

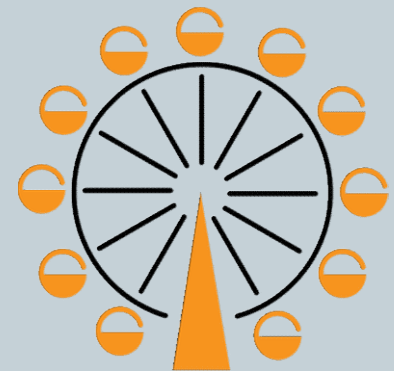
FAIR data: what it means, how we achieve it, and the role of RDA



Sarah Jones
Associate Director, Digital Curation Centre
Rapporteur of FAIR data expert group

sarah.jones@glasgow.ac.uk

Twitter: @sjDCC



What is FAIR?



A set of principles that describe the attributes data need to have to enable and enhance reuse, by humans and machines

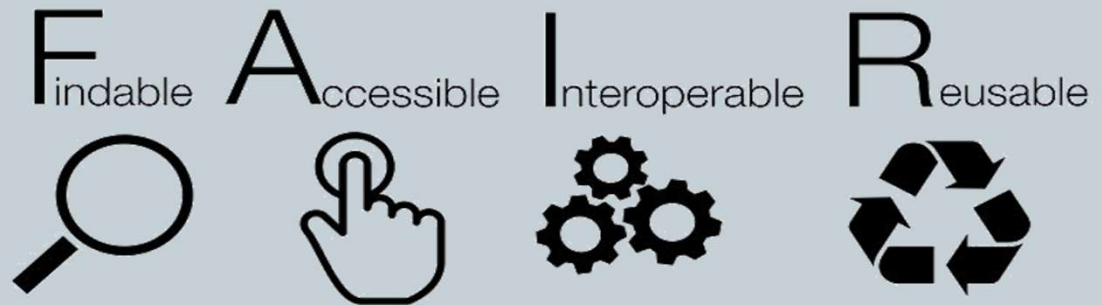


Image CC-BY-SA by [SangyaPundir](#)

Origins of FAIR



- Emerged from a workshop held in Leiden in 2014
- Come from life sciences but intended for all data
- Issued by FORCE11 community
- Echo previous principles on open data & curation



[OECD Principles and Guidelines for Access to Research Data from Public Funding](#) (2007)

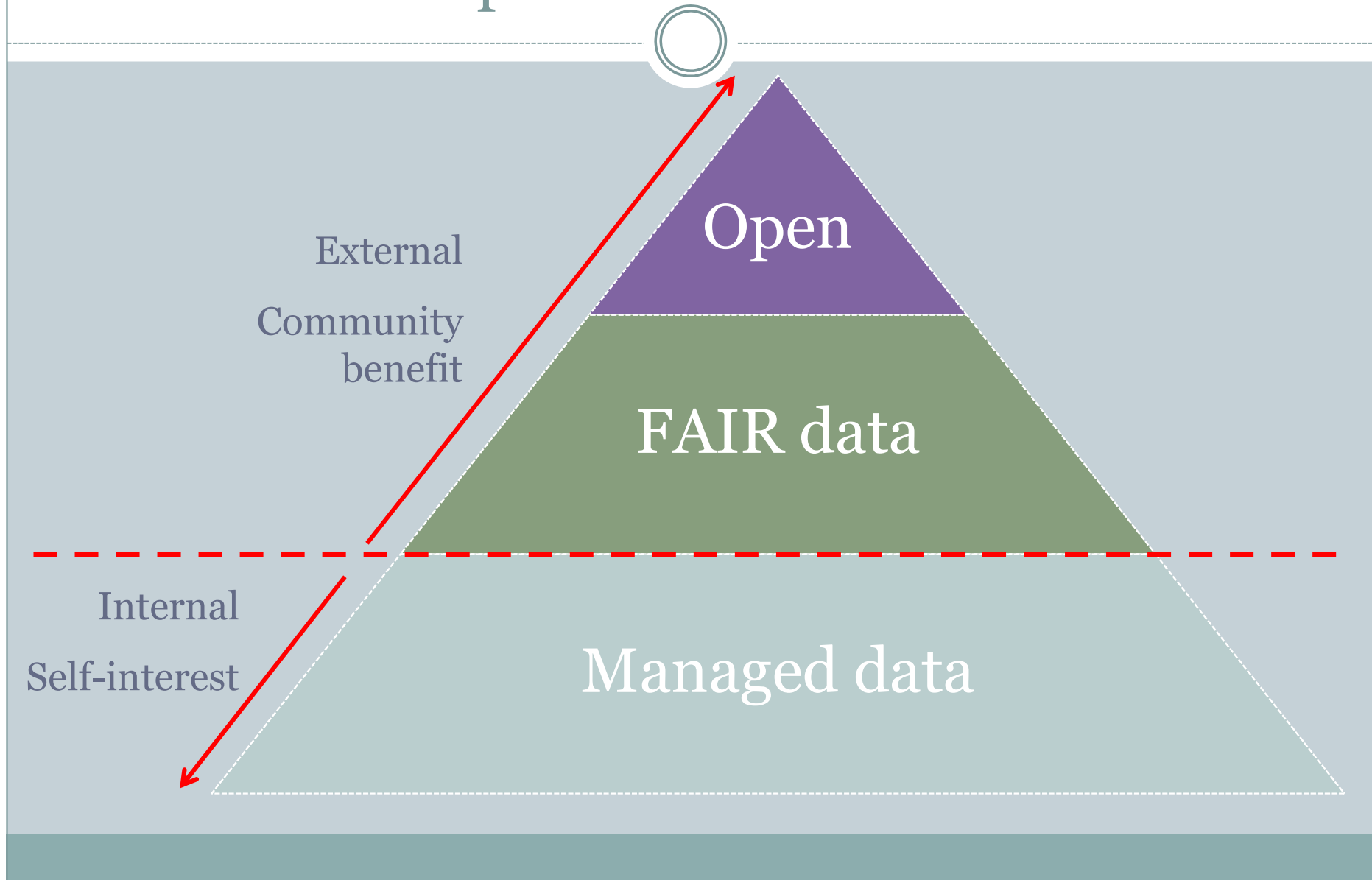


[Science as an Open Enterprise](#) (2012)
notion of '*intelligent openness*' where data are accessible, intelligible, assessable and useable

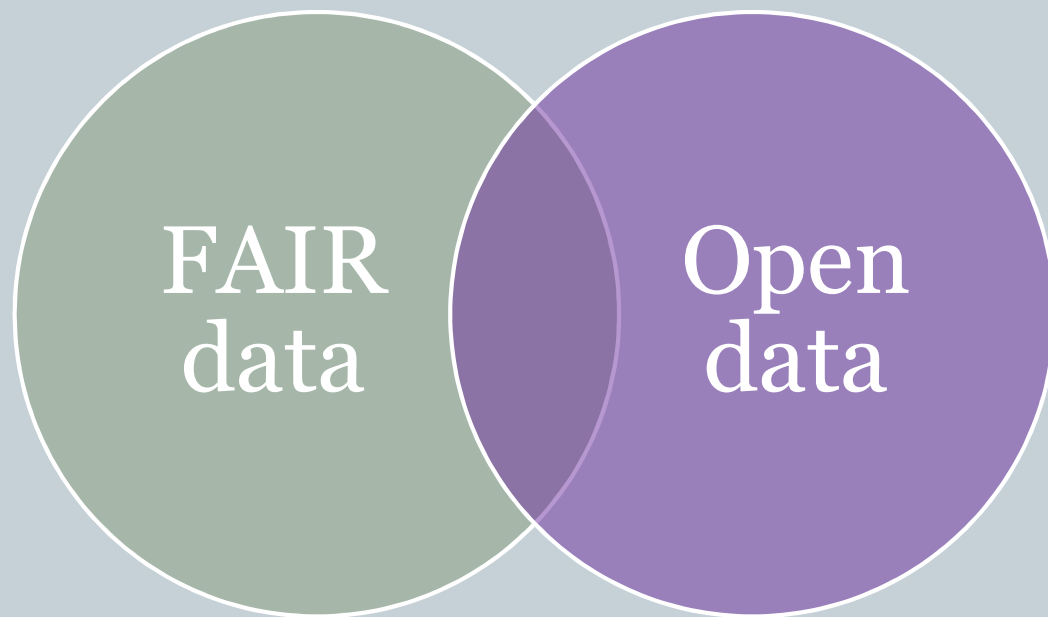


[G8 Science Ministers Statement](#) (2013)

How do Open and FAIR intersect?



Open data and FAIR data



FAIR and Open are not synonymous. Data can be both, one or neither.

And both are on a scale



Implementing FAIR – the EC expert group



Simon Hodson, CODATA
Chair of FAIR Data EG



Rūta Petrauskaitė,
Vytautas Magnus University



Peter Wittenburg, Max Planck
Computing & Data Facility



Sarah Jones, Digital Curation
Centre (DCC), Rapporteur



Daniel Mitchen, Data
Science Institute,
University of Virginia



Françoise Genova,
Observatoire Astronomique
de Strasbourg



Leif Laaksonen, CSC-IT
Centre for Science



Natalie Harrower,
Digital Repository of
Ireland – year 2 only



Sandra Collins,
National Library of
Ireland – year 1 only

Remit of the FAIR data expert group



1. To develop recommendations on what needs to be done to turn each component of the FAIR data principles into reality
2. To propose indicators to measure progress on each of the FAIR components
3. To provide input to the proposed European Open Science Cloud (EOSC) action plan on how to make data FAIR
4. To contribute to the evaluation of the Horizon 2020 Data Management Plan (DMP) template and development of associated sector / discipline-specific guidance
5. To provide input on the issue of costing and financing data management activities

Report framework



1. Concepts – why FAIR?
2. Creating a culture of FAIR data
3. Creating a technical ecosystem for FAIR data
4. Skills and capacity building
5. Measuring change
6. Funding and sustaining FAIR data
7. FAIR Data Action Plan

Culture and technology for FAIR



- Science/research is a cultural system with considerable technological dependency.
- Culture and technology for FAIR data are deeply interrelated.
- Fundamentally important to address cultural and technological requirements for FAIR data holistically.



How the Action Plan is structured



A short tweetable recommendation

- Underpinned by practical and specific action points
- Action points to be allocated to stakeholders

FAIR Data Action Plan applies to EC, member states, and international level, but we also place in context of EOSC to inform this roadmap

Primary recommendations and actions



Step 1: Define and apply FAIR appropriately

Rec. 1: Definitions of FAIR

Rec. 2: Mandates and boundaries for Open

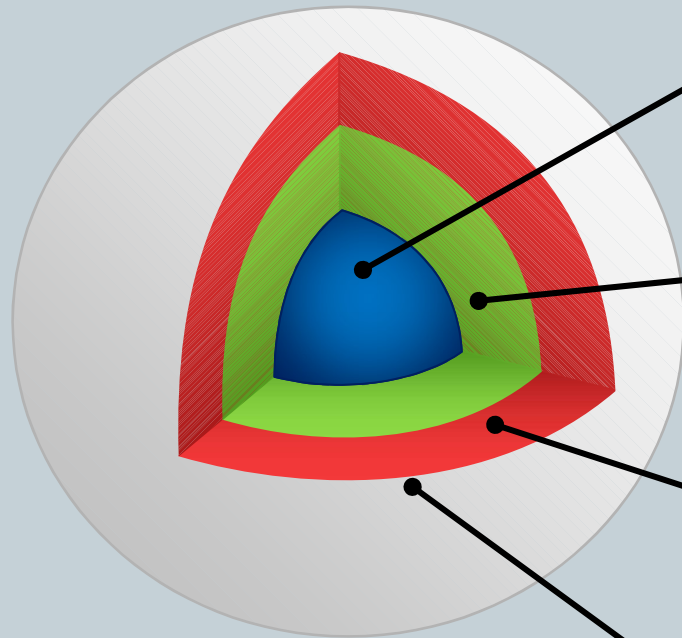
Rec. 3: A model for FAIR Data Objects

FAIR and Open



- Ensure FAIR also covers appropriate openness, assessability, long-term stewardship, timely sharing, data selection and legal interoperability.
- Recognise practices varies. Communities need to self-organise and define what FAIR data means.
- Make Open Data mandate explicit using the maxim ‘as open as possible as closed as necessary’ **but** apply proportionately with genuine best efforts to share

FAIR Data Objects



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 used 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. Basic 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.

Primary recommendations and actions



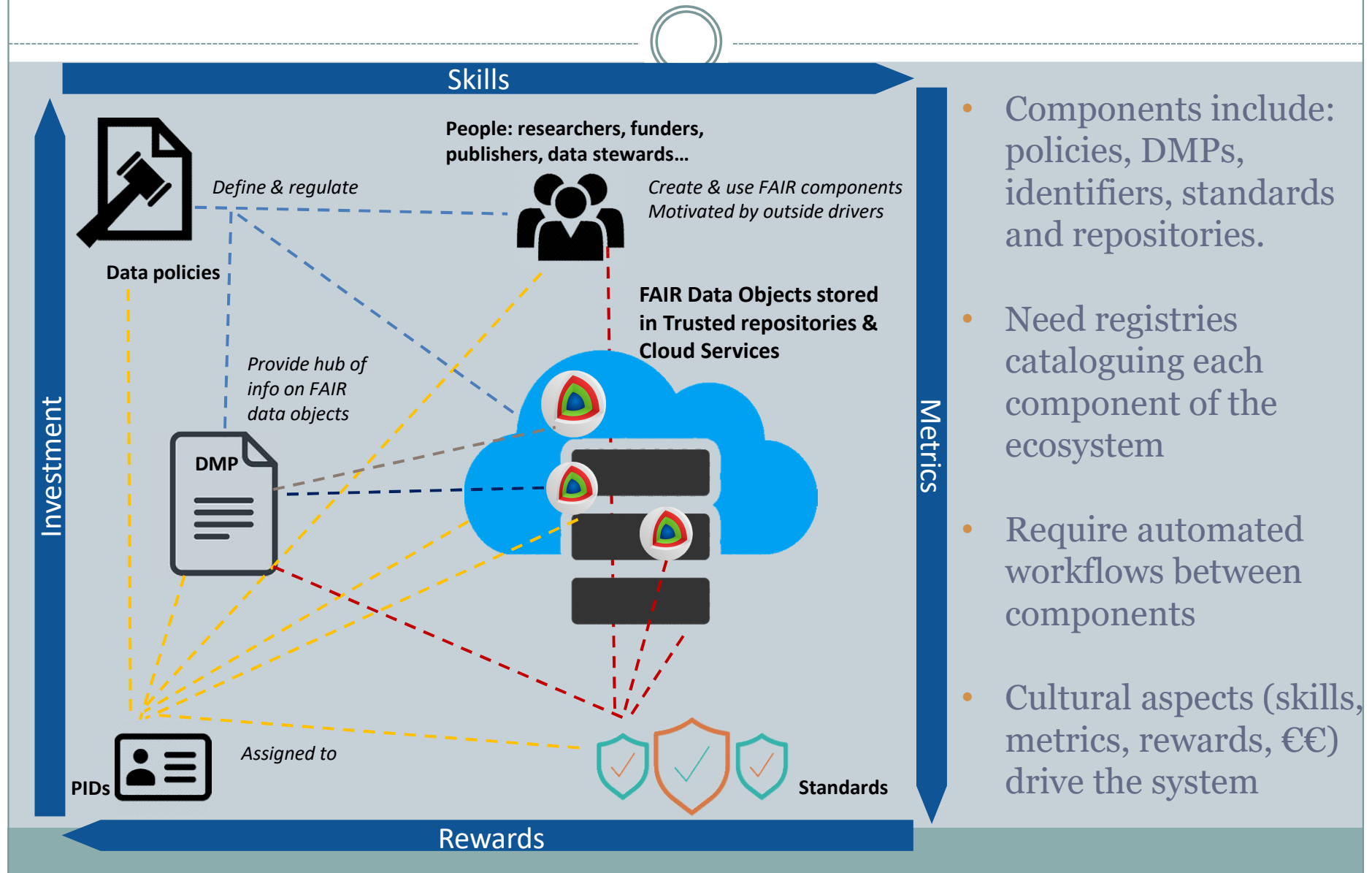
Step 2: Develop and support a sustainable FAIR data ecosystem

Rec. 4: Components of a FAIR data ecosystem

Rec. 5: Sustainable funding for FAIR components

Rec. 6: Strategic and evidence-based funding

FAIR data ecosystem



Sustainable & strategic funding



- The components of the FAIR ecosystem need to be maintained at a professional service level with sustainable funding.
- Funders of research data services should consolidate and build on existing investments in infrastructure and tools, where they demonstrate impact and community adoption.
- Funding should be tied to metrics to ensure interoperable data services and reusable data.

Primary recommendation and actions



Step 3: Ensure robust FAIR data and trusted services

Rec. 7: Disciplinary interoperability frameworks

Rec. 8: Cross-disciplinary FAIRness

Rec. 9: Develop robust FAIR data metrics

Rec. 10: Trusted Digital Repositories

Rec. 11: Develop metrics to assess and certify services

Interoperability frameworks



- Research communities must be supported to develop and maintain their disciplinary interoperability frameworks.
- These include principles and practices for data management and sharing, community agreements, data formats, metadata standards, tools and data infrastructure.
- Interoperability frameworks should be articulated in common ways and adopt global standards where possible to enable interdisciplinary research.
- Important role for international, cross-disciplinary initiatives and fora in development of practices, protocols and standards.

Metrics: for data & services



- A set of metrics for FAIR Data Objects should be developed and implemented, starting from the basic common core of descriptive metadata, PIDs and access.
- Metrics and certification are needed that assess services. Such schemes should not compete with existing repository accreditation, but should measure and accredit components such as identifier services, standards and vocabularies.



Primary recommendations and actions



Step 4: Embed a culture of FAIR in research practice

Rec. 12: Data Management via DMPs

Rec. 13: Professionalise data science and stewardship roles

Rec. 14: Recognise and reward FAIR data and FAIR stewardship

Data management via DMPs



- Any research project should include data management as a core element necessary for the delivery of its scientific objectives, addressing this in a Data Management Plan.
- The DMP should be regularly updated to provide a hub of information on the FAIR data objects.
- Also: use information held in Data Management Plans – structure data to enable information exchange across the FAIR data ecosystem

Roles: data science & data stewardship



- Two cohorts of professionals to support FAIR data:
 - data scientists embedded in those research projects
 - data stewards who will ensure the curation of FAIR data
- Coordinate, systematise and accelerate the pedagogy
- Accreditation of courses to professionalise roles
- Support skills transfer schemes

Recognition & rewards: data & services



- Data should be recognised as a core research output and included in the assessment of research contributions and career progression.
- The provision of data-related infrastructure and services must also be recognised as an essential part of the FAIR ecosystem.
- Require evidence of FAIR data and accredited services and reward accordingly.

Next steps for the Expert Group



- Launch interim report next week at EOSC Summit
- Run consultations over summer via
 - Online platform
 - Webinars
 - Events e.g. FORCE18, International Data Week
- Revise and finalise report and Action Plan
- Formally launch at Austrian Presidency event on 23 Nov

Role of the RDA in FAIR



For FAIR to really succeed, global agreements need to be in place. RDA is an ideal international forum to:

- Clarify definitions of FAIR
- Support research communities to develop disciplinary interoperability frameworks
- Identify commonalities across FAIR implementation to promote the formation and adoption of global standards
- Promote the exchange of good practice and lessons
- Collect use cases and examples...

Explore / define FAIR implementations



- What does FAIR mean in different disciplines?
- What are the best practices in data sharing?
- What are the appropriate boundaries of Open?
- How do we recognise skills and formalise career pathways?
- How do we certify FAIR data and services?
- What metrics denote which services should be sustained?
- How do registries interact and know of each other?
- How do we test that the ecosystem is fit-for-purpose?

Offer guidance and support



- Pointing researchers to the most appropriate services for them (Re3data, FAIRsharing...)
- Raising awareness and adoption of standards (Metadata Standards Catalogue, DCAT)
- Training researchers in data science (CODATA / RDA schools)
- Professionalising data stewardship (Community of Practice, accreditation...)
- Supporting certification of repositories & other services (CoreTrustSeal)

Context specific FAIR Action Plans



- Support the definition of more detailed FAIR Data Action Plans at research community and member state level
- Potential role for European nodes?



Thanks - questions?



<http://tinyurl.com/FAIR-EG>
www.force11.org/group/fairgroup/fairprinciples