"Turning FAIR Data into Reality" Progress and plans from the European Commission FAIR Data Expert Group

Françoise Genova

Observatoire Astronomique de Strasbourg

francoise.genova@astro.unistra.fr

On behalf of the FAIR EG

What is FAIR?

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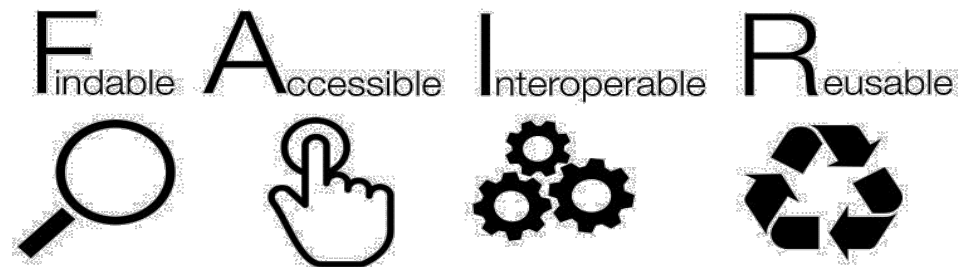

Image CC-BY-SA by SangyaPundir

FAIR Data EG: membership

Simon Hodson, CODATA Chair of FAIR Data EG

Rūta Petrauskaité, 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

FAIR Data EG: Original Objectives

1. To develop **recommendations** on what needs to be done to turn each component of the FAIR data principles into reality (EC, member states, international level)

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**

http://tinyurl.com/FAIR-EG

FAIR Data EG: Extension and Scope

Was to have run March 2017-March 2018. Extended to end Dec 2018 in order to contribute more directly into the FAIR Data Action Plan.

1. Develop **the FAIR Data Action Plan**, by proposing a list of concrete actions as part of its Final Report.

2. Additional **workshops** with experts from relevant European and international interoperability initiatives, input to FAIR Data Action Plan.

3. **Launch and disseminate FAIR Data Action Plan** and support Commission communication in November 2018.

Progress to date

Report outline

- Concepts – why FAIR?

- Creating a culture of FAIR data

- Creating a technical ecosystem for FAIR data

- Skills and capacity building

- Measuring change

- Facilitating change: FAIR Data Action Plan

Consultation

- Ran a consultation last Summer to get input into report drafting

- Shared report outline and encouraged engagement via:
  - Webinars moderated by group members
  - GitHub site

- Many contributions from community

https://github.com/FAIR-Data-EG/consultation

H2020 Data Management Plan survey

- Ran a survey between May-July 2017 in collaboration with OpenAIRE

- Asked about attitudes to DMPs, H2020 template and support needed

- Received 289 responses

- 50% were researchers

- 60% were (also) research support

Full collection on Zenodo:
- Survey report
- Raw data
- Analysed dataset
- Infographic

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

Also webinar slides and video at:
www.openaire.eu/openaire-webinar-results-survey-on-h2020-dmp-template

H2020 DMP survey recommendations

1. Clarify EC requirements for DMPs

2. Revise the DMP template structure

3. Simplify the DMP content and terminology

4. Provide discipline-specific guidelines and example answers

5. Encourage the publishing of DMPs and collate examples

6. Facilitate the inclusion of RDM costs in grant applications

7. Improve DMP review practices and share guidelines

Structuring the FAIR Data Action Plan

A short tweetable recommendation

– Underpinned by several practical and specific action points

– Action points to be linked to stakeholders and timeframes

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

Implementation of FAIR Data Action Plan

- We are to create a common FAIR Data Action Plan

- Use this as rubric to support the definition of related FAIR Data Action Plans at disciplinary, member state and other levels

- Workshop at EOSC Summit on 11th June will be start of this process

Emerging recommendations (excerpts)

FAIR Data

FAIR and Open – FAIR does not imply Open!

- Making data FAIR ensures it can be found, understood and reused – even if this is within a restricted or closed system.

- Essential to clarify the limits of open: open is the default, with **proportionate exceptions** for **privacy, commercial interests, public interests and security.**

Broad definition of FAIR

- FAIR works remarkably well to communicate the attributes and principles necessary to give data value and facilitate reuse.

- Mostly attributes that fall under 'reusable' or relate to the system around the FAIR data: timely release, assessable, stewarded for the long-term in a trusted and sustainable digital repository, responsibilities of users etc.

Broad application of FAIR

- The FAIR principles (and the Action Plan) necessarily apply to a number of digital objects related to the data, as well as the data themselves, e.g.

  - Metadata (and the standard defining the metadata; and the registry listing the standard…)

  - Code (and the metadata/documentation about the code; and the repository where the code can be found…)

  - Applies beyond the digital world, i.e. to the metadata describing analogue or physical research resources.
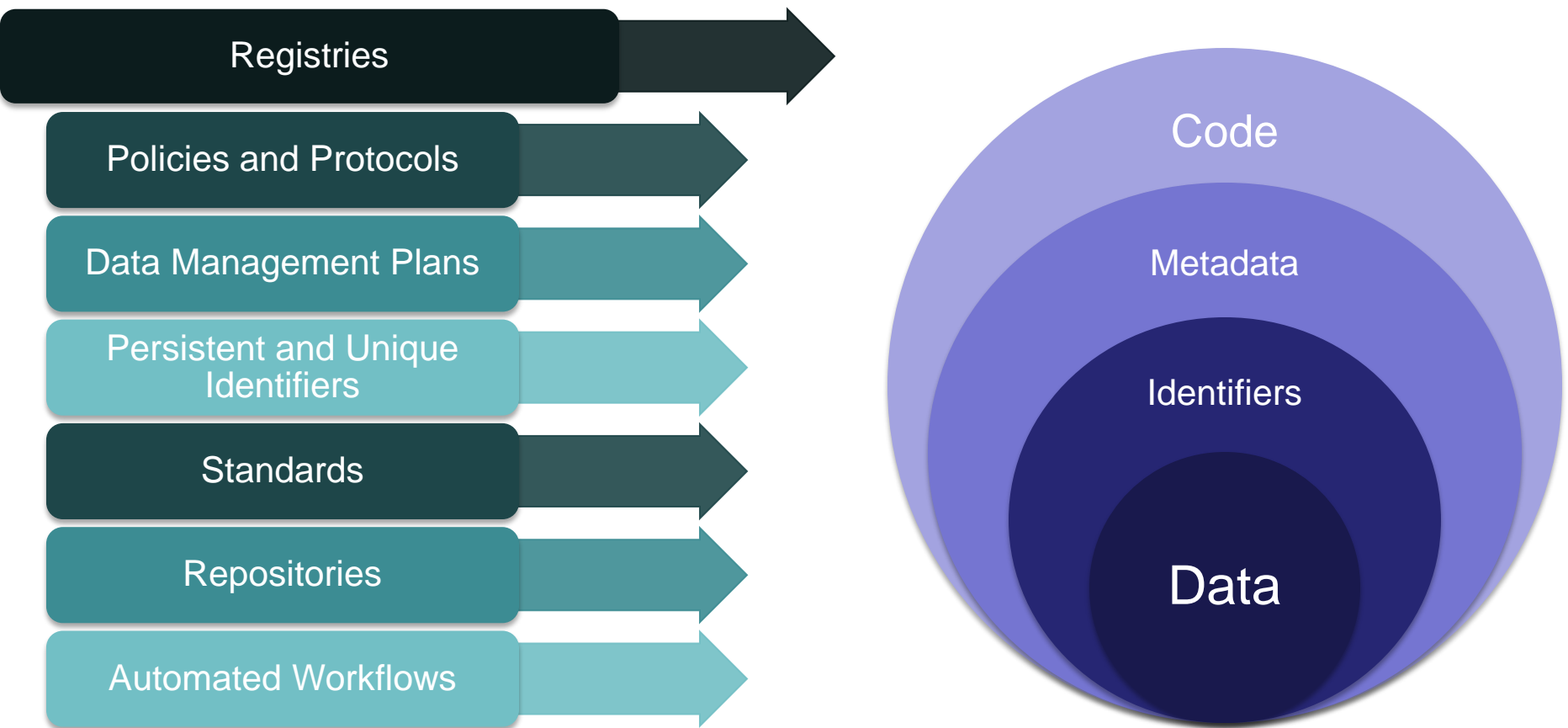
Culture and Technology for FAIR

- 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.

- **Enable research communities to develop their FAIR framework**

Components of a FAIR ecosystem

Registries

Policies and Protocols

Data Management Plans

Persistent and Unique Identifiers

Standards

Repositories

Automated Workflows

Code

Metadata

Identifiers

Data

Trusted Digital Repositories

- FAIR data depends upon an ecosystem of trusted digital repositories (including databases, domain and generic data repositories and data services).

- Data repositories should be incentivised, supported and funded to take the necessary steps towards accreditation (with CoreTrustSeal as a minimum standard).

- Mechanisms need to be developed to ensure that the repository ecosystem as a whole is fit for purpose, not just assessed on a per repository basis.

  – This includes sustainability, provision of domain and multidisciplinary repositories, data services and handover / end of life processes.

FAIR Data EG: Timescale

May - June

Interim report due end May

Launch at EOSC Summit on 11th June in Brussels

June - October

Consult over Summer

Workshop arranged for EOSC Summit

Webinars and other stakeholder events to be planned

November

Final report and FAIR Data Action Plan due

Official launch and formal communications at Austrian Presidency event in Vienna

How and where to engage with us?

- EIRG, in Sofia, 14-15 May http://e-irg.eu/workshop-2018-05-programme

- EOSC Summit, 11 June

- Webinar in mid-late June or in July – i.e. appropriate dates following the EOSC summit

- International Data Week (IDW2018) sessions and possibly side events (Gaborone, Botswana, 4-8 November) http://internationaldataweek.org and https://www.scidatacon.org/IDW2018

ADDITIONAL SLIDES

DMP survey
Key findings

Overall experience

of 189 respondents

- positive 60%
- not applicable 24%
- negative 16%

74%

understand the
FAIR concept
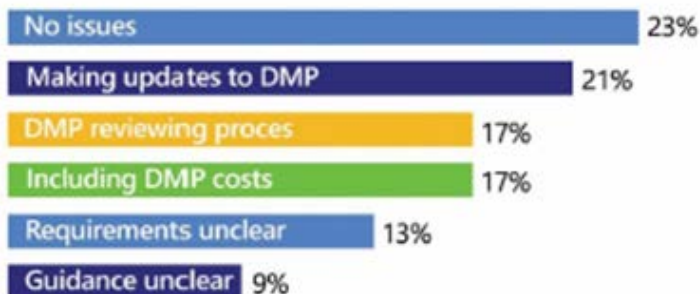
yet practical implementation remains difficult

Almost half
would openly
publish a DMP

48%

Yes if....
And even more would
do so if certain
conditions were met
such as confidentiality.

16%

Issues encountered when following H2020 guidelines

No issues	23%
Making updates to DMP	21%
DMP reviewing proces	17%
Including DMP costs	17%
Requirements unclear	13%
Guidance unclear	9%

Top 5 priorities for a DMP tool:

1. Suggest relevant **standards** for my field and data type
2. **Drop-down options** based on good practice per discipline
3. Give more **examples** or suggested answers
4. Include **disciplinine-specific** guidance and tailoring
5. Recommend **repositories or tools** that I can use

Emerging recommendations

FAIR Data and Open Data

- FAIR does not of itself imply and necessitate Open. It needs to be augmented with the principle:

   'As open as possible, as closed as necessary'

- Data can be accessible under restrictions and still be FAIR. Data shared in 'data safe havens' must be FAIR.

- Making data FAIR ensures it can be found, understood and reused – even if this is within a restricted or closed system.

- Essential to clarify the limits of open: open is the default, with **proportionate exceptions** for **privacy, commercial interests, public interests and security.**

24

Broad definition of FAIR

- FAIR works remarkably well to communicate the attributes and principles necessary to give data value and facilitate reuse.

- Can and should be augmented with certain key concepts that relate to the system necessary for FAIR. But important to resist the temptation of adding letters to FAIR (e.g. FAIR TLC etc).

- Mostly attributes that fall under 'reusable' or relate to the system around the FAIR data: timely release, assessable, stewarded for the long-term in a trusted and sustainable digital repository, responsibilities of users etc.

- Most significant challenge in FAIR are in **Interoperability and Reusability.**

Broad application of FAIR

- The FAIR principles (and the Action Plan) necessarily apply to a number of digital objects related to the data, as well as the data themselves, e.g.

  – Metadata (and the standard defining the metadata; and the registry listing the standard…)

  – Code (and the metadata/documentation about the code; and the repository where the code can be found…)

  – Applies beyond the digital world, i.e. to the metadata describing analogue or physical research resources.

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.

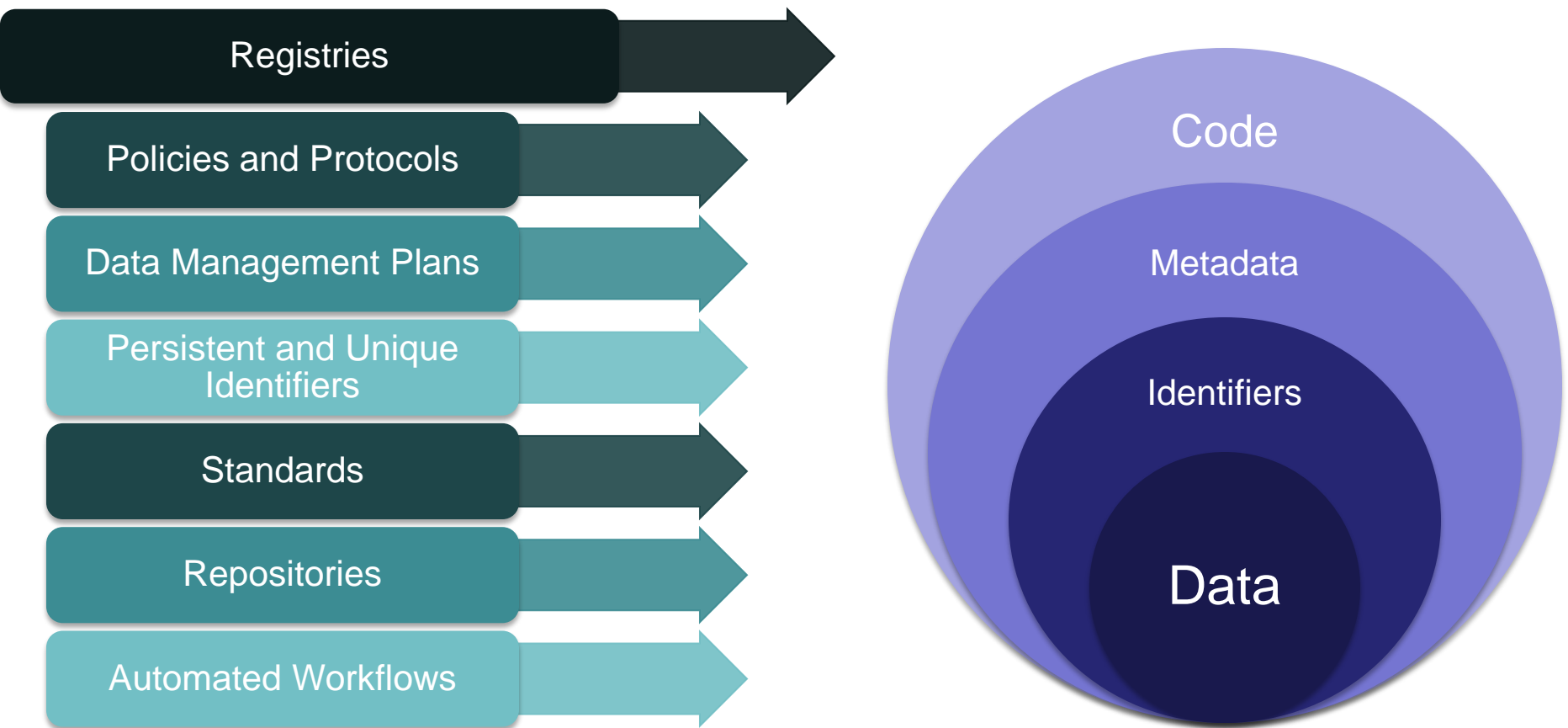# Enable research communities to develop their FAIR data frameworks

- Most significant challenge in FAIR are in **Interoperability and Reusability.**

- Essential to develop enabling mechanisms that support research communities to develop and implement FAIR data frameworks.

- One mechanism is to learn from and share examples from those domains which have had 'FAIR' resources for some time before the term was coined (e.g. aspects of the practice in linguistics and astronomy, genomics and use of remote sensing data).

Enable research communities to develop their FAIR data frameworks

Enabling mechanisms include:

- Collection and sharing of case studies where 'FAIR' data (before or after the term) has facilitated domains.

- Mechanism to encourage and facilitate the development of community agreements for FAIR practices.

- Mechanisms to encourage and facilitate the development of domain and interdisciplinary standards (with particular attention to collection/study level description, value-level concepts and provenance information.

- Important role for international, cross-disciplinary initiatives and fora in development of practices, protocols and standards.

Components of a FAIR ecosystem

Registries

Policies and Protocols

Data Management Plans

Persistent and Unique Identifiers

Standards

Repositories

Automated Workflows

Code

Metadata

Identifiers

Data

Trusted Digital Repositories

- FAIR data depends upon an ecosystem of trusted digital repositories (including databases, domain and generic data repositories and data services).

- Data repositories should be incentivised, supported and funded to take the necessary steps towards accreditation (with CoreTrustSeal as a minimum standard).

- Mechanisms need to be developed to ensure that the repository ecosystem as a whole is fit for purpose, not just assessed on a per repository basis.

    – This includes sustainability, provision of domain and multidisciplinary repositories, data services and handover / end of life processes.

Skills and Competencies

- As well as improving data skills in all researchers, steps need to be taken to develop two cohorts of professionals to support FAIR data:

  - **data scientists** embedded in those research projects which need them; and,

  - **data stewards** who will ensure the management and curation of FAIR data.

- A concerted effort should be made to coordinate, systematise and accelerate the pedagogy and availability of training for data skills, data science and data stewardship, including:

  - promoting and sharing curriculum frameworks and OERs;

  - a coordinated and adequately-funded train-the-trainers programme (for data science and data stewardship);

  - a (lightweight) programme of certification and endorsement for organisations delivering this training.

Metrics, Rewards and Recognition

- Metrics to support and encourage the transition to FAIR data practices should be developed and implemented.

    – The design of these metrics needs to be thorough and mindful of unintended consequences.

    – Metrics should also be monitored and regularly updated.

- Certification, evaluation or endorsement schemes are needed for the essential components of the FAIR data ecosystem.

- Metrics, rewards and recognition for research contributions need to give appropriate and significant weight to 'publication' of FAIR data

    – All journal editorial boards and research communities should require and recognise the availability of data underpinning 'published' findings.