

The FAIR principles as guiding principles in open access to research data



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

sarah.jones@glasgow.ac.uk

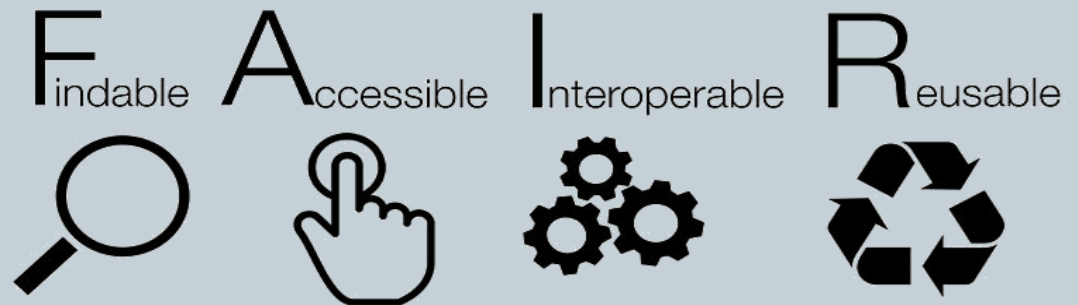
Twitter: @sjDCC



What is FAIR?



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



What is Open?



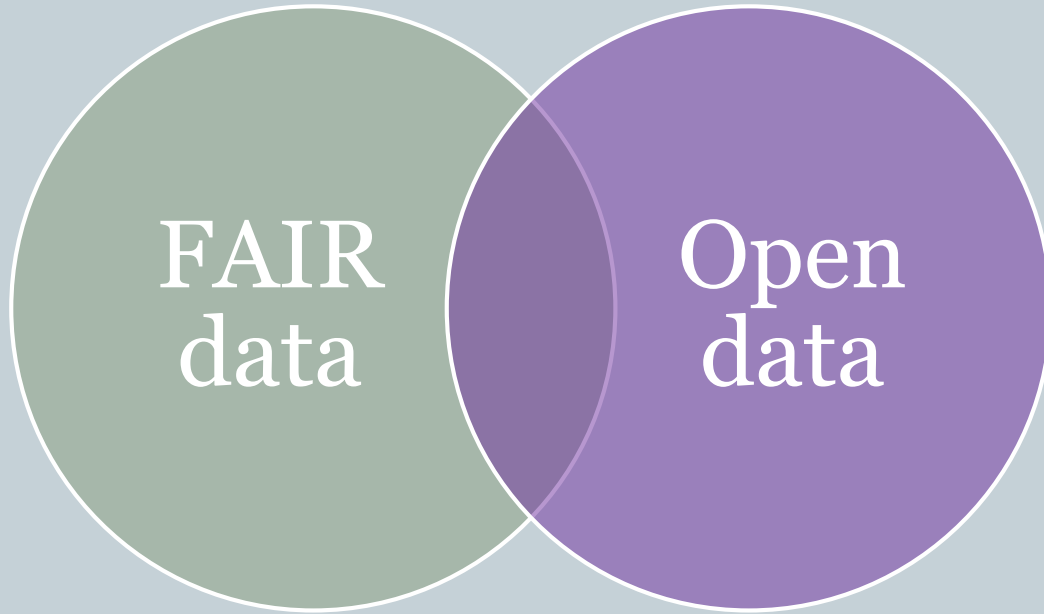
“Open data and content can be freely used, modified, and shared by anyone for any purpose”

Open Definition



OPEN KNOWLEDGE

Open data and FAIR data



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

And both are on a scale



The FAIR principles as guiding principles in open access to research data *for better sharing of*

Sarah Jones

Associate Director, Digital Curation Centre

Rapporteur of FAIR data expert group

sarah.jones@glasgow.ac.uk

Twitter: @sjDCC



FAIR requirements & case studies



<https://doi.org/10.5281/zenodo.1250535>

<https://doi.org/10.5281/zenodo.1245568>

Emerging implementation initiatives



<http://www.copdess.org/enabling-fair-data-project>

GO FAIR: a bottom-up international approach

for the practical implementation of the European Open Science Cloud (EOSC) as part of a global Internet of FAIR Data & Services

<https://www.go-fair.org>



Pilot project and several infrastructure investments

<https://eosc-pilot.eu>

<http://tiny.cc/EOSC-projects>

Why has FAIR gained traction?



- Meaningful and memorable articulation of concepts
- Natural desire to want to be ‘fair’
- Easy to understand at a high-level, but...

To what extent do the following statements represent your experience of using the H2020 template?

	Agree		Neither agree nor disagree		Disagree	
I don't understand what FAIR means	10%	17	16%	28	74%	125

“We understand the basic principle of FAIR, but the terminology is often difficult to grasp immediately. Things could be explained better in plain language”

“The term interoperable is quite confusing sometimes and mixed with re-use.”

OpenAIRE & FAIR data Expert Group H2020 DMP survey

<https://doi.org/10.5281/zenodo.1120245>

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 Mietchen, 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



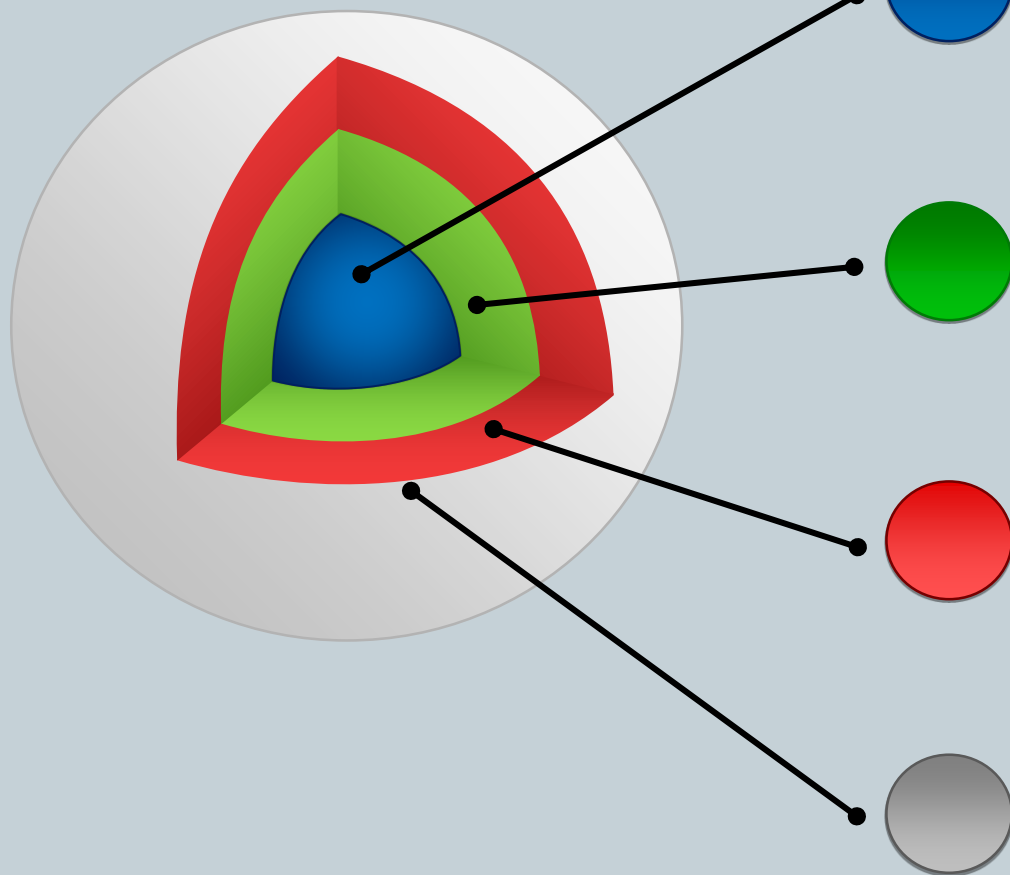
Sandra Collins,
National Library of
Ireland – year 1 only

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

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 (ORCID), 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 needs to be accompanied by the code used to process and analyse the data.

METADATA

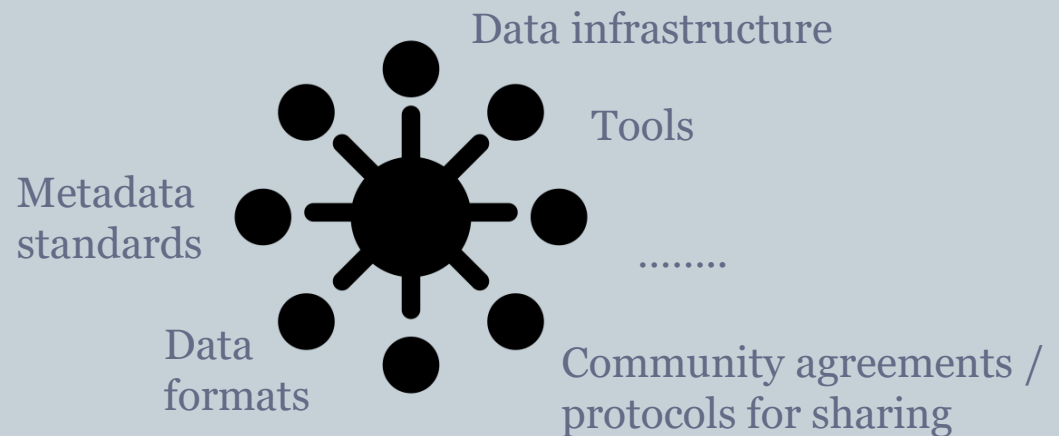
Contextual documentation

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.

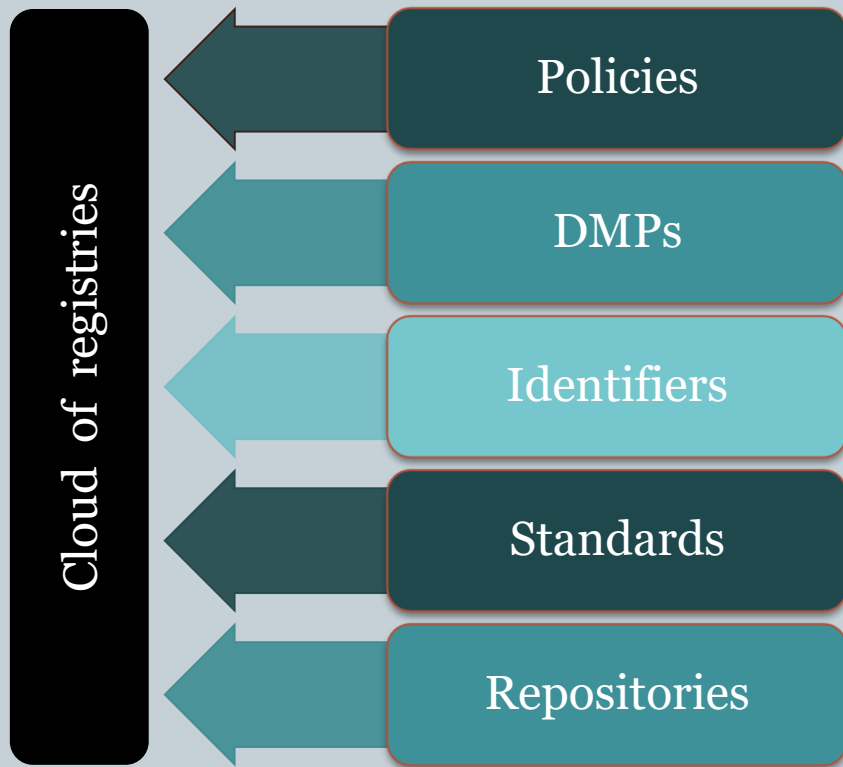
Interoperability frameworks



- Research communities must be supported to develop and maintain their disciplinary interoperability frameworks.
- Interoperability frameworks should be articulated in common ways and adopt global standards where possible to enable interdisciplinary research.

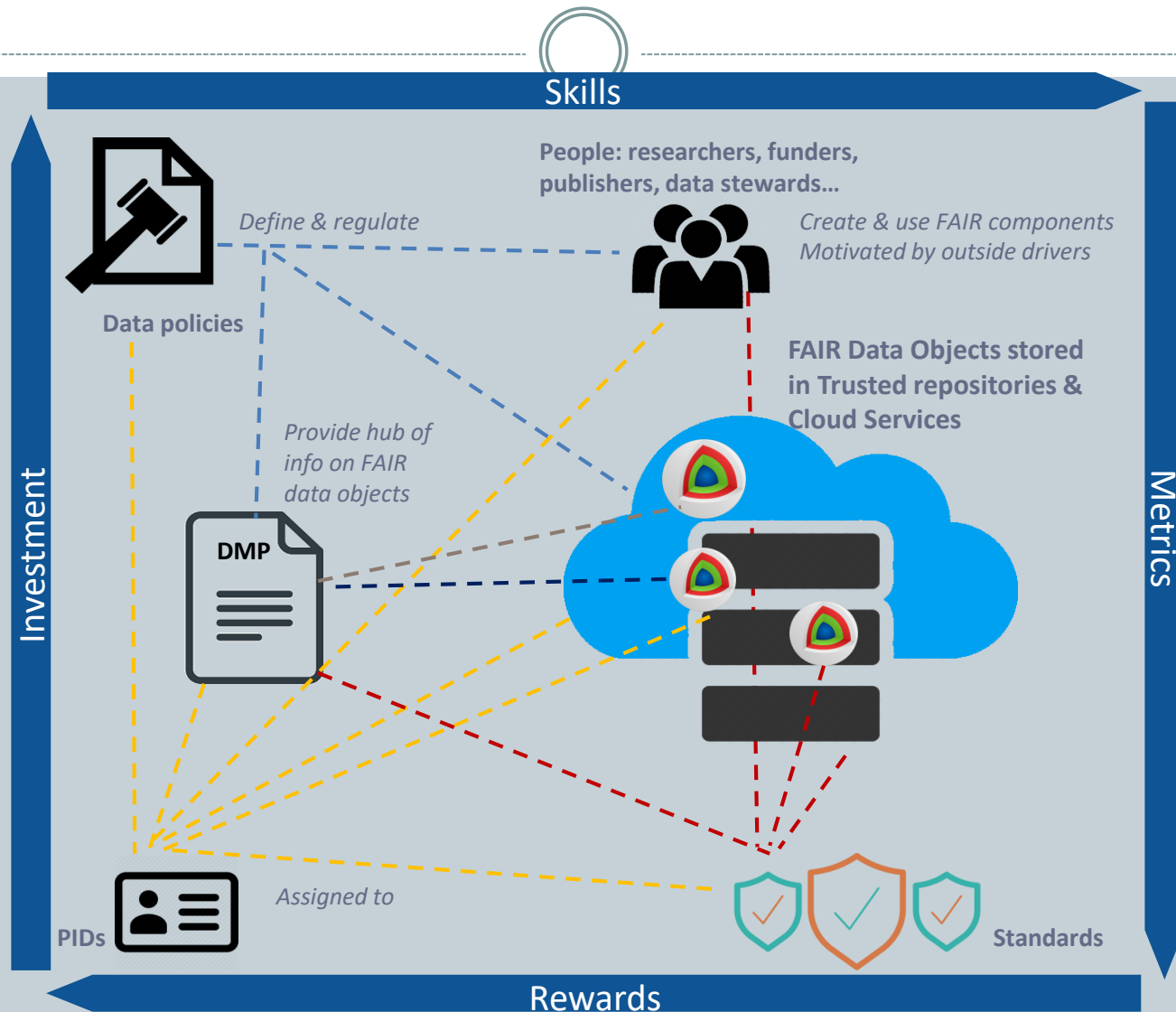


Components of the ecosystem



- Propose 5 key building blocks
- Need registries cataloguing each component of the ecosystem
- Ideally there would be automated workflows between components

FAIR data ecosystem

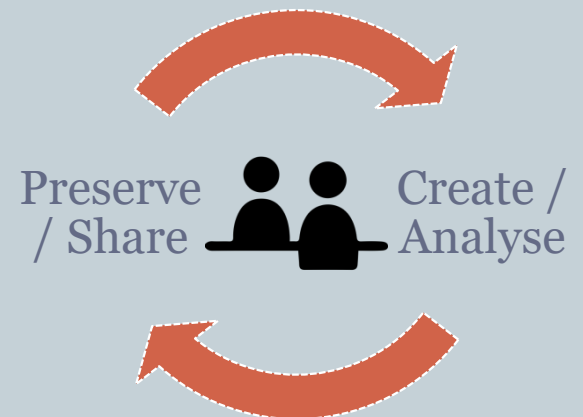


Cultural aspects (skills, metrics, rewards, investment) drive the system

Roles: data science & data stewardship



- Two cohorts of professionals to support FAIR data:
 - data scientists embedded in research projects
 - data stewards who will ensure the curation of FAIR data
- Coordinate, systematise and accelerate the pedagogy
- Support formal and informal learning
- Ensure researchers have foundational data skills



Recognition & rewards: data & services



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

Central role of Data Management Plans



- Any research project should include data management as a core element, addressing this in a Data Management Plan.
- The DMP should be regularly updated
- Use information in DMPs – structure data to enable information exchange across the FAIR data ecosystem



DMP acting as a hub of info on FAIR data objects, connecting to the wider elements of the ecosystem

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 also needed to assess services. Such schemes should not compete with existing repository accreditation, but should measure and accredit components such as identifier services, standards and vocabularies.



Sustainable & strategic funding



- The components of the FAIR ecosystem should be maintained at a professional level with sustainable funding.
- Support to help FAIR data services develop sustainable business models.
- Funders should coordinate investments and build on existing infrastructure and tools.

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

Context specific FAIR Action Plans



- Support the definition of more detailed FAIR Data Action Plans at research community and member state level
- National plans or Nordic plans?



Next steps for the Expert Group



- Currently responding to consultation feedback
 - Over 380 comments on Action Plan
 - Over 150 comments on report
 - Stakeholder groups covered include funders, publishers, universities, research infrastructures, projects, representative groups...
- Report and Action Plan to be launched at Austrian Presidency on 23rd November in Vienna
- Implementation via EOSC and future programmes

Thanks - questions?



<http://tinyurl.com/FAIR-EG>

www.force11.org/group/fairgroup/fairprinciples