



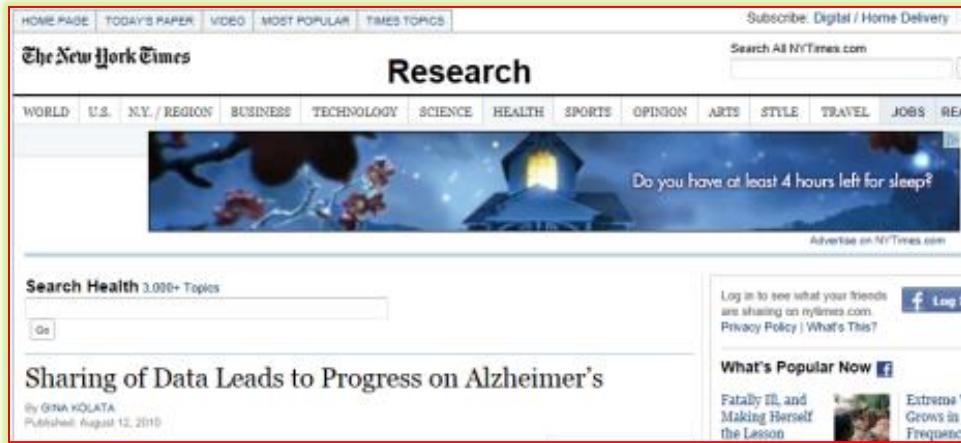
Update on the Research Data Alliance, etc.

Dr. Francine Berman

Chair, Research Data Alliance / US

Edward P. Hamilton Distinguished Professor of
Computer Science, Rensselaer Polytechnic
Institute

Data Sharing Driving New Discovery and Advances

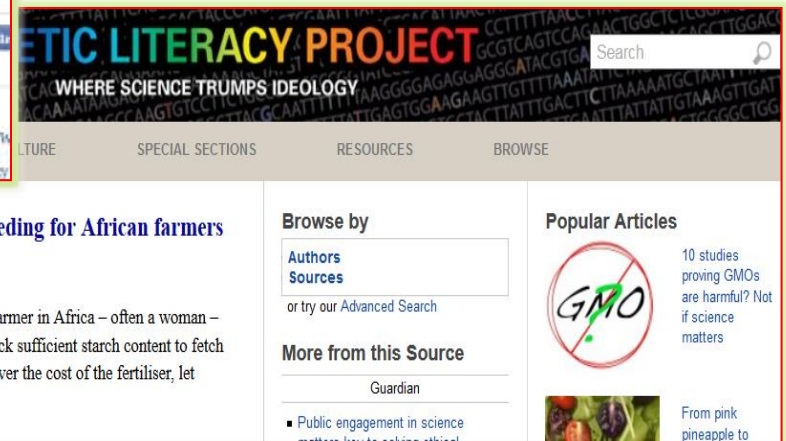


The screenshot shows the 'Research' section of The New York Times website. At the top, there are navigation links for 'HOME PAGE', 'TODAY'S PAPER', 'VIDEO', 'MOST POPULAR', and 'TIMES TOPICS'. The main header reads 'The New York Times Research' with a search bar. Below the header, there are category links: 'WORLD', 'U.S.', 'N.Y. / REGION', 'BUSINESS', 'TECHNOLOGY', 'SCIENCE', 'HEALTH', 'SPORTS', 'OPINION', 'ARTS', 'STYLE', 'TRAVEL', 'JOBS', and 'REAL ESTATE'. A featured article is titled 'Do you have at least 4 hours left for sleep?' with a background image of a house at night. Below this, there is a search bar for 'Search Health' with '3,000+ Topics'. A main article headline reads 'Sharing of Data Leads to Progress on Alzheimer's' by Gina Kolata, published August 12, 2010. To the right, there is a 'What's Popular Now' section with links to 'Fatally Ill and Making Herself the Lesson' and 'Extreme Weather Grows in Frequency'.

Sharing genomic data speeds cassava breeding for African farmers

Jean-Luc Jannink | April 3, 2015 | Guardian

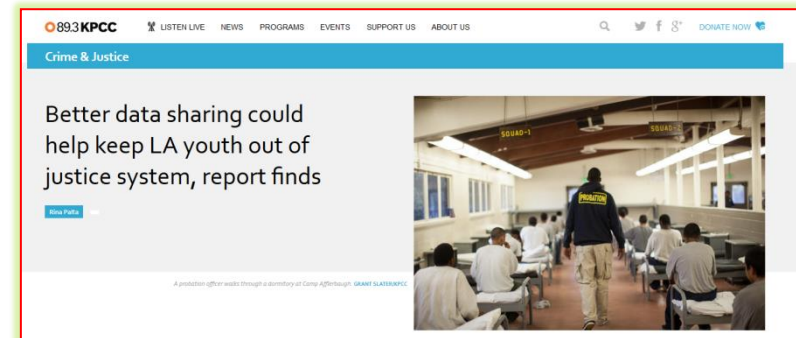
When the time comes to harvest cassava, a subsistence farmer in Africa – often a woman – only to be told they lack sufficient starch content to fetch a price that the farmer's offered doesn't cover the cost of the fertiliser, let alone the cost of the seedling.



The screenshot shows the 'Genetic Literacy Project' website. The main header features the text 'GENETIC LITERACY PROJECT' in large, colorful letters, with 'WHERE SCIENCE TRUMPS IDEOLOGY' below it. A search bar is located in the top right corner. Below the header, there are navigation links: 'LITERATURE', 'SPECIAL SECTIONS', 'RESOURCES', and 'BROWSE'. On the right side, there is a 'Browse by' section with a dropdown menu for 'Authors Sources' and a link to 'or try our Advanced Search'. Below this is a 'More from this Source' section with a link to 'Guardian' and a bullet point for 'Public engagement in science matters: from the science of the street to the science of the state'. To the right of this is a 'Popular Articles' section with a 'GMO' logo and the text '10 studies proving GMOs are harmful? Not if science matters'. Below the logo is an image of a pineapple with the text 'From pink pineapple to...'.



The screenshot shows the 'InformationWeek Healthcare' website. The main header features the 'InformationWeek Healthcare' logo and a search bar. Below the header, there are navigation links: 'Software', 'Security', 'Cloud', 'Mobility', 'Social Business', 'Big Data', 'Windows', 'Global CIO', 'Government', 'Healthcare', 'Education', 'Financial', 'SMB', and 'More'. Below the navigation links, there are several articles and advertisements. One article is titled 'Sharing Psychiatry EHR Data Cuts Readmission Rates'. Another article is titled 'Now you can afford them.' with a 'Watch video' button. There is also a 'Get InformationWeek Daily' section with the text 'Don't miss each day's hottest technology news, sent directly to your inbox, including occasional breaking news alerts'.



The screenshot shows the 'KPCC' website. The main header features the 'KPCC' logo and navigation links: 'LISTEN LIVE', 'NEWS', 'PROGRAMS', 'EVENTS', 'SUPPORT US', and 'ABOUT US'. Below the header, there is a 'Crime & Justice' section with the headline 'Better data sharing could help keep LA youth out of justice system, report finds'. Below the headline is a photo of a courtroom with the text 'A probation officer walks through a courtroom of Camp Afterschool, GRANT BLANKENHORN'. There is also a 'DONATE NOW' button in the top right corner.

Fran Berman, Research Data Alliance

Both Technical and Social Infrastructure Needed to support Data Sharing



Adopted Policy



Systems Interoperability



Common Types, Standards, Metadata



Sustainable Economics



Adopted Community Practice



Training, Education, Workforce

Prioritizing Infrastructure for Effort and Investment Challenging

*Stephanie A. Miner, the Syracuse mayor, said **[infrastructure is] too often overlooked** when politicians want to spend money on economic development. **“You don’t cut ribbons for new water mains, but that’s really what matters.”***

NY Times, February 15, 2014

The New York Times

Doctors Train to Spot Signs of A.D.H.D. in Children

Obama's Vote-Getting Tactics Struggle to Find the Uninsured

Obama Orders New Efficiency for Big Trucks

Public Defenders, Bolstered by a Work Analysis and Rulings, Push Back Against a Tide...

F.B.I. Joins Noose Is Le

Cost savings. Increased by the power of nysERDA

Get your Solutions Kit now.

U.S.

A Severe Winter Breaks Budgets as Well as Pipes

By JESSE MCKINLEY and RICHARD PÉREZ-PEÑA FEB. 15, 2014

SYRACUSE — Century-old water mains here have ruptured behind City Hall, popped in residential areas and split under the city’s bar and restaurant district. The mayor says she has personally reported three breaks, while exhausted crews work 18-hour shifts in

Plan, Design, Research, Build, Finance

Getting the World Involved in Building / Coordinating Data Sharing Infrastructure: the Research Data Alliance

- **Research Data Alliance (RDA):** Global **community-driven organization** whose mission is to build the **social and technical bridges (infrastructure)** that enable data sharing.
- **Research Data Alliance Vision:** *Researchers and innovators **openly share data across technologies, disciplines, and countries** to address the grand challenges of society.*



RDA: Accelerate Data Sharing and Interoperability Across Cultures, Communities, Scales, Technologies

■ **Technical parts of the data engine:**

- Data type registries reference model
- Wheat data interoperability framework



Common Types,
Standards, Metadata

■ **Rules of the road:**

- Common agreement on data citation
- Common practice for data repositories

Systems
Interoperability



Policy and Practice

■ **Better drivers**

- Summer schools in data science and cloud computing in the developing world (with CODATA)
- Data management plan development and monitoring



Sustainable
Economics



Training, Education,
Workforce

RDA Approach: Solve Problems and Facilitate Progress

RDA Members come together as

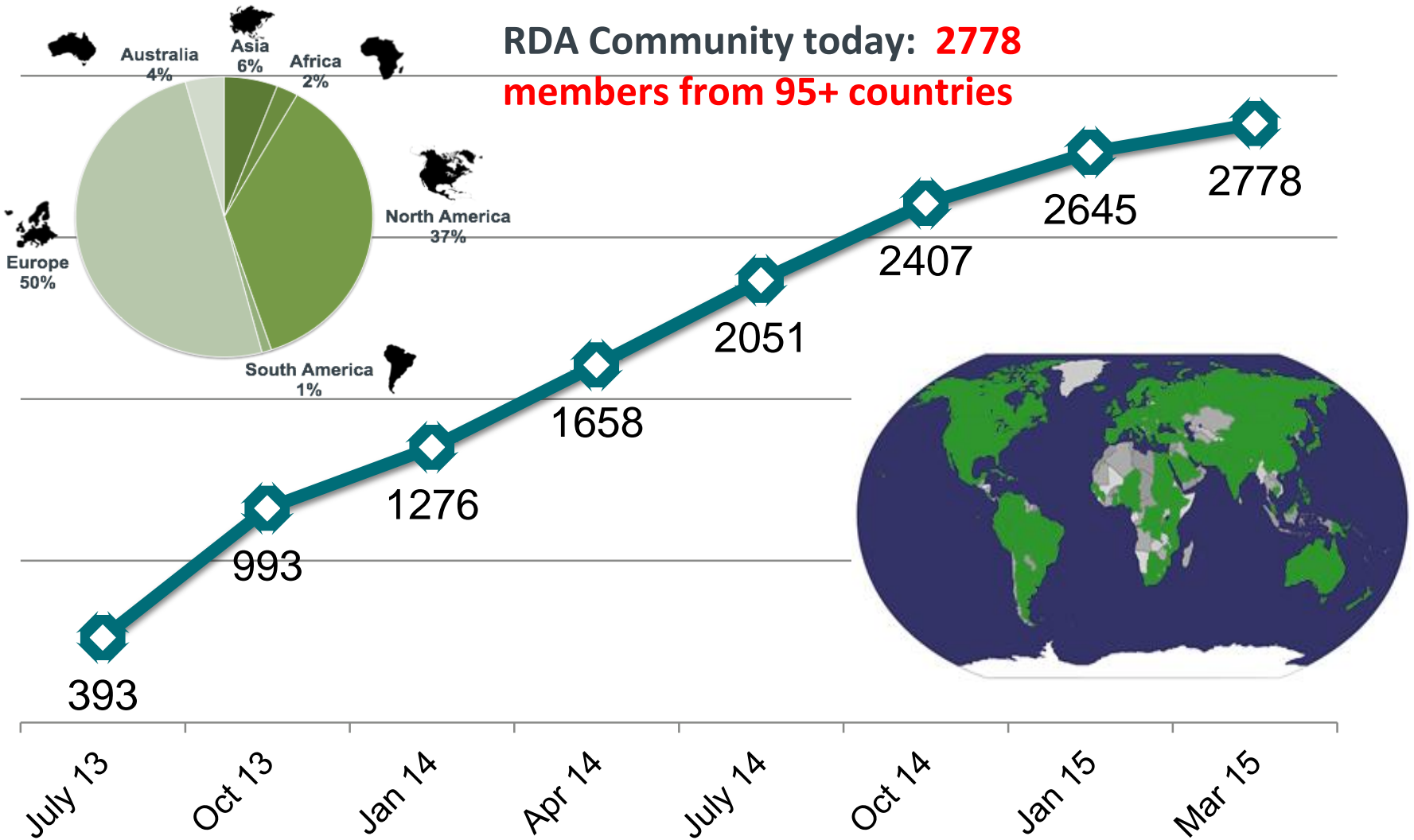
- **Working Groups** – 12-18 month efforts to **build, adopt, and use** specific pieces of infrastructure
- **Interest Groups** – longer-lived discussion forums that spawn Working Groups as specific pieces of needed infrastructure are identified.

RDA culture focuses on the pragmatic:

- **Working Groups must incorporate adopters** – no “build it and they will come”
- **Infrastructure must solve someone’s problem** but not necessarily everyone’s problems – not aiming for universal “esperanto” infrastructure
- **Amplify impact** when possible
 - community proactively enables **additional adopters** (communities, areas, organizations, projects that were not part of the original cohort) for whom RDA infrastructure work products are useful
 - RDA seeks to **collaborate with other organizations** to achieve their goals and strengthen the data community – RDA not looking for “world domination”

RDA Community @ 2: Precipitous Growth

RDA Community today: **2778**
members from **95+ countries**



RDA Community at Work: Interest Groups as of March 2015



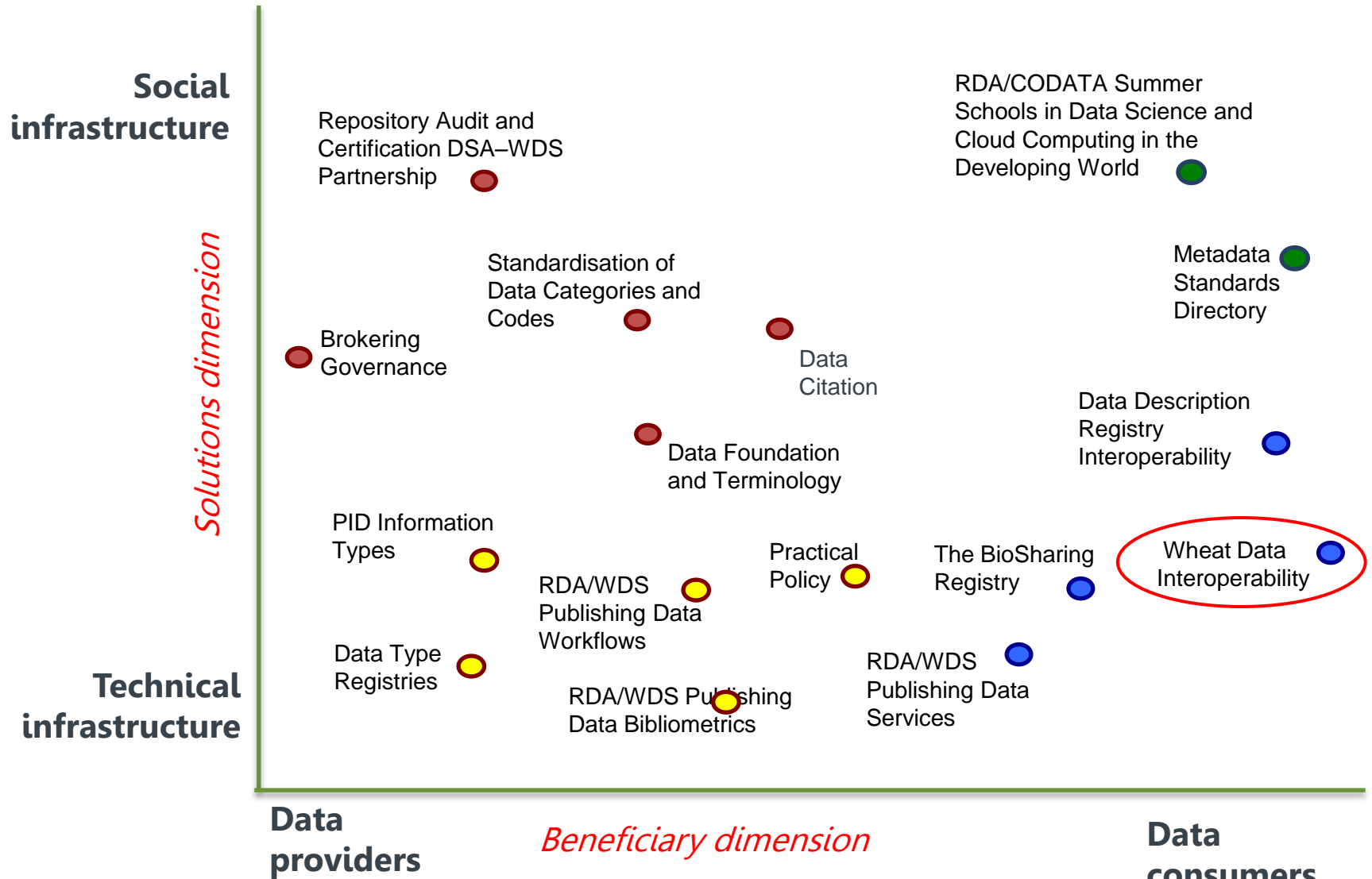
- | | | |
|---|---|---|
| 1. Agricultural Data Interoperability IG | 13. Domain Repositories Interest Group | Interest Group |
| 2. Active Data Management Plans* | 14. Education and Training on handling of research data | 27. RDA/CODATA Legal Interoperability IG |
| 3. Big Data Analytics IG | 15. ELIXIR Bridging Force IG | 28. RDA/CODATA Materials Data, Infrastructure & Interoperability IG |
| 4. Biodiversity Data Integration IG | 16. Engagement IG | 29. RDA/WDS Certification of Digital Repositories IG |
| 5. Brokering IG | 17. Federated Identity Management | 30. RDA/WDS Publishing Data Cost Recovery for Data Centres |
| 6. Community Capability Model IG | 18. Geospatial IG* | 31. RDA/WDS Publishing Data IG |
| 7. Data Fabric IG | 19. Libraries for Research Data | 32. Reproducibility IG |
| 8. Data for Development | 20. Long tail of research data IG | 33. Research data needs of the Photon and Neutron Science community |
| 9. Data Foundations and Terminology IG* | 21. Marine Data Harmonization IG | 34. Research Data Provenance |
| 10. Data in Context IG | 22. Metabolomics | 35. Service Management IG |
| 11. Development of cloud computing capacity and education in developing world research* | 23. Metadata IG | 36. Structural Biology IG |
| 12. Digital Practices in History and Ethnography IG | 24. PID Interest Group | 37. Toxicogenomics Interoperability IG |
- * in review

Domain Repositories Interest Group

(co-Chairs: George Alter/ICPSR, Peter Doorn/DANS, Ruth Duerr/NSIDC, Bob Hanisch/NIST + VAO)

- **Why:** Repositories critical for stewardship and preservation of research data. Common practice and policy can provide greater potential leverage and effectiveness. Exchange of ideas can improve user experience, economic sustainability.
- **What:** **RDA Domain Repositories Interest Group** brings together active data repositories serving many scientific disciplines. Discussions focus on sharing / creating good practice and collaborations around **data curation, dissemination, preservation** and **institutional sustainability**.
- **Value added:** RDA Domain Repositories Interest Group **working with other RDA groups** (data citation, metadata, certification of digital repositories) to adopt/amplify RDA infrastructure useful for their repositories
- **Impact:** Interest Group
 - helping build / strengthen individual repository organizations
 - creating community collaboration among repositories world-wide
 - developing a community that will improve stewardship options for domain researchers

RDA Working Groups Span a Broad Spectrum



Fran Berman, Research Data Alliance

Wheat Data Interoperability Working Group

co-Chairs: Esther Dzale Yeumo Kabore/French National Institute for Agricultural Research, Devika Madalli/Indian Statistical Institute, Johannes Keizer/Food and Agriculture Office of the UN

- **Why:** Wheat information systems needed to answer complex questions such as “*What genes and traits are relevant for understanding the impact of climate change on wheat plant productivity?*”. Diverse data on yield, market pricing, soil analysis, genomic and phenotypic information, etc. must be integrated / coordinated to address complex questions.
- **What:** **RDA Wheat Data Interoperability Working Group** developing a **common integration framework** for describing, representing, linking and publishing wheat data with respect to open standards to support wheat data sharing, use and re-use.
- **Work Products:** RDA Group will
 - Create common **standards and vocabularies** for wheat data management.
 - Create **framework** for Wheat Information System that integrates genomic annotations, phenotypes, genetic maps, physical maps, germplasm.
 - Facilitate access, discovery, use and re-use of Wheat Information System through **development / adoption of common metadata, vocabularies/ontologies/formats, good practice.**
- **Impact:** Working Group deliverables will be incorporated into the **Wheat Information System** of the Global Wheat Initiative and other international efforts, including the **Coherence in Information for Agricultural Research for Development (CIARD)** movement. Next steps: Framework will be adapted to other crops such as **Rice and Maize.**

Next Steps for RDA: Stay Pragmatic, Focus on Impact

More Infrastructure

Continuing pipeline of infrastructure deliverables adopted, used, coordinated and amplified to accelerate data sharing

More effective Community

Increasing coordination and collaboration between domains, sectors, organizations, communities. Effective advocacy for national and international data issues and communities.

Impact-focused Outreach

Stronger partnerships with industry, governments, domains, organizations.

Substantive engagement of students and early career professionals, greater spectrum of international cultures.

- **Next Plenaries** (Plenaries are both community and working meetings. Meetings held twice yearly around the world.):
 - September, 2015: **Paris, France (P6)**
 - March, 2016: **Tokyo, Japan (P7)**
 - September, 2016: ~ **Washington, DC (P8)**
 - March, 2017: **Barcelona, Spain (P9)**

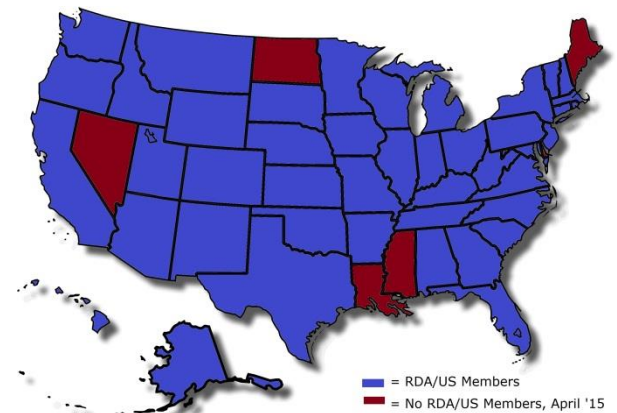
Joining RDA:

Go to rd-alliance.org and register

- Must agree to RDA principles (openness, community-driven, etc.)
- Free for individuals



RDA/US: Collaborate Globally, Contribute Locally



- **RDA/US = all U.S. members of RDA**
 - Currently ~1000 members of RDA in 46 states
- **RDA/US Mission:** To build RDA community in the U.S. and leverage RDA momentum to advance the U.S. data community
- **Current Activities:**

~~Community Development~~

- **Student / Early Career** programs (supported by **NSF, Sloan Foundation**)
- ~~Targeted outreach~~ to data-enabled communities and organizations (supported by **NSF**)

Community support:

- RDA Deliverables **Adoption Amplification seed projects** (supported by **NSF**)
- **International Plenaries participation support** for RDA/US (**NSF**)
- **Coordination meetings** for RDA Working Groups (supported by **NIST, NSF**)

Organizational Support

- **U.S. RDA Plenaries** hosting (supported by **NSF** and sponsors)
- **RDA/US development** – leadership and community building (supported by **NSF**)

RDA/US team:

Steering Committee

- Fran Berman, RPI -- Chair
- Larry Lannom, CNRI – Vice-Chair
- Beth Plale, IU – Vice-Chair
- Kathy Fontaine, RPI – *Managing Director*

Student Resident: Candice Lanius, RPI

Communications Manager: Yolanda Meleco

Administrative Coordinator: Jamie Lupo-Petta

NSF-supported RDA/US Student / Early Career Pilots



- **NSF RDA/US Fellows Pilot 2013-2015:**
 - Grad students and early career professionals
 - Attend Plenary N, work with an IG or WG, report on work at Plenary N+1
- **NSF RDA/US Interns Pilot 2013-2015 :**
 - Undergrad and grad students
 - Work with specific Interest or Working Group over the Summer, report on work at Fall Plenary
- **Purpose:**
 - Engage students and early career researchers in RDA
 - Link RDA efforts with other community efforts
 - Build/strengthen the professional network of Fellows and Interns
 - Build/strengthen the generational pipeline within the data community

Interns and Fellows Span Disciplines and RDA constituent communities

2014-2015 Pilot Participants

- **RDA/US Fellows**
 - 2014-2015 cohort from **Arizona Geological Survey, RPI, California Digital Library, UNC, Chapel Hill, Ronin Institute** for Independent Scholarship
- **RDA/US Interns**
 - 2014-2015 cohort from **University of Washington** School of Information, **Tufts / Perseus** Digital Library Project, **Indiana University** School of Informatics and Computing, **RPI** School of Humanities, Arts and Social Science, **Florida State University** School of Information, **UCSD** Department of Sociology, **UIUC** Department of Civil and Environmental Engineering, **Drexel** College of Computing and Informatics
 - Working with 7 WGs and IGs

RDA/US Resident – Humanities Ph.D. Student Candice Lanius part of RDA/US Headquarters Leadership Team



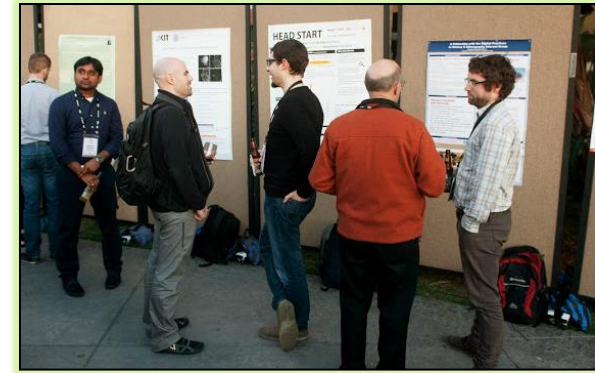
Projects:

- Data Architecture Vocabularies
- Research Data Provenance
- Digital Data in History and Ethnography
- Researcher Engagement in Data
- ORCID and Data Deposit
- Metadata Standards Directory
- Data Type Registries



Pilot → Program: RDA Data SHARE Fellows Program

- Funded by **Sloan Foundation** for 2015-2018. Program Leads: Beth Plale (PI), Kathy Fontaine, Inna Kouper
 - Program provides support for 12-18 months of effort.
 - Projects focus on evaluating, trial use, or improvement of products developed within a Working Group, developing and testing adoption strategies, or facilitation of interaction between RDA groups.
 - **Program enables Fellows to do substantive work and broaden / strengthen their networks and community.**
 - Program includes external evaluation to ensure quality and impact
- **Open to U.S. grad students, postdoctoral researchers, and early career researchers**



RDA/US Scholars Program

RESEARCH DATA ALLIANCE
UNITED STATES

Home Data Share Program **RDA/US Fellows** RDA/US Interns About RDA

The **Research Data Alliance (RDA)** invites applications for its newly redesigned fellowship program. The program's goal is to engage early career researchers in the US in Research Data Alliance (RDA), a dynamic and young global organization that seeks to eliminate the technical and social barriers to research data sharing.

The successful Fellow will engage in the RDA through a 12-18 month project with the bulk of the work to take place in the summer under the guidance of a mentor from the RDA community. The project is carried out within the context of an RDA Working Group (WG), Interest Group (IG), or Coordination Group (i.e., Technical Advisory Board), and is expected to have mutual benefit to both Fellow and the group's goals. Fellows receive a stipend and travel support and must be currently employed or appointed at a US institution.

Fellows have a chance to work on real-world challenges of high importance to RDA, for instance:

- Engage with social sciences experts to study the human and organizational barriers to technology sharing
- Apply a WG product to a need in the Fellow's discipline
- Develop plan and disseminate RDA research data sharing practices
- Develop and test adoption strategies
- Study and recommend strategies to facilitate adoption of outputs from WGs into the broader RDA membership and other organizations
- Engage with potential adopting organizations and study their practices and needs
- Develop outreach materials to disseminate information about RDA and its products
- Adapt and transfer outputs from WGs into the broader RDA membership and other organizations

The program involves carrying out a project over one or two summers and travel to RDA plenaries during the duration of the fellowship (international and domestic travel). Fellows will receive a \$5000 stipend for each summer of the fellowship. Fellows will be paired with a mentor from the RDA community. Through the RDA Data Share program, fellows will participate in a cohort building orientation workshop offering training in RDA and data sciences. This workshop is held at the beginning of the fellowship. RDA Data Share program coordinators will work with Fellows and mentors to clarify roles and responsibilities at the start of the fellowship.

Criteria for selection: The Fellows engaging in the RDA Data Share program are sought from a variety of backgrounds: communications, social, natural and physical sciences, business, informatics, and computer science. The RDA Data Share program will look for a T-shaped skill set, where early signs of cross discipline competency are combined with evidence of teamwork and communication skills, and a deep competency in one discipline.

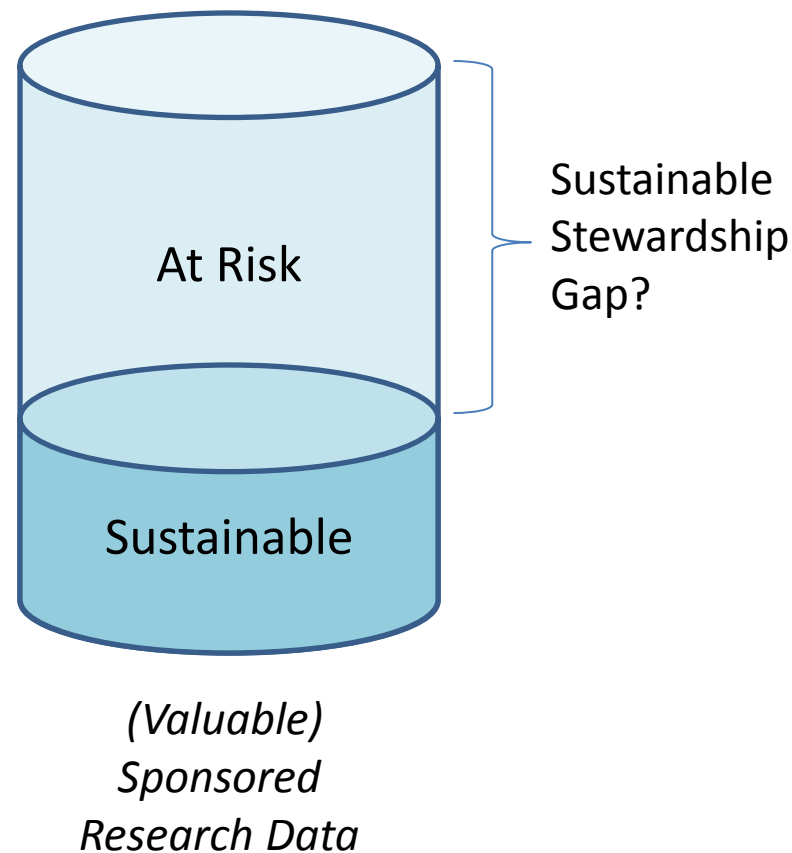
Additional criteria include: interest in and commitment to data sharing and open access; demonstrated ability to work in teams and within a limited time framework; and benefit to the applicant's career trajectory.

Eligibility: Graduate students and postdoctoral researchers at institutions of higher education in the United States, and early career researchers at U.S.-based research institutions who graduated with a relevant master's or PhD and are no more than three years beyond receipt of their degree. Applications from traditionally underserved populations are strongly encouraged to apply.

To apply: Interested candidates are invited to submit their resume/cumulative vitae and a 300-500 word statement that briefly describes their education, interests in data issues, and career goals to datashare-enquiry@list.indiana.edu. Candidates are encouraged to browse the RDA website <https://rd-alliance.org/> and pages of interest and working groups to identify relevant topics and mutual interests.

Etc. ... Sustainable Stewardship

- **Research data → innovation.**
 - Research increasingly expected to be available to the broader research community and general public *now* and *in the future*.
- **Preservation and stewardship of research data often ad hoc with much of it at risk**
 - *How much is sustainable?*
 - *What data is at risk?*
 - *What should we do about it?*
- **Lack of understanding about the sustainable stewardship gap hampers evidence-based discussion, prioritization and potential strategic investments.**



Sustainable Stewardship Pilot

- **Focus:** *Develop and pilot a comparative study of the amount and sustainability of research data generated from support from major sponsored public and non-profit organizations.*
- **Funding:** Alfred P. Sloan Foundation
- 2015-2016: **Pilot the study** on a representative cohort of data to determine:
 - **How much and what kind?:** What is the size and what are the characteristics of a research data baseline sample that might be sustainably preserved?
 - **How is it being cared for?:** What are the attributes of existing publicly accessible data stewardship organizations hosting the baseline sample; what are the most important attributes of successful stewardship organizations that can be replicated in the future?
 - **What are the policy and financial implications?:** What does the baseline sample tell us about the larger universe of research data? What strategies and investments are likely to be effective to provide and sustain stewardship for valued community data?

Pilot Approach

- Develop methodology / metrics by which to **measure and characterize the sustainable stewardship gap** – work with planning group to ensure:
 - Pilot survey represents information that can provide **compelling evidence needed to “move the stewardship and preservation needle”** for public, private, academic stakeholders and decision-makers
 - Pilot survey focuses on **information that can be obtained and used for analysis** and comparison
 - Pilot survey provides **useful results that can make an impact**
 - Pilot sets the stage for **larger “full-blown” sustainable stewardship study**

Who's Involved? [Planning Group]

- Myron Gutmann, U. of Colorado (PI, co-lead)
- Fran Berman, RPI (co-lead)
- Vint Cerf and Guha Ramanathan, Google
- Phil Bourne, NIH
- Cliff Lynch, Coalition for Networked Information
- George Alter, ICPSR
- Peter Fox, RPI
- Trisha Cruse, DataONE
- Sayeed Choudhury, Johns Hopkins University
- Margaret Hedstrom, U. of Michigan
- Chris Borgman, UCLA
- Brian Lavoie, OCLC
- Andy Maltz and Elizabeth Cohen, Science and Technology Council, Academy of Motion Picture Arts and Sciences
- John Gantz, IDC

Concrete steps

- **Sustainability:**

Please share your views on

- what information your organization needs to increase the priority of data stewardship and preservation
- What your community considers “valuable” data and how it could be characterized
- What level of sustainability and services your community needs to leverage the potential of data.

- **Data Infrastructure and Sharing:**

- Develop and implement a data management plan for your own data.
 - Make it available through a community repository
 - Cite your data in publications
 - Include the data bill in project budgets
- Work with colleagues in the RDA to amplify your efforts
 - Join or start an Interest Group
 - Join or start a Working Group
 - Involve your students in the RDA community
 - Help RDA/US create workshops that can positively impact community data sharing infrastructure

Thank you



Fran Berman, Research Data Alliance