Driving the paradigm shift towards Open Science: FAIR data

Kostas Repanas

European Commission

Directorate-General for Research and Innovation

Open Science unit

EUA-FAIRsFAIR webinar

Dec 8th, 2020

Open Science at the EC

- **Open Science** means **sharing knowledge and tools as early as possible**, not only **between researchers** and **between disciplines**, but also with **society at large**.

- Open Science improves the **quality**, **efficiency** and **creativity** of research and the **trust by society in science**. In particular, OS is beneficial for science, scientists and funders, e.g.:

  - tackles the reproducibility crisis;

  - faster response to societal challenges e.g. Coronavirus, Ebola;

  - access to and sharing results yields higher impact through collaborations;

  - generates new research findings & reduces inequalities;

  - large opportunity costs of non-FAIR data— €10.2bn/year (source: Cost-benefit analysis of FAIR research data, 2017).

- The Commission acts as **policy maker** (propose legislation and encourage MS), a **funder** (we set requirements to our projects) and a **capacity builder** (we fund 'enabling' projects).

European Commission

Evolution of our policies across the FPs

2008

FP7
OA **Pilot**
Deposit and open access

2014

H2020
OA **Mandatory**
Deposit and open access

& ORD/DMP **Pilot**

2017

H2020
OA **Mandatory**
Deposit and open access

& ORD/DMP **by default (exceptions)**

Horizon Europe

Planned under **Horizon Europe (2021)**:

- Open Science (OA, RDM, Citizen Engagement, etc.) embedded across the FP. OS to play a role in the:
  - **Evaluation** of proposals (methodology)
  - **Grant Agreement**
  - **Reporting**—during the project's lifetime

- Strengthening of the obligations with respect to open access and focus on responsible RDM in line with FAIR

Research data under Horizon Europe

- The governing principle will be to manage research data **responsibly**, **in line with FAIR**, and for data to be "**as open as possible and as closed as necessary**". We are **likely** to see the following:

  - At proposal stage, beneficiaries will be **evaluated** on **preliminary RDM considerations**

  - All projects that **generate/collect/re-use** research data will have to establish and regularly update a **Data Management Plan** (living document and machine-actionable)

  - Beneficiaries will have to **deposit data in a repository** and link their data to publications they underpin, if applicable

    - Valid repositories will be those that provide **PIDs** for the data and ensure **rich metadata** in line with FAIR.

    - For some actions, an additional obligation to deposit in a repository that is federated under EOSC.

  - Beneficiaries to ensure **open access ASAP** under CC BY or CC0 licences (or equivalent), unless exceptions apply (duly justified in the DMP)

European Commission

Other research outputs under Horizon Europe

- Data is not the only research output that should be managed → **digital**: Software, algorithms, protocols, workflows, models, and **physical**: reagents, antibodies, hardware, etc. all need to be properly managed:

    - We would like to see other research outputs also **described in the DMP**.

    - There will be a **strong encouragement** for beneficiaries to **deposit and provide open access via a repository** to these outputs, unless legitimate interests or constraints apply.

- Research outputs, either digital or physical, should also be findable, interoperable, accessible and re-usable → We need a better understanding of

    - What **FAIR** means for digital objects, other than data (e.g. software)

European Commission

FAIR data against COVID-19

- This effort represents a practical use case of **FAIR data in action**.

- The European Commission launched on 20 April the **European COVID-19 Data Platform** together with EMBL-EBI, ELIXIR, and other partners, and with the support of MS.

- The Platform is a thematic priority pilot to realise the **EOSC vision** and to showcase the added value of **FAIR data sharing** to advance science and benefit researchers

- Responds to the need to capitalise on the quick and wide **sharing**, **re-use**, **processing** of **research data** and **metadata** on the SARS-CoV-2, and the COVID-19 disease.

- Very strong focus on ensuring that data and metadata are as **open** and as **FAIR** as possible.

- Crucial lesson: Researchers need **quick** and **unrestricted** access to multiple data sources to accelerate their research. **FAIR data** is an essential component in this.

FAIR is great. What's next?

- There is **increased familiarity** across the board with the FAIR principles, and a firm **conviction about the positive impact** of FAIR on e.g. enhancing the transparency & quality of the scientific practice, enhanced productivity, etc.

- There's a need to **mainstream FAIR practices** across the scientific community. Funders across Europe and beyond are **incorporating the FAIR principles** into their policies & requirements.

- **Stumbling block**: scarcity of **comprehensive** and **comparable** evaluation mechanisms/metrics for what have up to now been **abstract** or **imprecisely defined requirements**.

- A number of initiatives are filling this gap:

  - RDA WG on the **FAIR Data Maturity Model**

  - The **FAIRware initiative**

European Commission

FAIR Data Maturity Model

- The FDMM WG has delivered **a set of indicators** (Lego building blocks of FAIRness), **priorities** on these indicators (taking into account different readiness levels across communities), and **guidelines** (explaining in detail what each indicator is about, and how it can be satisfied).

- The outputs of the WG are the **foundation** on which FAIR assessment/evaluation methodologies can be built with two important purposes:

  - **Self-assessment**: To allow **researchers** to **estimate** the FAIR maturity of their research outputs and how to progressively improve.

  - **Evaluation**: To allow **funders** to **monitor compliance** against set requirements

The FAIRware initiative

- This is an effort led by **RoRI** (Research on Research Institute) supported by funders across the **UK**, **Europe** and **Canada**.

- The goal is to develop **new open source software tools and systems** to **mainstream uptake of the FAIR principles** (e.g. including automatic FAIRness evaluation tools).

- Involvement of H2020 project **FAIRsFAIR**

  - **FAIRsFAIR** is a Coordination and Support Action funded by the EC aiming at the mainstreaming of the FAIR principles across communities. It is developing and implementing global standards for FAIR certification of repositories, and linked to ongoing EOSC efforts.

  - Its involvement in this initiative ensures a significant **pool of existing knowledge** on which to continue building, and prevents related activities from happening in isolation.

FAIR assessment

FAIRsFAIR has developed F-UJI, a service based on REST, and is piloting a programmatic assessment of the FAIRness of research datasets in five trustworthy data repositories.

F-UJI Automated FAIR Data Assessment Tool

1

The F-UJI assessment is based on **16 out of 17 core FAIR object assessment metrics** developed within FAIRsFAIR and each corresponding to a part or the whole of a FAIR principle. F-UJI adheres to existing web standards and PID resolution services best practices and utilises external registries and resources such as re3data[1] and Datacite[2] APIs, SPDX License List[3], RDA Metadata Standards Catalog[4], and Linked Open Vocabularies (LOV)[5] For information on the practical tests implemented against the metrics, see Devaraju, Huber, et al., 2020.

The source code is now available with a free license through Github. Any feedback on improving the tool and associated metrics can be added as an issue on Github.

FAIR-Aware is an online tool which helps researchers and data managers assess how much they know about the requirements for making datasets findable, accessible, interoperable, and reusable (FAIR) before uploading them into a data repository.

FAIR Aware

2

https://www.fairsfair.eu/tools-software

The tool comprises 10 carefully designed questions, each generously supplied with additional information and practical tips which extend users' understanding of the FAIR principles as they work through the questionnaire with a target dataset in mind.

European Commission

EOSC-Synergy

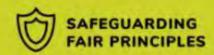The intermediate report (D3.3) provides:

- Requirements for FAIR data repositories

- Metrics for programmatic assessment based on FAIRsFAIR and RDA recommendations

- An implementation plan divided into 3 phases: evaluate, establish and endorse

https://www.eosc-synergy.eu/wp-content/uploads/public-EOSC-SYNERGY-WP3-D3.3-FINAL.pdf

How do we move forward?

- **FAIR data** are an essential component in **Open Science policies** to use the wealth in data to generate new and impactful research and tackle the reproducibility crisis

- As a funder, the European Commission, will impose **stringent requirements** on all Horizon Europe projects with respect to publications, data, and other research outputs

- Specifically, with respect to data, the focus will be on **responsible research data management in line with the FAIR principles**

- The community-backed work developed under the **FDMM WG**, and new work on the **FAIRware initiative** are key steps towards **evaluation / assessment mechanisms of FAIRness**

- **FAIR is a journey: Assessment of FAIRness** that does not punish beneficiaries, but educates, raises awareness and ultimately increases data reuse

European Commission

Thank you

European Commission