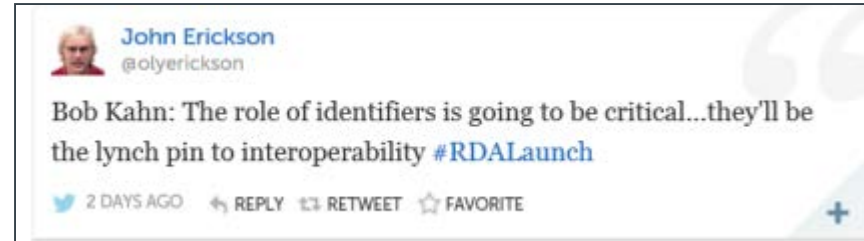
- *Focus:* Share and create interoperability framework for data from agricultural research, in particular wheat
  - *Wheat data combines Germplasm data, genetic and phenotypic data, statistical data, bibliographical data, etc.*
- Group members from India, France, Greece, UN Food and Agricultural Organization



# Create-Adopt-Use: RDA Working Groups Creating a Pipeline of Impact-focused Deliverables

## Current Working Groups and Groups Under Review

- PID Information Types
- Data Type Registries
- Data Foundation and Terminology (pending)
- Practical Policy (pending)



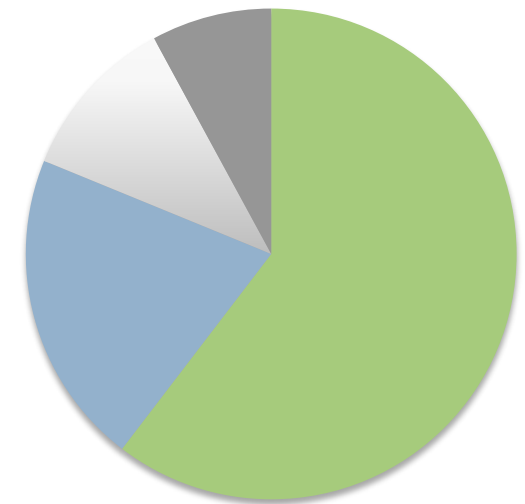
## RDA Working Group: Persistent Identifier (PID) Information Types

- **Group developing a framework for information types associated with unique identifiers; framework will be adopted at DKRZ (German supercomputing Center) and other sites**
  - *Agreement about the information associated with PIDs would allow programmers to implement the same API independent of the PID type being used*
  - *Provides a framework for other defined concepts and a prototype API for requesting PID information*

# Current Status (as of 4/13): RDA Community = ~700 participants from 44 countries

- Albania
- Australia
- Austria
- Bangladesh
- Belgium
- Bulgaria
- Brazil
- Canada
- China
- Congo
- Czech Republic
- Denmark
- Estonia
- Finland
- France
- Germany
- Greece
- Iceland
- India
- Iran
- Ireland
- Italy
- Japan
- Krygrystan
- Kuwait
- Netherlands
- New Zealand
- Norway
- Palestine
- Poland
- Portugal
- Russia
- Serbia
- Singapore
- South Africa
- South Korea
- Spain
- Sweden
- Switzerland
- Taiwan
- Turkey
- United Arab Emirates
- United Kingdom
- United States (31%)

## RDA by Sector



- Academics (61%)
- Private Sector (21%)
- Public Sector (11%)
- Unknown (8%)

# Why is the RDA Important to the US?

- **Internationally:** Leading role for US in RDA is strategically important, and increases global competitiveness
- **Nationally:** US has its own data challenges that must be resolved by the RDA/US data community
- *RDA provides a vehicle for implementing the underlying infrastructure required to make new policy approaches work.*



Office of Science and Technology Policy  
Executive Office of the President  
New Executive Office Building  
Washington, DC 20502

**FOR IMMEDIATE RELEASE**  
March 29, 2012

Contact: Rick Weiss 202 456-6037 [rweiss@ostp.eop.gov](mailto:rweiss@ostp.eop.gov)  
Lisa-Joy Zgorski 703 292-8311 [lisajoy@nsf.gov](mailto:lisajoy@nsf.gov)

**OBAMA ADMINISTRATION UNVEILS "BIG DATA" INITIATIVE:  
ANNOUNCES \$200 MILLION IN NEW R&D INVESTMENTS**

EXECUTIVE OFFICE OF THE PRESIDENT  
OFFICE OF SCIENCE AND TECHNOLOGY POLICY  
WASHINGTON, D.C. 20502

February 22, 2013

MEMORANDUM FOR THE HEADS OF EXECUTIVE DEPARTMENTS AND AGENCIES

FROM: John P. Helder, Director

SUBJECT: Increasing Access to the Results of Federally Funded Scientific Research

**I. Policy Principles**


The Administration is committed to ensuring that, to the greatest extent and with the fewest constraints possible and consistent with law and the objectives set out below, the direct results of federally funded scientific research are made available to and useful for the public, industry, and the scientific community. Such results include peer-reviewed publications and digital data.

Scientific research supported by the Federal Government is a key driver of progress in areas such as health, energy, and the environment. Access to digital data sets resulting from research resources and efforts on understanding and forecasting the future, as well as many biotechnology innovations, publications and scientific data in digital services related to custom, personalized, publications and data for re-use through the impact and accountability of the Federal Government's research and innovation, growth and job creation.

The Administration also recognizes that the coordination of peer review, that are essential scholarly publications. It is critical that the important that Federal policy not adversely affect the Federal Government to do so.

To achieve the Administration's commitment to research and digital scientific data, Federal agencies have clear and coordinated policies for its

of digital data, the Obama Research and Development Initiative." By rights from large and complex help solve some of the Nation's most



FUNDING AWARDS DISCOVERIES NEWS PUBLICATIONS STATISTICS ABOUT NSF PASTLAME

National Science Foundation  
Directorate for Engineering (ENG)

ENG HOME ENG FUNDING ENG AWARDS ENG DISCOVERIES ENG NEWS ABOUT ENG

Engineering  
design element

NSF Data Management Plan Requirements

Beginning January 18, 2011, proposals submitted to NSF must include a supplementary document of no more than two pages labeled "Data Management Plan" (DMP). This supplementary document should describe how the proposal will conform to NSF policy on the dissemination and sharing of research results. Proposals that do not include a DMP will not be able to be submitted. For more information about this new requirement, please see the [Grant Proposal Guide, Chapter II.C.2.i](#) and the [Data Management and Sharing Frequently Asked Questions \(FAQ\)](#).

Please note: the Engineering Directorate (ENG) has additional guidance for proposals submitted to ENG programs, [https://nsf.gov/eng/central/ENG\\_DMP\\_Policy.pdf](https://nsf.gov/eng/central/ENG_DMP_Policy.pdf). Questions and/or suggestions about this new requirement may be addressed to Dr. Maria K. Burka at [mburka@nsf.gov](mailto:mburka@nsf.gov).

# RDA Plenary 2: Opportunity for US Data Leadership

- **RDA Plenary 2: September 16-18 at National Academies in Washington DC**



## RDA Plenary is an opportunity

- To convene data-enabled organizations in a “neutral space” that facilitates the implementation and use of global infrastructure and global collaboration
- To highlight US efforts, contributions, and impact with a larger portion of the US data community
- For the RDA community to progress its work and expand its efforts



**Thank You**

**[enquiries@rd-alliance.org](mailto:enquiries@rd-alliance.org)**

**[plale@indiana.edu](mailto:plale@indiana.edu)**

**research data sharing without barriers**  
**[rd-alliance.org](http://rd-alliance.org)**

*Beth Plale*