RESEARCH DATA ALLIANCE
EUROPE

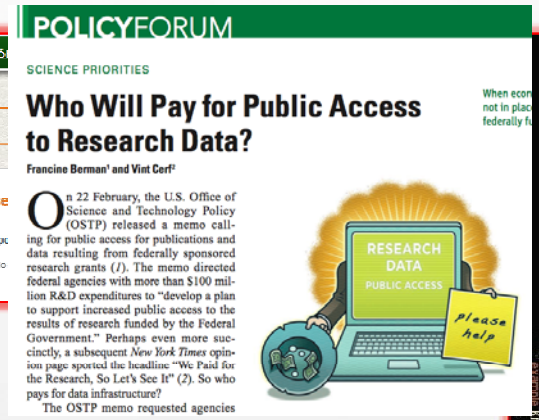
Finland and the global research data interoperability efforts

Leif Laaksonen / CSC – IT Center for Science

03.12.2013

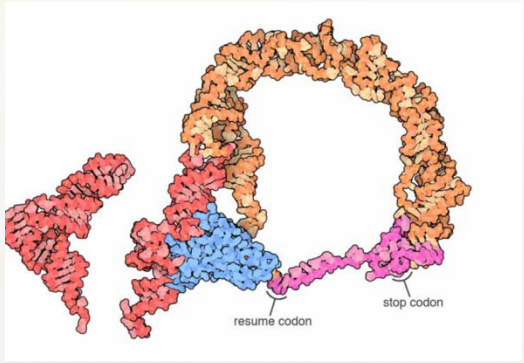
RESEARCH DATA ALLIANCE EUROPE

Digital Data has hit the "Tipping Point" as a Global Priority



Community coordination accelerates data-driven discovery and innovation

- *"Just do it"* -- Focused efforts help communities drive tangible progress



PROTEIN DATA BANK

Creation / adoption of data sharing policies have accelerated research innovation

Development of a publicly available shared data collection enabling new results for Alzheimer's

Development and adoption of parallel communication protocols through the MPI Forum drove a generation of advances

Transfer-Messenger RNA image (PDB Molecule of the Month for January 2013) courtesy of David Goodsell and the RCSB PDB

IETF®

Now 25 years old, the Internet Engineering Task Force's mission "to make the Internet work better" has resulted in key specifications of Internet standards that support innovation

MPI Forum image from http://lists.mpi-forum.org/mpi-forum/2008/12/0198.php

Promoting Access to Public Research Data for Scientific, Economic, and Social Development Data Science Journal 3 (2004) 135.

- **Technological issues**: Broad access to research data, and their optimum exploitation, requires appropriately designed technological infrastructure, broad international agreement on interoperability, and effective data quality controls;
- **Institutional and managerial issues**: While the core open access principle applies to all science communities, the diversity of the scientific enterprise suggests that a variety of institutional models and tailored data management approaches are most effective in meeting the needs of researchers;
- **Financial and budgetary issues**: Scientific data infrastructure requires continued, and dedicated, budgetary planning and appropriate financial support. The use of research data cannot be maximized if access, management, and preservation costs are an add-on or after-thought in research projects;
- **Legal and policy issues**: National laws and international agreements directly affect data access and sharing practices, despite the fact that they are often adopted without due consideration of the impact on the sharing of publicly funded research data;
- **Cultural and behavioral issues**: Appropriate reward structures are a necessary component for promoting data access and sharing practices. These apply to those who produce and those who manage research data.

Three complementing premises

- ***Data from publicly funded research are a public good produced in the public interest***
- ***Factual data are central to the scientific research process***
- ***Data access and sharing issues are international in scope***

Research Data policy in Finland

- Research and Innovation Council
- Ministry of Education and Culture
- Academy of Finland through the Finnish Research Infrastructure Committee (FIRI Committee)
- The TTA services and its working groups
- Council of Finnish Academies:
  - Finnish National Committee for Research Data Management based on their involvement in Alliance for Permanent Access (APA) and CODATA/World Data System (WDS).
- Universities, CSC, projects, (e-)infrastructure initiatives

International organisations dealing with standardisation, research data policy and education and training issues

- The Internet Engineering Task Force (IETF)
- The World Wide Web Consortium (W3C)
- DataCite (DataCite helps researchers to find, access, and reuse data)
- Committee on Data for Science and Technology (CODATA) and World Data System (WDS). Part of International Council for Science (ICSU)
- The Organisation for Economic Co-operation and Development (OECD)
- United Nations Educational, Scientific and Cultural Organization (UNESCO)
- Alliance for Permanent Access (APA/APARSEN)
- Research Data Alliance (RDA) complementing the domain
- A big number of national, continental and international research projects and infrastructures.
- And many more…

International Council for Science : Committee on Data for Science and Technology (CODATA)

- Advancing Informatics for Microbiology
- Anthropometric Data and Engineering
- Data at Risk
- Data Citation Standards and Practices
- Earth and Space Science Data Interoperability
- Exchangeable Materials Data Representation to Support Scientific Research and Education
- Fundamental Physical Constants
- Global Information Commons for Science Initiative
- Linked Open Data for Global Disaster Risk Research
- Octopus: Mining Space and Terrestrial Data for Improved Weather, Climate and Agriculture Predictions
- Global Roads Data Development
- Preservation of and Access to Scientific and Technical Data in/for/with Developing Countries (PASTD)

Forward from October 2010

N. Kroes EC: Data is currency of modern Science

Global attempts to improve data sharing and interoperability

European Collaborative Data Infrastructure with Common data Services

National Research Data Services

Combine grass-roots based thinking with broad engagement in implementing

Why introducing the Research Data Alliance?

- The Research Data Alliance (RDA) is a new community organization forming to facilitate **specific, short-term efforts that accelerate the sharing and exchange of research data promoting exchange between disciplines**
- Supported by the funders (Australian government, European Commission, USA NSF & NIST)
- **Working groups will serve as accelerants** to data sharing practice and infrastructure. Work products / deliverables to include
  - 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.
- Linkage to the GSO on Research Infrastructures (Data WG) in the so called “G8 extended configuration" to include "Outreach Countries (O6)" like India, China, Brazil, South Africa, Mexico and Australia.

RDA Guiding Principles

- **Openness** – Membership is open to all interested individuals who subscribe to the RDA's Guiding Principles. RDA community meetings and processes are open, and the deliverables of RDA Working Groups will be publicly disseminated.
- **Consensus** – The RDA moves forward by achieving consensus among its membership. RDA processes and procedures include appropriate mechanisms to resolve conflicts.
- **Balance** – The RDA seeks to promote balanced representation of its membership and stakeholder communities.
- **Harmonization** – The RDA works to achieve harmonization across data standards, policies, technologies, infrastructure, and communities.
- **Community-driven** – The RDA is a public, community-driven body constituted of volunteer members and organizations, supported by the RDA Secretariat.
- **Non-profit** - RDA does not promote, endorse, or sell commercial products, technologies, or services.

Political Support

- “To the greatest extent and with the fewest constraints possible publicly funded scientific research data should be open, while at the same time respecting concerns in relation to privacy, safety, security and commercial interests, whilst acknowledging the legitimate concerns of private partners.”
- “Open scientific research data should be easily discoverable, accessible, assessable, intelligible, useable, and wherever possible interoperable to specific quality standards”
  - From the science ministers of “G8 extended configuration" to include "Outreach Countries (O6)" like India, China, Brazil, South Africa, Mexico and Australia: https://www.gov.uk/government/news/g8-science-ministers-statement

RDA Working Groups

- ***Form the Foundation for RDA Community Impact!***
- ***An Interest Group can be established prior to a Working Group for community discussion of issues and areas that facilitate data-driven research. Many groups are forming at www.rd-alliance.org***
- ***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.**

Individual Member Participation

- Register to the on-line community, agree to the RDA Principles 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

Organisational Member Participation

- Organisational Members are approved by the RDA Council and 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).**
- Affiliation with likeminded organisations in order to coordinate efforts in mutual areas of interest and to avoid unnecessary duplication and conflict.
  - **There are no financial considerations on either side.**
  - **Memorandum of Understanding (MoU) requested**

Groups that met at the RDA Plenary 2 (2013)

BOLD = new since last Plenary

Birds-of-a-Feather

- **Linked Data**
- **Chemical Safety Data**
- **Education and Skills Development in Data Intensive Science**
- **Libraries and Research Data**
- **Cloud Computing and Data Analysis Training for the Developing World**

Working Groups

- Data Type Registries
- **Metadata Standards**
- **Practical Policy**
- Persistent Identifier Types
- Data Foundations and Terminology
- **Data Categories and Codes**

Interest Groups

- Agricultural Data
- **Big Data Analytics**
- **Data Brokering**
- Certification of Trusted Repositories (joint with ICSU-WDS)
- **Long tail of Research Data**
- Marine Data Harmonization
- Community Capability Model
- **Data Publishing (joint with WDS)**
- **Toxicogenomics Interoperability**
- **Research Data Provenance**
- Data Citation
- Metadata
- **Economic Models and Infrastructure for Federated Materials Data Management**
- Engagement
- Preservation e-Infrastructure
- Legal Interoperability (joint with CODATA)
- **Global Registry of Trusted Data Repositories and Services**
- **Digital Practices in History and Ethnography**

Data Citation Harmonization Summit

- DataCite,force11, CODATA/ICST, ESIP, DCC, etc.

RDA/WDS Interest Group on Certification of Trusted Digital Repositories

- Certification breeds trust
- Discussions on setting up a registry of trusted digital repositories
- Compare and join DSA and WDS certifications

Education and Skills Development in Data Intensive Science

- Proceed with formal process
  - Representatives from data analytics, e-infrastructure, libraries, data archives
- Involve stakeholders
  - LERU, LIBER, industry
- What are the basic skills for a data scientist?

Long-tail of Research Data

- Develop good practices for long-tail research
  - Understanding of long-tail data
  - Cost and funding models
  - Understanding willingness to deposit
  - Facilitating discovery

RDA/CODATA IG on Legal Interoperability

- Producing whitepaper
- Proposal for Workshop at NAS 2014
- Case studies
  - Best practices
  - Illustrative of legal frameworks
  - Inter-comparison of case studies

Current WGs and IGs

- It is the beginning and only a few have been fully set up but many topics are being discussed
- Working groups (9): Community Capability Model, Data Citation, Data Foundation and Terminology, Data Type Registries... https://rd-alliance.org/workinggroup-list.html
- Interest groups (18): Agricultural Data Interoperability, Big Data Analytics... https://rd-alliance.org/interestgroup-list.html

High-profile, high impact RDA Plenaries

- **RDA Plenary 1: March 18-20, 2013**
  - **at Gothenburg, Sweden**
- **RDA Plenary 2: September 16 - 18, 2013**
  - **at the National Academy of Sciences in Washington, DC, USA**
- **RDA Plenary 3: March 26 – 28, 2014**
  - **at the Croke Park Conference Centre Dublin, Ireland**
  - **https://www.rd-alliance.org/rda-third-plenary-meeting.html**

- **Agenda included**
  - Leading scientists and high-level international speakers
  - Birds of a Feather Sessions for RDA Interest Groups
  - All-hands sessions for technical, organizational issues, governance
  - Data Organization and International Data Panels, etc.
- ***RDA Plenary 4 in Amsterdam late September 2014***

- **Information:**
  rd-alliance.org
- **Questions, comments:**
  enquiries@rd-alliance.org
- **Registration and Forum:**
  forum.rd-alliance.org
- RDAEurope-info@postit.csc.fi | europe.rd-alliance.org