Sustainable FAIR Data management is challenging for RIs and it is challenging to solid Earth scientists

Massimo Cocco, Daniele Bailo, Keith Jeffery, Rossana Paciello, Valerio Vinciarelli and Carmela Freda

massimo.cocco@ingv.it
massimo.cocco@epos-eric.eu
The Data Generation Universe

Research Infrastructures (RIs) sample a limited portion of the Data Generation Universe, while ensuring:

- Quality Controlled data
- Standardized Data & Metadata
- Metadata Curation & Integration
- Data Curation and Integration
- Services curation and integration
- Access to data and products
- Visualization of integrated data
- Access to multidisciplinary data
- Generation of new scientific products
- Data qualification
- Service qualification
All Research Data

The data universe

Managed data

Open access data

FAIR Data

RIs represent a suitable environment to manage data

Sketch from S Venkataraman, DCC
RIs experiences corroborate that good data requires good management of the research data lifecycle

- **Data Management (plan)** requires the full control over the entire Data Life Cycle
- Data Management involves **Technical** and **Legal** aspects and it implies **Governance**
- This requires resources, thus involving a **Financial** dimension
- DLC starts with **data creation/collection**
- It is followed by **processing, visualization and analysis**
- Final stage is **publication/dissemination** of data and **open access**
- It is then necessary to have the **archiving and curation** in place, following the **FAIR principles**
- This requires **integration** and **interoperability**
RIs work on **data interoperability** since decades according to their architecture.
FAIR guiding principles for research data have further and better defined the FAIRness horizon.
FAIR DATA PRINCIPLES

Findable:
F1. (meta)data are assigned a globally unique and persistent identifier;
F2. data are described with rich metadata;
F3. metadata clearly and explicitly include the identifier of the data it describes;
F4. (meta)data are registered or indexed in a searchable resource;

Accessible:
A1. (meta)data are retrievable by their identifier using a standardized communications protocol;
A1.1 the protocol is open, free, and universally implementable;
A1.2. the protocol allows for an authentication and authorization procedure, where necessary;
A2. metadata are accessible, even when the data are no longer available;

Interoperable:
I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2. (meta)data use vocabularies that follow FAIR principles;
I3. (meta)data include qualified references to other (meta)data;

Reusable:
R1. metadata are richly described with a plurality of accurate and relevant attributes;
R1.1. (meta)data are released with a clear and accessible data usage license;
R1.2. (meta)data are associated with detailed provenance;
R1.3. (meta)data meet domain-relevant community standards;

https://www.force11.org/group/fairgroup/fairprinciples
Data Management and FAIR principles demands for FAIR Data Management Plans

- technical infrastructure and stewardship support
  - possible
- user experience
  - easy
- communities
  - normative
- incentives
- required
- policies

[Matt Spritzer, COS]
FAIR principles requires practices and solutions

**Maturity / FAIRness “roadmap”**

- **Data**
  - Information harmonisation
  - Formats harmonisation
  - Quality check

- **Metadata**
  - (RICH) Standard
  - Ontologies
  - vocabularies
  - Catalogue
  - Provenance

- **Interoperability**
  - Protocols
  - SOA / web services
  - PID
  - AAAI

- **Services**
  - Portals / Human computer interface
  - User defined Workflows
  - Visualisation
  - Virtual research environment

---

FAIR Data Management through sustainable practices

• The adoption of PRACTICES following FAIR data principles involves:
  ➢ Suitable **Technical** solutions for data integration and interoperability (Integrated metadata, shared metadata standards - 15 metadata formats to be standardized in DCAT-, SOA Interoperability services, data provenance ...)
  ➢ Effective **Legal** solutions (shared Data Policies, access rules, licensing of data - CC:BY - and metadata - CC:0)
  ➢ Effective **Governance** of integrated Data engaging data and service providers and IT experts (community building, awareness, trust)
  ➢ A **Financial** dimension to ensure sustainable management ensuring resources for data archiving, storage and long-term preservation, data qualification, access and management of persistent identifiers, ....

• Turning FAIR principles into practice requires:
  ✓ adequate human resources and skills
  ✓ Shared and adopted Data Management Plans
  ✓ ICT solutions and resources
  ✓ Adopting a suitable timeline, being aware that it does not necessarily correspond to the expectations of the authorities promoting FAIRness.
Conclusive Remarks

• FAIRness requires clear methodologies and technical activities in order to move from principles to reality
• FAIR Data Management needs active participation of data producers and practitioners
• FAIR Data Management needs a suitable timeline involving all interested stakeholders
• The sustainability of FAIR DATA Management is still a challenge for Research Infrastructures (RIs)
• FAIR DATA Management is even more challenging to Earth scientists
Open Science & FAIR Data require Resources

Thank You
For your attention
Thank You