# The Research Data Alliance

Françoise Genova, CDS/CNRS, Strasbourg Astronomical Observatory
Co-chair of RDA Technical Advisory Board
Co-lead of RDA Europe French National Node

## THE RESEARCH DATA ALLIANCE

www.rd-alliance.org

building the social and technical bridges that enable open sharing of data

### 25 FLAGSHIP OUTPUTS

OF WHICH 4
ICT
TECHNICAL
SPECIFICATIO

### 75 ADOPTION CASES

ACROSS MULTIPLE DISCIPLINES, ORGANISATIO

## 93 GROUPS WORKING ON GLOBAL DATA INTEROPERABILITY CHALLENGES

of which 32 WORKING GROUPS & 61 INTEREST GROUPS

### 7,180 INDIVIDUAL MEMBERS FROM 137 COUNTRIES

67% Academia & Research 15% Public Administration 11% Enterprise & Industry

45 ORGANISATIONAL MEMBERS & 8 AFFILIATE MEMBERS



### Vision

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

### **Mission**

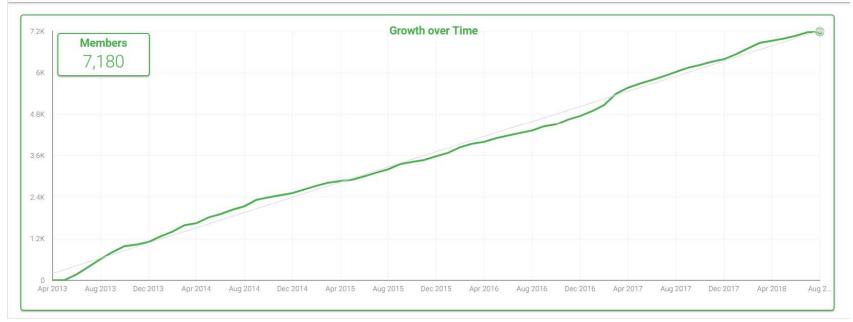
RDA builds the social and technical bridges that enable open sharing of data.







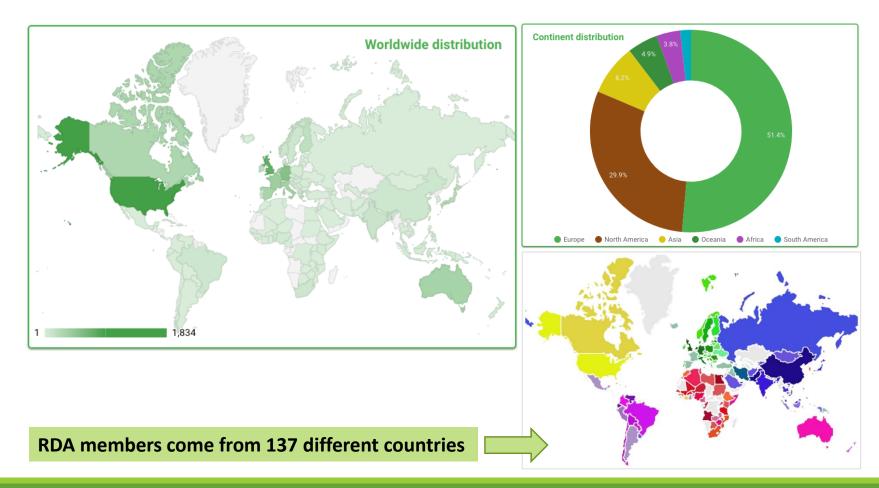
### **RDA Worldwide Growth**







### RDA Geographical Distribution





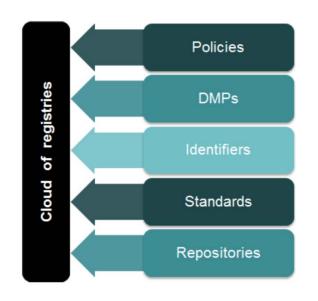
### The Research Data Alliance

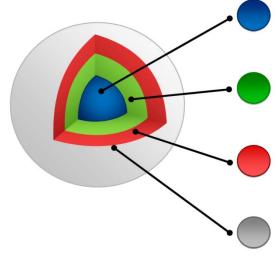
- Community-driven, international, consensus building, problem solving, neutral forum
- Different stakeholder profiles of research data sharing represented –
   The relevant community
  - Researchers
  - Data managers
  - Data stewards
  - Librarians
  - Policy makers & research funders
  - Private sector Publishers

## The RDA community

- Identifies open questions and work on them in the IGs and WGs
- Shares expertise and best practices
- Produces outcomes and formal recommendations, which are progressively recognized by the European Commission as ICT recommendations

## Very diverse open questions (EC FAIR EG)





#### DATA

#### The core bits

At its most basic level, data is a bitstream or binary sequence. For data to have meaning and to be FAIR, it needs to be represented in standard formats and be accompanied by Persistent Identifiers (PIDs), metadata and code. These layers of meaning enrich the data and enable reuse.

#### **IDENTIFIERS**

#### Persistent and unique (PIDs)

Data should be assigned a unique and persistent identifier such as a DOI or URN. This enables stable links to the object and supports citation and reuse to be tracked. Identifiers should also be applied to other related concepts such as the data authors (ORCIDs), projects (RAIDs), funders and associated research resources (RRIDs).

#### STANDARDS & CODE

#### Open, documented formats

Data should be represented in common and ideally open file formats. This enables others to reuse the data as the format is in widespread use and software is available to read the files. Open and well-documented formats are easier to preserve. Data also need to be accompanied by the code use to process and analyse the data.

#### METADATA

#### Contextual documentation

In order for data to be assessable and reusable, it should be accompanied by sufficient metadata and documentation. Basis metadata will enable data discovery, but much richer information and provenance is required to understand how, why, when and by whom the data were created. To enable the broadest reuse, data should be accompanied by a 'plurality of relevant attributes' and a clear and accessible data usage license.

Components of the FAIR data ecosystem

Model for FAIR data objects

(Interim report of the FAIR Data EC Expert Group)

## Very diverse open questions (EC FAIR EG)

- Research culture
- Technical ecosystem
- Skills and capacity building

### Plus

- Metrics
- Cost and investments
- Analysis, recommendations and action plan (interim report/many comments)

## A few examples of topics tacked by RDA

- Agriculture
  - Agriculture IG
  - Uses RDA to discuss interoperability in their field (involving FAO & GODAN)
    - Wheat Data Interoperability WG
    - Rice Data Interoperability WG
    - On-Farm Data Sharing WG
    - Agrisemantics WG
  - Capacity Development for Agriculture Data WG
- Materials Sciences (involves NIST)
  - International Materials Resource Registry WG Registry based on the IVOA/astronomy one

## A few examples of topics tackled by RDA

- Certification of data repositories
  - Merged Data Seal of Approval & World Data System, which initially addressed different communities – agreement on criteria and method > Core Trust Seal
- Data Citations, Data repositories, Data Management Plans, etc.
- Also Overarching Groups
  - Disciplinary Collaboration Frameworks
  - Data Fabric management of digital objects across the data life cycle

## Among RDA capacities

- Synchronise between the national, regional/European and international levels
- Attract and engage experts from a variety of research fields
- The place to organise interaction platforms among data professionals on a wide variety of topics
- The place to discuss disciplinary interoperability frameworks for fields which do not have one already
- The place to build synergies between solutions from different fields

- Lessons learnt
  - Sociological/cultural aspects are very important
  - Technological aspects should be driven by user needs
- Very broad questions