



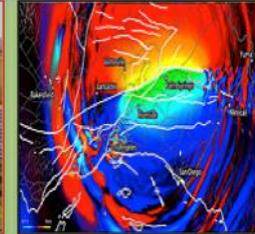
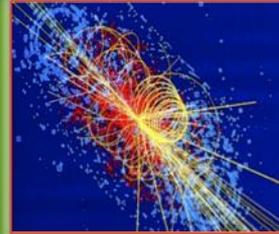
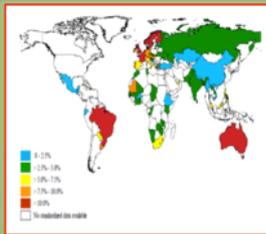
Update on the Research Data Alliance

Dr. Kathleen Fontaine
Managing Director, RDA / US

CENDI Meeting
Washington, DC
January 8, 2015

research data sharing without barriers
rd-alliance.org

Many Infrastructure Building Blocks Needed to Accelerate Research Progress



Research Dissemination and Reproducibility

Data Use and Re-use

Data Access (now) and Preservation (later)

Data Discovery and Data Sharing

Data Access and Distribution Policy

Institutional Data Sharing Practice

Data Discovery Tools

Digital Object Identifiers

Common Metadata Standards

Data Citation Standards

Data Preservation Practice

Data Analytics Algorithms

Data Scientists and Expert Support

Curation Practice and Policy

Auditing, Certification and Reporting Practice

Sustainable Economic Models

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

3

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, ...

Research Data Alliance

- Created to accelerate the development of research data sharing infrastructure worldwide
 - build social, organizational and technical infrastructure
 - reduce barriers to data sharing and exchange
 - accelerate the development of coordinated global data infrastructure
 - broad range of infrastructure pieces developed based on need



- **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 Medium-term (3-5 years) Strategic Goals⁶

- **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

A bit of history ...

First RDA organizational telecon: August 2012

Global Data Planning Meeting: October 2012



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 DAIF: 2011-2012

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

RDA was moving towards “steady state”. The RDA community had over 1300 members from 64+ countries as of March 2014.....

continued.....

A bit more history.....

First set of Working Group Deliverables initial release

WGs write up the final results for RDA review.

Adoption Day prior to March Plenary; highlights broader uses of WG outputs.

RDA Fourth Plenary

RDA Fifth Plenary

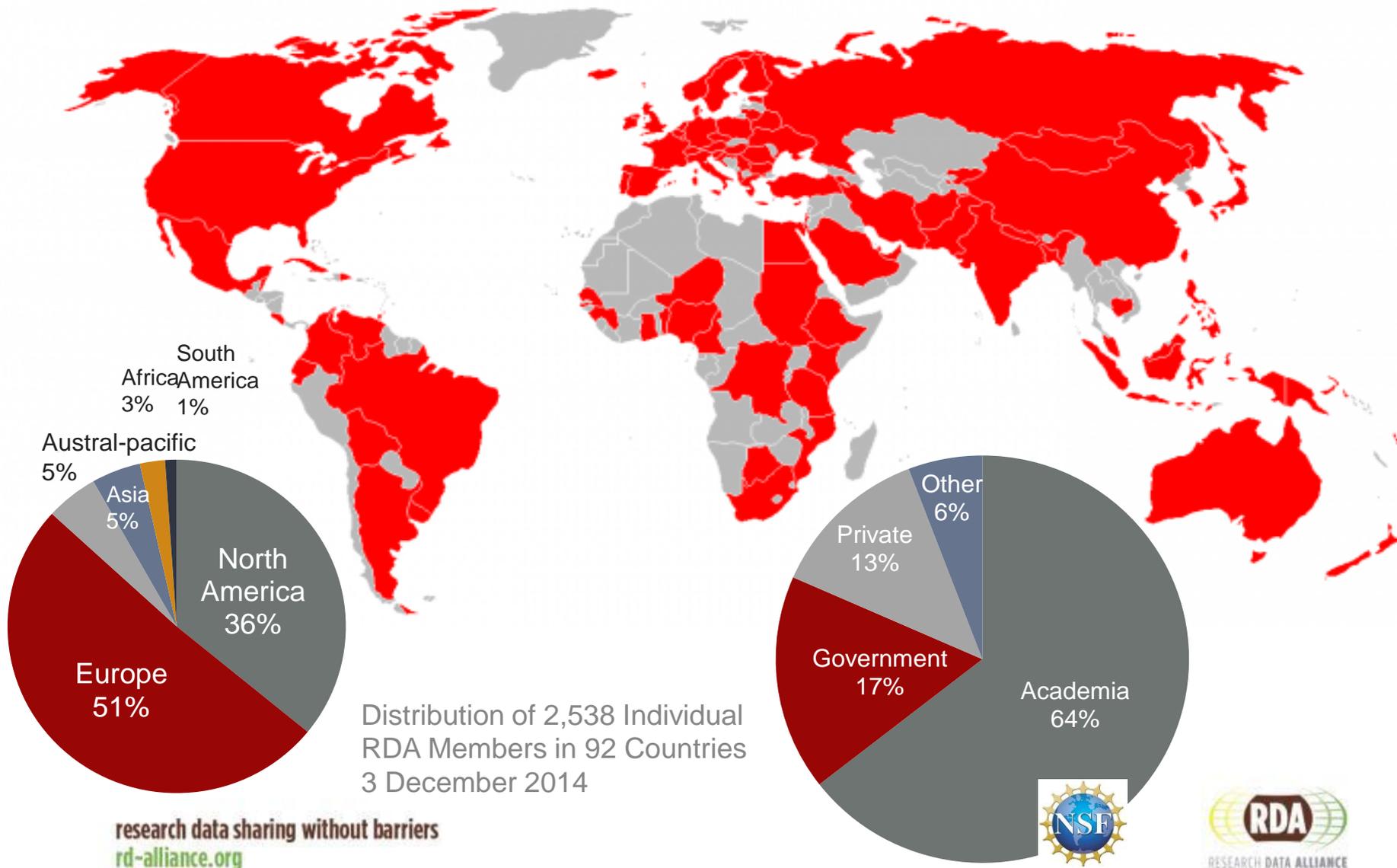
September 2014

March 2015

The RDA community now has over 2500 members from 92 countries as of December 2014.....

N
O
W

The Research Data Alliance Community



RDA Organizational Structure

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

■ **Organizational Members:**

- Alliance for Permanent Access
- American University Library
- Australian National Data Service
- Barcelona Supercomputing Center - Centro Nacional de Supercomputación
- Columbia University Library
- Corporation for National Research Initiatives (CNRI)
- Computer Sciences Corporation (CSC)
- Digital Curation Center
- EIROForum IT Working Group
- eResearch Services and Scholarly Application Development Division of Information Services, Griffith University
- European Data Infrastructure (EUDAT)
- National Institute of Advanced Industrial Science and Technology (AIST), Japan
- International Association of STM Publishers
- Internet2

- Microsoft Research
- NZ eScience Infrastructure
- Purdue University Libraries
- Research Data Canada
- Scholarly Publishing and Academic Resources Coalition (SPARC)
- Washington University in St. Louis Libraries
- Science and Technology Facilities Council

■ **Affiliates**

- Committee on Data for Science and Technology (CODATA)
- International Council of Scientific Unions (ICSU) World Data System
- Open Researcher and Contributor ID (ORCID)
- DataCite
- Consortia Advancing Standards in Research Administration Information (ASRAI)
- Global Alliance for Genomics and Health

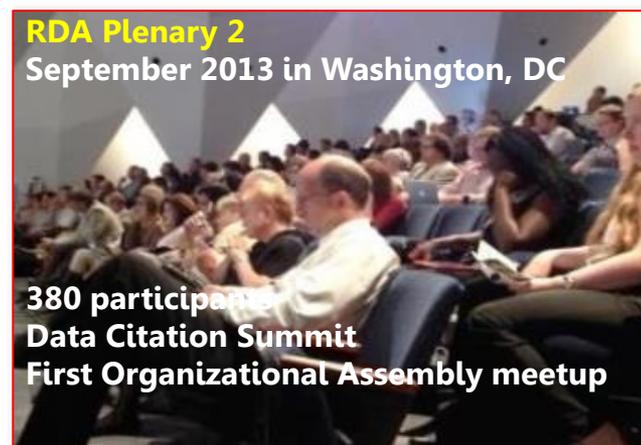
Approach: Create -> Adopt -> Use

- **Birds-of-a-Feather** – groups meeting at Plenaries to gauge broader interest in a topic
- **Interest Groups** – longer-lived discussion forums that spawn Working Groups as specific pieces of needed infrastructure are identified.
- **Working Groups** – 12-18 month efforts to build, adopt, and use specific pieces of 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**
- **Plenary meetings twice a year** – March and September to work and share

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)



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



RDA Working Groups

1. Brokering Governance WG
2. Data Citation WG
3. Data Description Registry Interoperability WG
4. Data Foundation and Terminology WG †
5. Data Type Registries WG †
6. Metadata Standards Directory Working Group WG
7. PID Information Types WG †
8. Practical Policy WG †
9. RDA/CODATA Summer Schools in Data Science and Cloud Computing in the Developing World WG
10. RDA/WDS Publishing Data Bibliometrics WG
11. RDA/WDS Publishing Data Services WG
12. RDA/WDS Publishing Data Workflows WG
13. Repository Audit and Certification DSA–WDS Partnership WG
14. Repository Platforms for Research Data WG*
15. Standardisation of Data Categories and Codes WG
16. The BioSharing Registry: connecting data policies, standards & databases in life sciences WG*
17. Wheat Data Interoperability WG



Deliverables Prepping for Review

Data Foundations and Terminology

Deliverables: Common vocabulary for data terms, formal definitions and open registry for data terms [http://smw-rda.esc.rzg.mpg.de/index.php/Main_Page]

Initial Adopters and Users: EUDAT, DKRZ, Deep Carbon Observatory, CLARIN, EPOS

Data Type Registries

Deliverables: System of data type registries, formal model for describing types, working model of a registry. [<http://typeregistry.org/registrar/#pages/Scope>]

Initial Adopters and Users: CNRI, International DOI Foundation, Deep Carbon Observatory

Practical Policy

Deliverables: Survey of policies in production use, test bed of machine actionable policies, deployment of 5 policy sets, policy starter kits [<https://www.rd-alliance.org/group/practical-policy-wg/outcomes/practical-policy-pp-output.html>]

Initial Adopters and Users: RENCI, DataNet Federation Consortium, CESNET, Odum Institute, EUDAT

Persistent Identifier Information Types

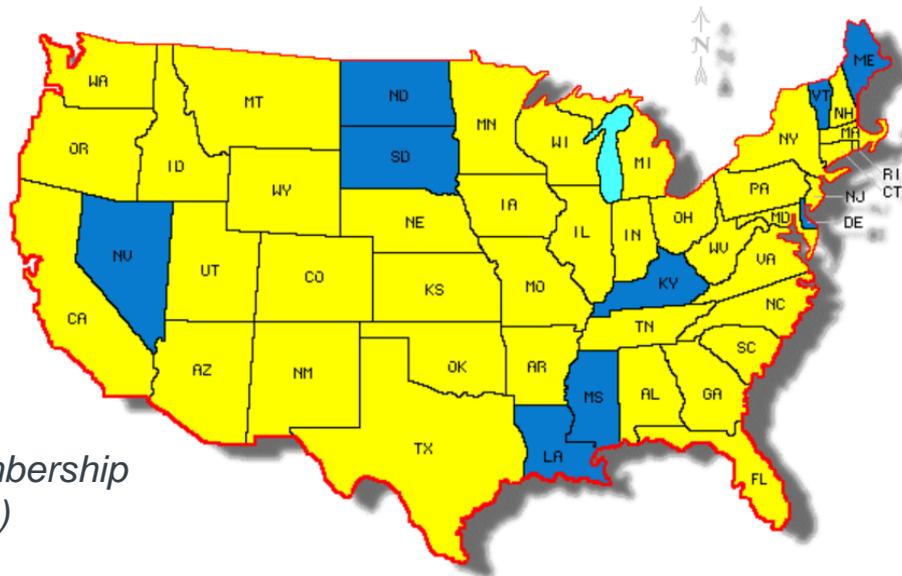
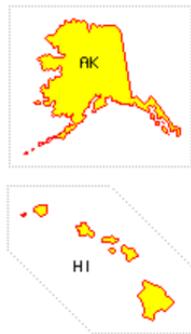
Deliverables: Minimal set of PID types, API [<https://www.rd-alliance.org/group/pid-information-types-wg/outcomes/pid-information-types-pit-output.html>]

Initial Adopters and Users: Data Conservancy, DKRZ

RDA/US – Contribute and Collaborate Globally and Locally

- RDA/US Goals
 - Leverage and implement RDA deliverables in the US to amplify impact
 - Build the RDA community within the US
 - Collaborate closely with other RDA “regions” on key programs and initiatives
 - Contribute to RDA “international” efforts and leadership
- Outreach
 - Over three dozen presentations and articles done for various groups
 - Panels, sessions, and conferences around special topics
- Deliverables Amplification
 - Five Working Groups presented their deliverables at Plenary 4
 - Path from output to adoption between now and March
 - Continual cycles of output and adoption as WGs mature
- Student/Early Career Engagement
 - First round of Fellowships and Internships awarded this summer
 - Awardees go to Plenary to present their work





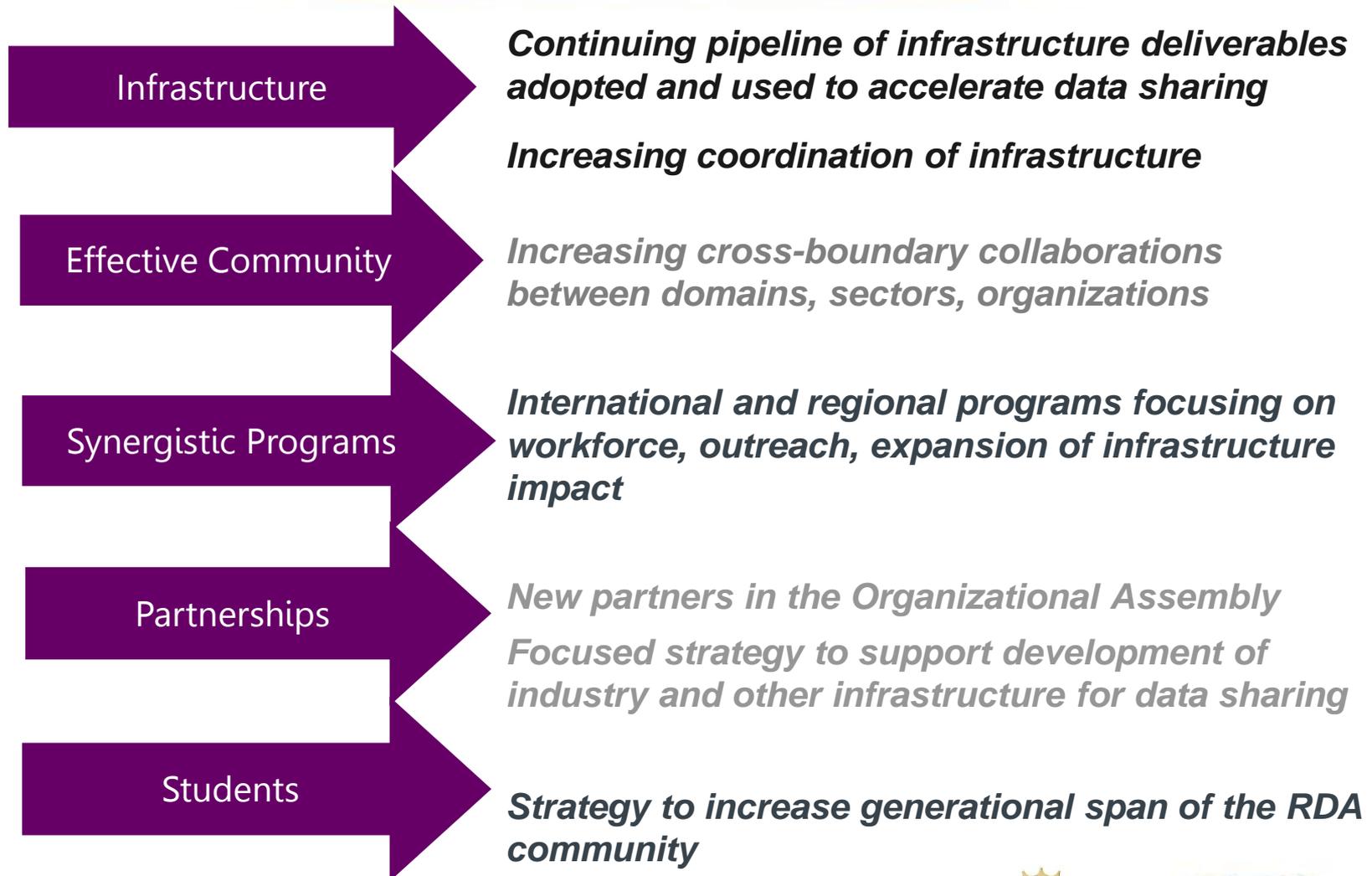
*RDA US membership
(yellow states)*

- NSF-supported RDA/US initiatives:
 - Outreach (RDA & RDA/US)
 - RDA Deliverables Amplification
 - Student / Early Career Engagement
- RDA/US Steering Committee
 - Fran Berman, RPI
 - Kathy Fontaine, RPI
 - Larry Lannom, CNRI
 - Beth Plale, IU

- **Paul Weiss, PhD student, Information School, University of Washington.** Project: *Data Architecture Vocabulary: Concept Analysis and Testing via semantic Media Wiki Tools*
- **Anna Krohn, master's student, Tufts University, Perseus Digital Library.** Project: *Research Data Provenance: Identifying Gaps and Overlap across RDA*
- **Isuru Suriarachchi, PhD student, School of Informatics and Computing, Indiana University Bloomington.** Project: *Research Data Provenance: Identifying Gaps and Overlap across RDA*
- **Brandon Costelloe-Kuehn, lecturer, Science and Technology Studies, Rensselaer Polytechnic Institute.** Project: *Digital Data and Infrastructure Practices in History and Ethnography*
- **Dong-Joon Lee, PhD student, School of Information, Florida State University.** Project: *Leveraging ORCID to Intermediate Data Deposit Among Distributed Repositories*
- **Rene Patnode, PhD student, Department of Sociology, UC San Diego.** Project: *Researcher Engagement in Data Exchanges: Needs and Practices within and across Scientific Domains*
- **Mostafa Elag, Postdoctoral researcher, Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign.** Project: *Data Type Registries: Populating the Registry*
- **Yuan Ling, PhD student, College of Computing and Informatics, Drexel University.** Project: *The RDA Metadata Standards Directory: Computer Actionable Policies Supporting Collaboration and Long-term Sustainability*

- **Rachael Black, Arizona Geological Survey** will investigate how and why individuals, or institutions, choose to become involved in the Research Data Alliance as well as how RDA members integrate what they learn from RDA into practice in their work.
- **Brandon Costelloe-Kuehn, Rensselaer Polytechnic Institute** will conduct an ethnographic study of data practices in many of the fields represented in the RDA. He will document, analyze, and work to reduce the socio-technical barriers to data sharing, examining a number of projects at the cutting-edge of digital research, interviewing project leads on: a) the contexts of these data sharing projects, b) the need for innovation in data practices and c) the design logic guiding new practices in digital scholarship.
- **John Kratz, California Digital Library**, will participate in the RDA/WDS Publishing Data and Long Tail of Research Data interest groups and the Publishing Services working group. He will contribute his perspective from bench biology, particularly with respect to data that does not fit neatly in a discipline-specific repository. He will also examine how RDA projects and products enable institutional repositories to meet the data publishing needs of bench scientists and identify unmet needs the RDA could productively address.
- **Sarah Ramdeen, University of North Carolina Chapel Hill**, will be working with the Preservation e-Infrastructure Interest Group. She will investigate what other groups are working on “content standard[s] of all the materials that needs to be preserved along with the data in order to make the data re-usable”. The PellIG has two goals: documenting which groups and disciplines are working in this area, and disseminating that information to the community. If time permits, Sarah would like to expand this to include identifying community needs related to best practices in provenance. Both for RDA as a whole and for specific communities within RDA.
- **Anne Thessen, Ronin Institute for Independent Scholarship**, will focus on developing standards for reporting information about species. She will use the data about a genus of dinoflagellates (a group of microbial eukaryotes) and explore effective ways to share it. At the end of the fellowship, she expects to have to have a good computable dataset and a data structure allows for querying members of this genus using morphological or ecological features.

Continuing Emphasis for the RDA: Regionally ↔ Globally



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



Thank You and Questions

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

