Research Data Alliance: Current Activities and Expected Impact

SGBD Workshop, May 2014
Herman Stehouwer
The Vision

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

RDA is building the social and technical bridges that enable open sharing of data. Researchers, scientists, data practitioners from around the world are invited to work together with us to achieve the RDA vision.
“Knowledge is the engine of our economy. And data is its fuel.”

(Neelie Kroes, Vice-President of the European Commission)

Data will become the key resource to tackle the big challenges in the world
RDA – Research Data Alliance

How to tackle big challenges if data is highly fragmented by disciplines, domains and countries?

**RDA** is **building bridges** to overcome the hurdles for easy data access, data **sharing** and **interoperability** by facilitating **collaboration** between experts from all over the world belonging to **different disciplines and organisations**.
RDA – How does it work?

Experts and Data practitioners come together in RDA Working and Interest Groups to overcome concrete hurdles.

They discuss matters at the Plenaries and on the RDA online platform.
May 2014 - The orchard needs to be managed

Applications draw up data from the ground like water

RDA is the trunk that brings the two together

Technologies make data discoverable and accessible
RDA Working Groups

- **Form the Foundation for RDA Community Impact!**

- **Working Groups envisioned as accelerants to data sharing practice and infrastructure in the short-term** with the overarching goal of advancing global data-driven discovery and innovation

- **RDA Working Group profile:**
  - **Short-term: 12-18 months**
  - **Focused efforts with specific actions adopted by specific communities**
  - **International participation**
  - **Open, voluntary, consensus-driven**
  - **Complementary to effective efforts elsewhere**

Potential outcomes / deliverables:
- New data standards or harmonization of existing standards.
- Greater data sharing, exchange, interoperability, usability and re-usability.
- Greater discoverability of research data sets.
- Better management, stewardship, and preservation of research data.
An **Interest Group** (IGs) can be established prior to a Working Group for community discussion of issues and areas that facilitate data-driven research.

IGs are longer-term groups defining common issues and interests.

WGs and IGs are collaborating intensively with groups in comparable initiatives such as IETF, CODATA, WDS, W3C.
9 Working groups including Community Capability Model, Data Citation, Data Foundation and Terminology, Data Type Registries...
https://rd-alliance.org/workinggroup-list.html

24 Interest groups including Agricultural Data Interoperability, Big Data Analytics...
https://rd-alliance.org/interestgroup-list.html

including joint groups with CODATA and WDS
Currently scientific communities are working on their own data organizations and even within communities there is a lot of heterogeneity -> different data models and terminology

• This WG is building a reference data model
  • Core for trust:
    ▪ Metadata
    ▪ PID
    ▪ Find, Access, Interpret, Re-use
  • Currently 21 approaches are being studied
    • Goal to come to a harmonized, reference model with a widely agreed terminology
Persistent identifiers (PID) are the core of proper data management and access.

This WG will provide a first solution for standardized PID types.

Later, the WG will design and implement an API for interaction with typed information.

Automated data management across disciplines and repositories can highly benefit from standardized types.
• There are so many data types in use, and new ones are continuously defined in science
  • The result is that often researchers see interesting data, but don’t know how to open, process or visualize the data
  • This WG is implementing a type registry for data, which explains how to open, visualize and process the data
  • New data types can be added to the register with information how to open, visualize and process the data
  • In september a worldwide setup for a type registry is expected from this work package
Registration of existing Metadata schemes

- Communities, initiatives need to make their choices explicit.
- Controlled vocabularies often are key to understand the categories.
- Communities can look for standards that come close to their needs.
- Ability to harvest other metadata, interpret and map it semantically.
- A first step to formalize the metadata domain to enable efficient machine processing of metadata.
Examples from WGs

**Practical Policy WG**

- Computer Actionable Policies
  - to identify a number of typical application scenarios for policies such as replication, preservation etc.
  - to collect exemplary practical policies for a first number of such application scenarios, register them, allow people to compare and re-use them and to extract options for commonalities and optimizations
  - to create awareness about ways to come to reproducible science, to come to trusted repositories and to allow proper certification
Expected Impact

- Five first groups
- Essential data life-cycle building blocks / bridges
<table>
<thead>
<tr>
<th>Community-Driven RDA Groups by Focus</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Domain Science - focused</strong></td>
</tr>
<tr>
<td>- Toxicogenomics Interoperability IG</td>
</tr>
<tr>
<td>- Structural Biology IG</td>
</tr>
<tr>
<td>- Biodiversity Data Integration IG</td>
</tr>
<tr>
<td>- Agricultural Data Interoperability IG</td>
</tr>
<tr>
<td>- Digital History and Ethnography IG</td>
</tr>
<tr>
<td>- Defining Urban Data Exchange for Science IG</td>
</tr>
<tr>
<td>- Marine Data Harmonization IG</td>
</tr>
<tr>
<td>- Materials Data Management IG</td>
</tr>
<tr>
<td><strong>Community Needs - focused</strong></td>
</tr>
<tr>
<td>- Community Capability Model IG</td>
</tr>
<tr>
<td>- Engagement IG</td>
</tr>
<tr>
<td>- Clouds in Developing Countries IG</td>
</tr>
<tr>
<td><strong>Reference and Sharing - focused</strong></td>
</tr>
<tr>
<td>- Data Citation IG</td>
</tr>
<tr>
<td>- Data Categories and Codes WG</td>
</tr>
<tr>
<td>- Legal Interoperability IG</td>
</tr>
<tr>
<td><strong>Data Stewardship - focused</strong></td>
</tr>
<tr>
<td>- Research Data Provenance IG</td>
</tr>
<tr>
<td>- Certification of Digital Repositories IG</td>
</tr>
<tr>
<td><strong>Community Needs - focused</strong></td>
</tr>
<tr>
<td>- Preservation e-infrastructure</td>
</tr>
<tr>
<td>- Long-tail of Research Data IG</td>
</tr>
<tr>
<td>- Publishing Data IG</td>
</tr>
<tr>
<td>- Domain Repositories IG</td>
</tr>
<tr>
<td>- Global Registry of Trusted Data Repositories and Services IG</td>
</tr>
<tr>
<td><strong>Base Infrastructure - focused</strong></td>
</tr>
<tr>
<td>- Data Foundations and Terminology WG</td>
</tr>
<tr>
<td>- Metadata Standards WG</td>
</tr>
<tr>
<td>- Practical Policy WG</td>
</tr>
<tr>
<td>- PID Information Types WG</td>
</tr>
<tr>
<td>- Data Type Registries WG</td>
</tr>
<tr>
<td><strong>Data Stewardship - focused</strong></td>
</tr>
<tr>
<td>- Metadata IG</td>
</tr>
<tr>
<td>- Big Data Analytics IG</td>
</tr>
<tr>
<td>- Data Brokering IG</td>
</tr>
</tbody>
</table>
RDA Plenary Meetings ...

**Plenary 1 – 18- 20 March 2013**
*Goteborg, Sweden*
- 240 participants
- 3 WG, 9 IG

**Plenary 2 - 18-20 September 2013**
in Washington, DC, USA
- 380 participants
- 6 WG, 17 IG, 5 BOF

**Plenary 3 - 26-28 March 2014**
in Dublin, Ireland
- 490 participants
- 16 WG, 35 IG and 20 BOF meetings
- 10 co-located workshops & meetings

**Plenary 4 - 22-24 September 2014**
*, Amsterdam, Netherlands*

- Working & interest groups get together and hold face-to-face discussions
- New groups proposals & Birds of a Feather
- RDA member networking
- Co-located events
How can you become a member?

- Register to the on-line community and become a Member of RDA.
- No fees involved for individual participation.
- Membership is open to any individual who subscribes to the RDA Guiding Principles.
- As a Member one may join and form Working and Interest Groups and participate in RDA elections.
- [https://www.rd-alliance.org/user/register](https://www.rd-alliance.org/user/register)
How can your Organisation become a member?

- Organisational Members can include R&D agencies, for-profit companies and non-profit foundations, community organizations, institutions, etc.
  - Annual membership fee based on size of organisation (in number of persons).

Why should your Organisation become a member?

- Affiliation with likeminded organisations in order to coordinate efforts in mutual areas of interest and to avoid unnecessary duplication and conflict.
What Organisations can you team up with?

**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
- 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

**Interested Affiliates**
- 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)
European Scientists feedback on RDA ….

Exchange and interaction with European Scientists on a wide range of topics including Sharing and Re-use of Data, Publishing and Citing Data, Infrastructures and Repositories

- A number of issues emerged as essential in helping to achieve the RDA vision:
  - Infrastructures must work and be persistent and sustainable, i.e. the work must be repeatable in a number of years.
  - Work must be reproducible.
  - Credit must be given for work and this credit must be useful in careers.
  - Data must be referable.
  - Infrastructures must be trustworthy and this trust must be earned.

...... for complete outputs see rd-alliance.org
Get Social with RDA

- **Twitter**
  @resdatall

- **Facebook**
  https://www.facebook.com/pages/Research-Data-Alliance/459608890798924

- **LinkedIn**
  www.linkedin.com/pub/research-data-alliance/77/115/7aa/

- **SlideShare**
  http://www.slideshare.net/ResearchDataAlliance
- RDA Collaborative Web Platform  rd-alliance.org
- Interaction with RDA  enquiries@rd-alliance.org
- RDA Europe  rda-europe@rd-alliance.org  |  europe.rd-alliance.org