Status of the Research Data Alliance Possible liaisons with astronomy

Françoise Genova, CDS/RDA/RDA Europe

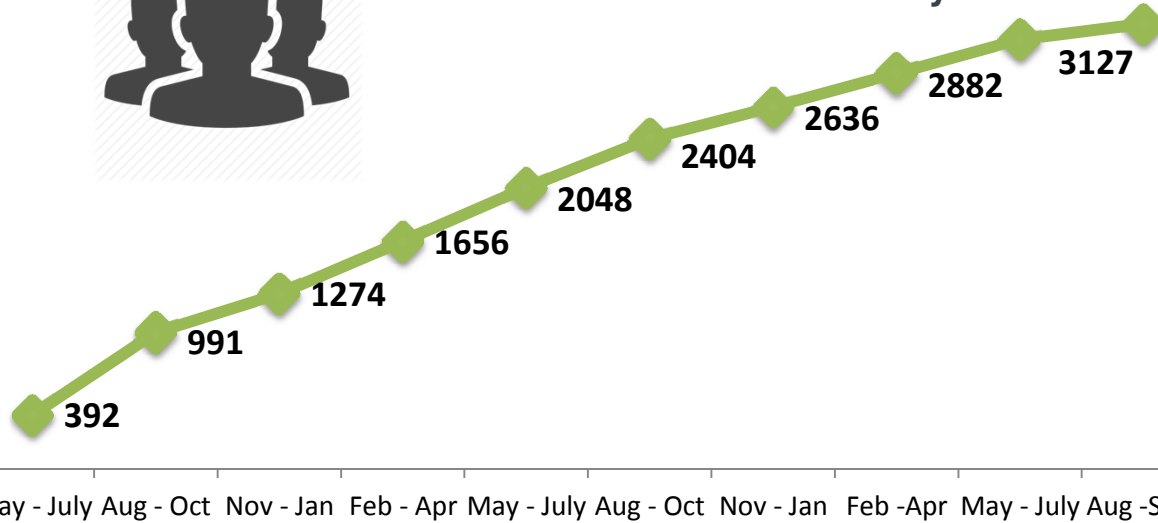research data sharing without barriers
rd-alliance.org

The RDA

- Facilitate scientific data sharing

- Created in March 2013 by the Australian Government, the European Commission, USA NSF and NIST

- A unique place to meet the international community which works on science data sharing (researchers, engineers, librarians)

- Fully bottom-up: activities are defined by the members

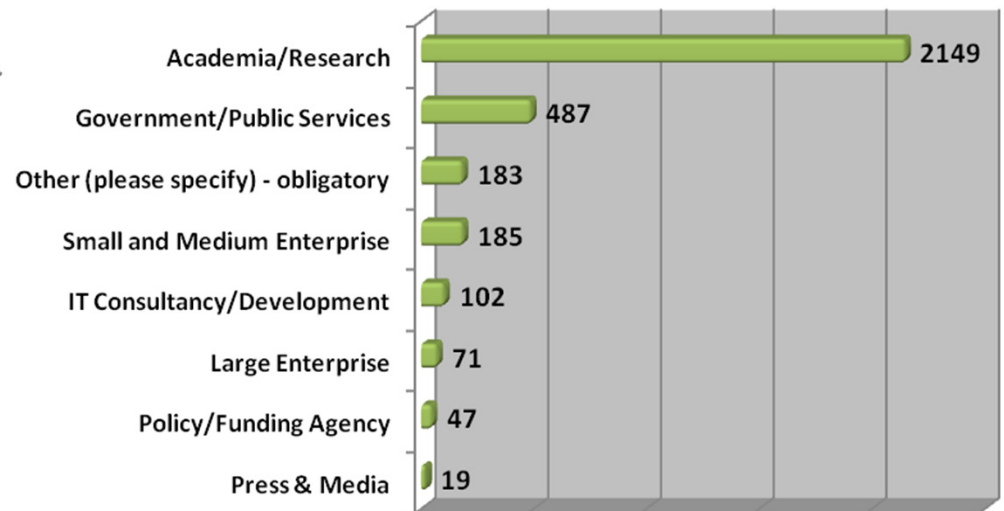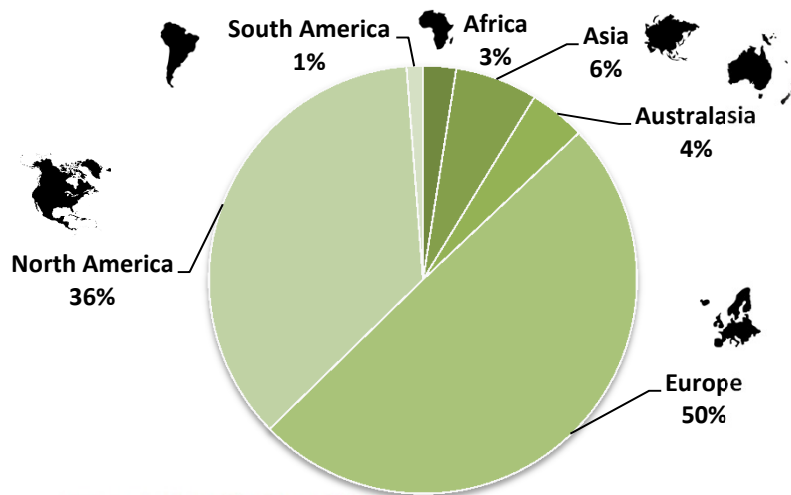- Working Groups (18 months, deliverables) & Interest Groups

research data sharing without barriers
rd-alliance.org

RDA
RESEARCH DATA ALLIANCE

The Research Data Alliance Community Today³

Total RDA Community Members: 3243

392 · 991 · 1274 · 1656 · 2048 · 2404 · 2636 · 2882 · 3127

May - July · Aug - Oct · Nov - Jan · Feb - Apr · May - July · Aug - Oct · Nov - Jan · Feb -Apr · May - July · Aug -Sept

from 103 countries

South America 1%
Africa 3%
Asia 6%
Australasia 4%
North America 36%
Europe 50%

Academia/Research — 2149
Government/Public Services — 487
Other (please specify) - obligatory — 183
Small and Medium Enterprise — 185
IT Consultancy/Development — 102
Large Enterprise — 71
Policy/Funding Agency — 47
Press & Media — 19

research data sharing without barriers
rd-alliance.org

RESEARCH DATA ALLIANCE

Growth…

- When it started: 5 WG
- Currently: 15 WG, 40 IGs

… Diversity!

- The first WGs were very technical
- Also « sociological » and « disciplinary » aspects

>>> Bottom-up! <<<

RDA
RESEARCH DATA ALLIANCE

RDA Interest (IG) and Working Groups (WG) by Focus 1

Domain Science - focused

- Toxicogenomics Interoperability IG
- Structural Biology IG
- Biodiversity Data Integration IG
- Agricultural Data Interoperability IG
- **Wheat Data Interoperability WG**
- Digital Practices in History and Ethnography IG
- Geospatial IG
- Marine Data Harmonization IG
- Metabolomics IG
- RDA/CODATA Materials Data Infrastructure and Interoperability IG
- Research Data Needs of the Photon and Neutron Science Community IG
- Defining Urban Data Exchange for Science IG
- **The BioSharing Registry: Connecting data policies, standards and databases in the life sciences WG**
- Urban Quality of Life Indicators IG

Community Needs - focused

- Community Capability Model IG
- Engagement IG
- RDA/CODATA Summer Schools in Data Science and Cloud Computing in the Developing World IG
- Development of Cloud Computing Capacity and Education in Developing World Research IG
- Data for Development IG
- Education and Training on handling of research data IG

research data sharing without barriers
rd-alliance.org

RDA
RESEARCH DATA ALLIANCE

RDA Interest (IG) and Working Groups (WG) by Focus 2

Reference and Sharing - focused

- **Data Citation WG**
- **Standardization of Data Cat. and Codes WG**
- RDA/CODATA Legal Interoperability IG
- Reproducibility IG
- **Data Description Registry Interoperability Working Group**
- **RDA / WDS Publishing Data Bibliometrics WG**

Data Stewardship and Services - focused

- Research Data Provenance IG
- Preservation e-infrastructure IG
- **RDA / WDS Publishing Data Services WG**
- **RDA / WDS Publishing Data Workflows WG**
- Long-tail of Research Data IG
- RDA/WDS Publishing Data IG
- **RDA/WDS Repository Audit and Certification**
- **DSA-WDS Partnership WG**
- Domain Repositories Interest Group
- Brokering Interest Group
- ELIXIR Bridging Force IG
- Libraries for Research Data IG*RDA / WDS Certification of Digital Repositories IG
- RDA / WDS Publishing Data Cost Recovery for Data Centres IG

Base Infrastructure - focused

- **Data Foundation and Terminology WG**
- **Metadata Standards Directory WG**
- **Practical Policy WG**
- **PID Information Types WG**
- **Data Type Registries WG**
- Data in Context IG
- Big Data Analytics IG
- **Brokering Governance WG**
- Federated Identity Management IG
- Metadata IG
- PID Interest Group
- Service Management IG
- Data Fabric IG

Diversity   *… but not chaos!*

- Bottom-up activities for grouping

- Meetings of the Group Chairs with the Technical Advisory Board at each Plenary – inspired by the IVOA!

- The TAB constantly works at understanding the activity landscape and suggesting liaisons between the Groups

research data sharing without barriers
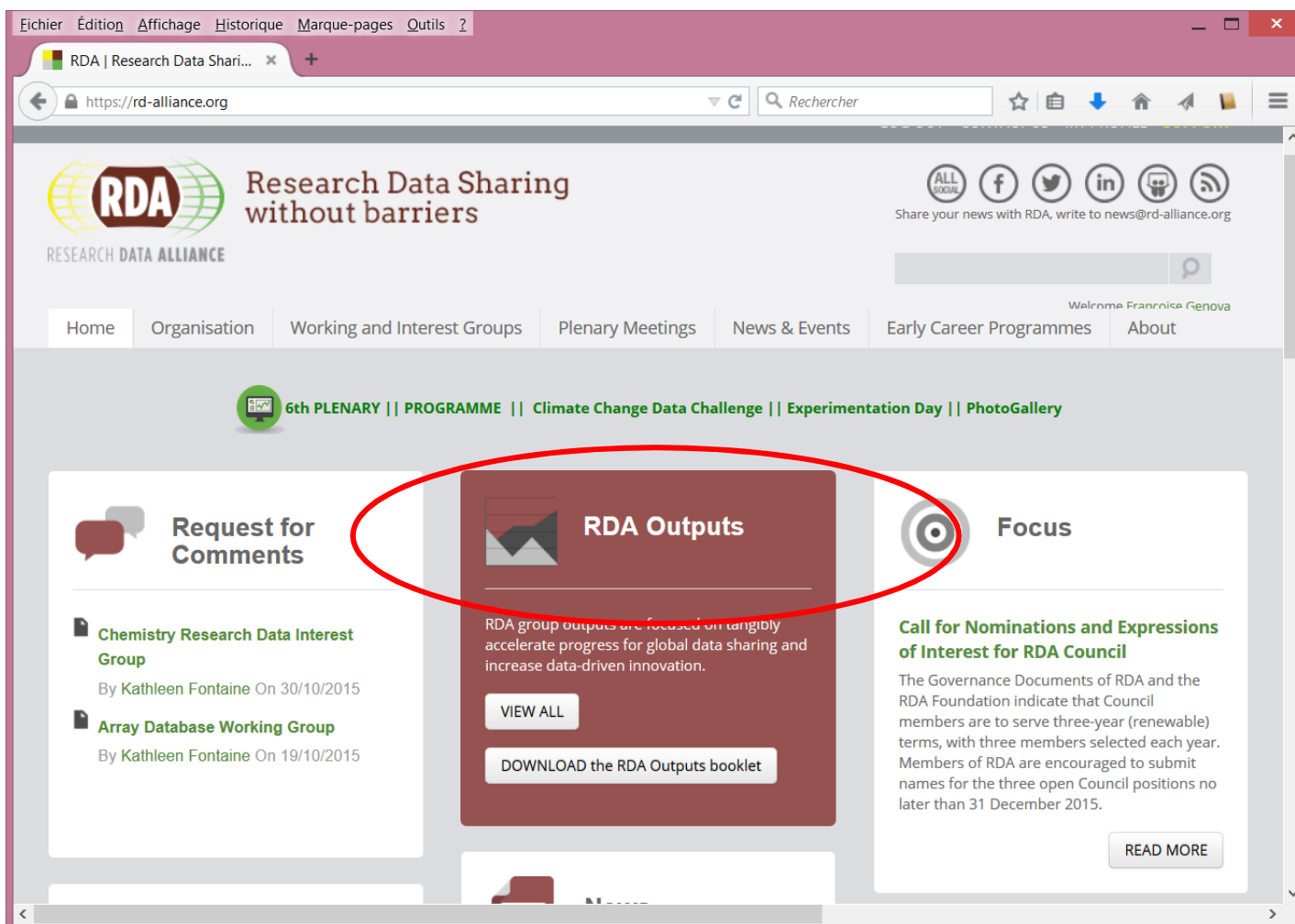rd-alliance.org

RDA
RESEARCH DATA ALLIANCE

DFIG – grouping of WG/IGs
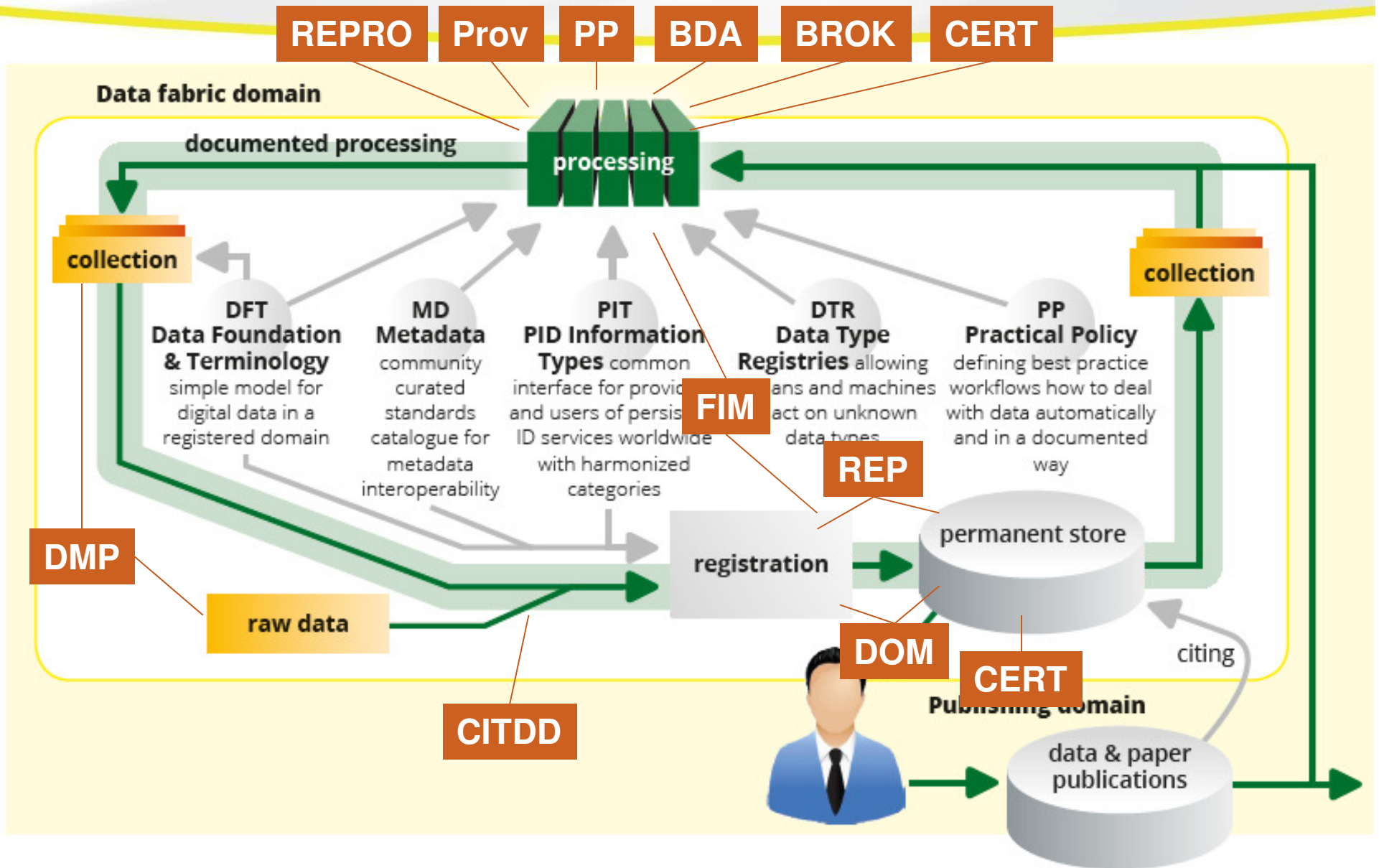
Publishing Cluster
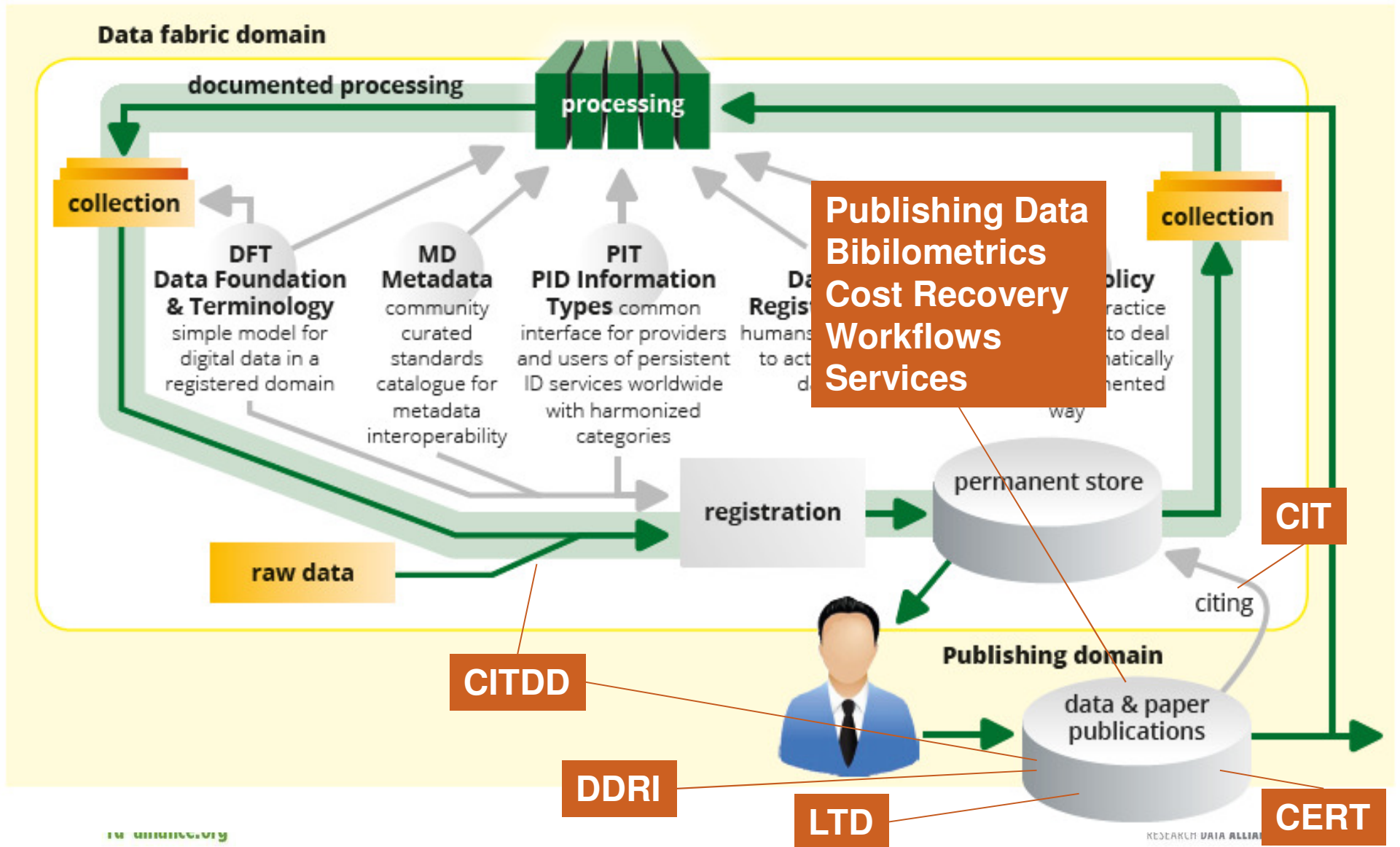
Lots of information on the web site

Diversity of interests and aims

- **Some disciplines use the RDA as a neutral forum to discuss the** <u>disciplinary aspects of interoperability</u>

- **Example: Agriculture**
  - Wheat Interoperability Working Group
  - Agriculture Interest Group

- **Astronomy: we have the IVOA and ADASS**

- **Many other topics of interest**

research data sharing without barriers
rd-alliance.org

A few examples

- Research Data Provenance
- Repository Audit and Certification DSA/WDS
- RDA/WDS WGs tackle data publication
- (Dynamic) Data Citation – VAMDC participate as testbed
- Active Data Management Plans
- Domain Repositories

RDA
RESEARCH DATA ALLIANCE

IG Research Data Provenance

WG Repository Audit & Certification DSA/WDS

- Preserving data in sustainable data centres is a key of scientific data sharing

- Trust is essential : trustworthy repositories

- Facilitate the work of the data centres (internal assessment/external evaluation)

- Common work of two basic certification frameworks, DSA and WDS, to define common criteria

European certification framework

- **Basic Certification** is granted to repositories which obtain DSA certification

- **Extended Certification** is granted to Basic Certification repositories which *in addition* perform a structured, externally reviewed and publicly available self-audit based on DIN 31644/nestorSeal

- **Formal Certification** is granted to repositories which *in addition to* Basic Certification obtain full external audit and certification based on ISO 16363

In conclusion

- Become a member to receive information
- Have a look at the Group activities and join those of interest for you
- Contact the chairs if you want to participate actively
- Think about creating new Groups – interested potential participants gathered through BoFs at the Plenaries
- The Plenaries are great also for informal interaction – lots of new ideas & good practices!

research data sharing without barriers
rd-alliance.org

RDA
RESEARCH DATA ALLIANCE

Additional viewgraphs

research data sharing without barriers
rd-alliance.org

- Certification
- RDA Outputs

DSA/WDS Certification WG: Common requirements

- 16 common criteria
- Each criterion comes with guidance

- Context
- Three topics addressed
    - Organisational infrastructure
    - Digital object management
    - Technology

Organisational infrastructure

- Mission/scope
- Licenses
- Continuity of access
- Confidentiality/Ethics
- Organisational infrastructure
- Expert guidance

Digital object management

- Data Integrity and authenticity
- Appraisal
- Documented storage procedure
- Preservation plan
- Data quality
- Workflows
- Data discovery and identification
- Data reuse

Common catalogue

- R0   Context

Please provide context for your repository

- R1   Mission/Scope   Organizational Infrastructure

The repository has an explicit mission to provide access to and preserve data in its domain

- R2   Licenses   Organizational Infrastructure

The repository maintains all applicable licenses covering data access and use and monitors compliance

RDA
RESEARCH DATA ALLIANCE

Common catalogue

- R3   Continuity of access   Organizational infrastructure

The repository has a continuity plan to ensure ongoing access to and preservation of its holdings

- R4   Confidentiality/ethics   Organizational Infrastructure

The repository ensures, to the extent possible, that data are created, curated, accessed, and used in compliance with disciplinary and ethical norms

- R5   Organizational infrastructure Organizational Infrastructure

The repository has adequate funding and sufficient numbers of qualified staff managed through a clear system of governance to effectively carry out the mission

Common catalogue

- R6   Expert guidance   Organizational Infrastructure

The repository adopts mechanism(s) to secure ongoing expert guidance and feedback (either in-house, or external, including scientific guidance, if relevant)

- R7   Data integrity and authenticity Digital Object Management

The repository guarantees the integrity and authenticity of the data

- R8   Appraisal   Digital Object Management

The repository accepts data and metadata based on defined criteria to ensure relevance and understandability for data users

RDA
RESEARCH DATA ALLIANCE

Common catalogue

- R9   Documented storage procedures Digital Object Management

The repository applies documented processes and procedures in managing archival storage of the data

- R10 Preservation plan   Digital Object Management

The repository assumes responsibility for long-term preservation and manages this function in a planned and documented way

- R11 Data quality   Digital Object Management

The repository has appropriate expertise to address technical data and metadata quality and ensures that sufficient information is available for end users to make quality-related evaluations

RDA
RESEARCH DATA ALLIANCE

Common Catalogue

- R12 Workflows   Digital Object Management

Archiving takes place according to defined workflows from ingest to dissemination

- R13 Data discovery and identification Digital Object Management

The repository enables users to discover the data and refer to them in a persistent way through proper citation

- R14 Data reuse   Digital Object Management

The repository enables reuse of the data over time, ensuring that appropriate metadata are available to support the understanding and use of the data

RDA
RESEARCH DATA ALLIANCE

Common Catalogue

- R15 Technical infrastructure   Technology

The repository functions on well-supported operating systems and other core infrastructural software and is using hardware and software technologies appropriate to the services it provides to its Designated Community

- R16 Security   Technology

The technical infrastructure of the repository provides for protection of the facility and its data, products, services, and users

RDA Outputs to date

- **Data Foundation & Terminology**: a model for data in the registered domain.

- **PID Information Types**: a common protocol for providers and users of persistent ID services worldwide.

- **Data Type Registries**: allowing humans and machines to act on unknown, but registered, data types.

- **Practical Policy**: defining best practices of how to deal with data automatically and in a documented way with computer actionable policy.