

Workshop report

FAIR data maturity model Working Group

Online meeting #10 — 16 September 2020

Project	RDA FAIR data maturity model working group	Date & Time	16 September 2020 07:00 — 08:30 UTC 16 September 2020 13:00 — 14:30 UTC
Type	Online meeting	Location	Google Meet
Meeting Chairs	Keith Russell Shelley Stall	Issue date	2020-10-12

Agenda

1. Welcome, objectives of the meeting
2. FAIR assessments and DMPs
3. FAIR use cases
4. Next steps

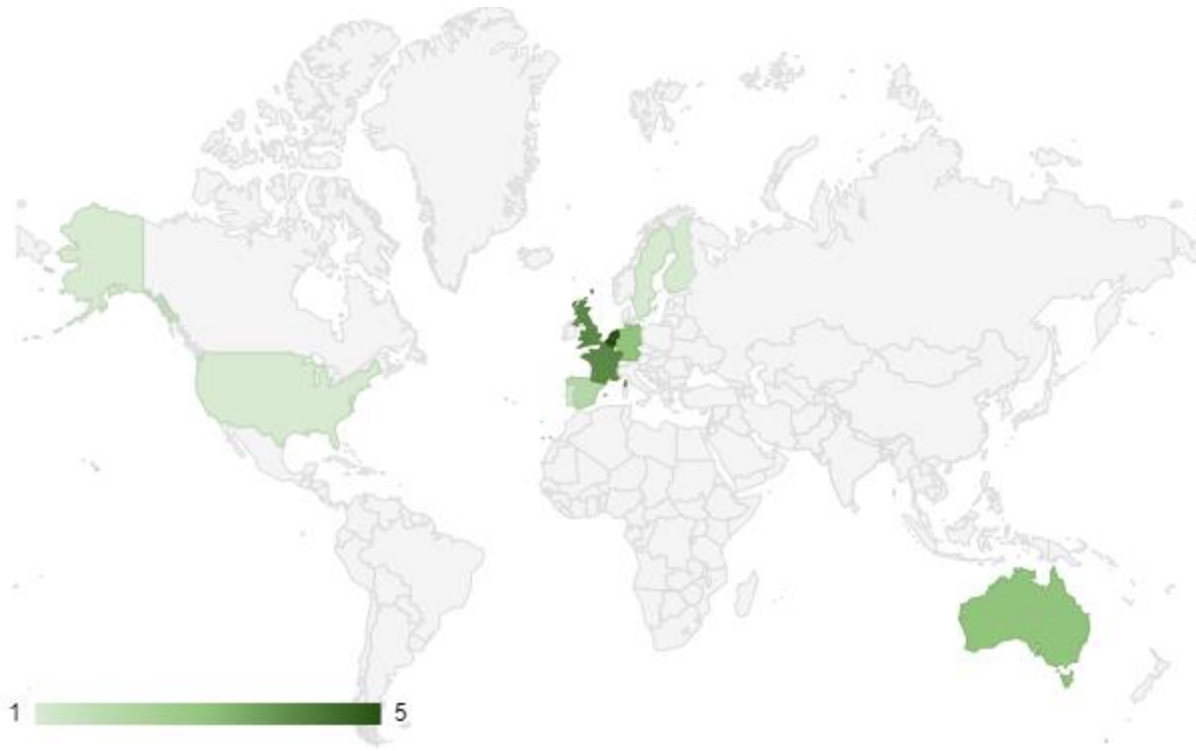
Participants

The workshop was well attended. Here below is a non-exhaustive list of the participants.

Name		Affiliation
Adam Dinsmore	GB	Wellcome Trust
Alice Frémand	UK	UK Polar Data Centre, British Antarctic Survey
Angus Whyte	GB	DCC / H2020 FAIRsFAIR
Anusuriya Devaraju	DE	PANGAEA / University of Bremen
Barbara Sierman	NL	KB National Library of the Netherlands
Carlos Casorrán Amilburu	BE	European Commission DG RTD
Christian Pichot	FR	INRAE - AnaEE
Christophe Bahim	BE	PwC, Editor team
David Carr	GB	Wellcome Trust
Dimitri Szabo	FR	INRA
Edit Herczog	BE	Chair, Vision & values SPRL

Fernando Aguilar	ES	CSIC
Ibrahim Emam	UK	Imperial College, Data Science Institute
Isabel Campos	PT	EOSC-synergy project (CSIC)
Jez Cope	UK	Data Services Lead at The British Library
Keith Russell	AU	Chair, ARDC
Konstantinos Repanas	BE	European Commission DG RTD
Maggie Hellström	SE	ICOS ERIC / ENVRI-FAIR
Max Dekkers	ES	Independent Consultant, Editor team
Mari Elisa Kuusniemi	FI	University of Helsinki
Mikaela Lawrence	AU	SIRO Australia
Mohamed Yahia	FR	Inist-CNRS / Datacite
Mustapha Mokrane	NL	DANS
Nicolas Loozen	BE	PwC, Editor team
Patricia Herterich	GB	Digital Curation Center
Peter Hayes	AU	Interplay Project and ANU
Rob Hooft	NL	Dutch Techcentre for Life Sciences
Romain David	FR	INRA
Shelley Stall	US	Chair
Sandor Brockhauser	DE	EuXFEL
Stefan Decker	DE	RWTH Aachen and Fraunhofer FIT
Wouter Addink	NL	Naturalis/DiSSCo
Yan Grange	NL	ASTRON

Here below is a map representing the provenance of the different participants



Meeting summary

The chairs welcomed the participants and introduced the agenda of the meeting. On the agenda of the meeting were four guest speakers who were invited to discuss FAIR and data management plans as well as concrete FAIR use cases. Before giving the floor to the speakers, the chairs presented the state of play. Currently, the FAIR data maturity model working group is well underway in its maintenance mode. Besides, the chairs unveiled the tentative schedule for the end of the year.

- Workshop #10
- A survey trying to bridge the gap between funders and communities
- Workshop #11
- The establishment of a work plan for 2021
- CODATA FAIR symposium

On top of that, the chairs reported two publications since the last meeting of the working group.

- Publication of the [FAIR Data Maturity Model: Specification and Guidelines](#) as an RDA recommendation, 25 June 2020
- EOSC-SYNERGY [Intermediate report on technical framework for FAIR principles implementation](#), 2 Sept. 2020

The editor of the working group introduced the presentations on the role of DMPs and how DMP tools can help achieve FAIRness.

Rob Hooft presented the Data Stewardship Wizard, an educational tool, which can be seen as a FAIR maturity model asking questions in the present tense. In other words, the tools try to quantify the FAIRness of existing data but also it tries to evaluate the data that does not exist yet. Indeed the funders want the data to be FAIR. The DS try to help to achieve FAIRness and see how there is to do to do so. It predicts how FAIR the data will become.

The DS is bespoke and can be tailored in the back-end (e.g. choosing the questions to ask and make the best DMP). All questions are multiple choice questions and are machine readable. Nevertheless, some questions which are not about the FAIR principles but have an influence on, are present in the predefined questionnaire. The questionnaire filled in is exportable in any possible format.

Rob Hooft clarified that the DS is born from the burden of DMPs for data researchers. A remark was made from the audience that there is currently not enough studies to support DMPs and that they are based on a lot of assumptions.

Mustapha Mokrane pointed out that data generated in a project can be very diverse and not necessarily requires the highest level of FAIR. Mustapha Mokrane questioned the value of an average FAIR rating. Sarah Jones commented that very few funder DMPs ask anything about the selection of data and which has long-term value or should be shared. Rob Hooft mentioned that the DS can ask the same questions repeatedly (for a project that delivers five different datasets).

Angus Whyte and Patricia Herterich introduced FAIRsFAIR and its interest in DMPs. Currently, [FAIRsFAIR is working on implementing metrics](#). FAIR assessments should not be only during the planning phase of a project but also throughout its lifecycle. More than that, the DMP should be iterative from the planning to the publication. The speakers introduced FAIRAware and [FUJI](#). FAIRAware is about raising awareness and educating the researchers whereas FUJI is about assessment datasets.

FAIRsFAIR produced a set of recommendations to foster machine-actionable DMPs. Angus and Patricia concluded their presentation by showing community interest in use cases as well as introduced DMPonline, a FAIRsFAIR online DMP tool, which allows data stewards in institutions to provide reviews when requested.

DMPs tools help understand better what is expected as well as what are the positive outcomes of FAIRness over time.

Maggie Hellström shared that the most "effective" argument when selling in DMPs and FAIR to researchers is that it safeguards their own work and makes reporting & publishing so much easier. She also observed that considering the public dissemination of DMPs, the timing is important; in some disciplines, a rich & detailed DMP may contain information that would allow the *scooping* of research ideas by competitors.

Rob Hooft observed that although data is under scrutiny in portals, they are only designed for humans. FAIR for humans but not FAIR for machines.

The following point was brought to light by the audience: facilitating automated assessment of (machine-actionable) FAIRness is in principle fine, but what about the timing (and frequency) of performing such tests? The FAIRness of datasets needs to be followed-up over time. Angus Whyte responded it should be up to research institutions and funders to ensure there is continuous evaluation of FAIRness of research outputs, with support from RI and other service providers

The editor introduced the next speakers, who were funding agencies representatives. They were invited to talk about the extent to which they are involved in FAIR assessments and in what context.

The editor introduced the next speakers, who were funding agencies representatives. They were invited to talk about the extent to which they are involved in FAIR assessments and in what context.

David Carr and Adam Dinsmore presented FAIRware which is a funder-led project to help accelerate implementation of the FAIR principles through the design and development of software tools for FAIR assessment. The goal is to enable funders to assess the FAIRness of research outputs at a portfolio level, and identify areas where additional support may be required to help implement the FAIR principles. It will also aim to enable researchers to assess the FAIR status of research outputs and provide actionable guidance.

FAIRware is one of five flagship projects of the Research on Research Institute (RORI) – an international consortium of research funders, academics, research organizations and technologists which aims to champion transformative research on research systems, cultures and decision-making. It is supported by a consortium of five funders - the Austrian Science Fund (FWF), the Canadian Institutes of Health Research (CIHR), the UK National Institute of Health Research, the Swiss National Science Fund (SNSF) and Wellcome

The next step will be to seek a supplier (or suppliers) to design and build a minimum viable product for FAIRware with the help of Wellcome Datalabs. The partners are committed to developing FAIRware as an open source community project. The partners have signed a collaboration agreement with the FAIRsFAIR project – who will provide expert input and advice and help ensure the work builds upon and aligns with previous and ongoing work in this space.

David and Adam announced that the RFP will be published in a couple of weeks (it was published on Monday 28 September – see: <https://wellcome.org/what-we-do/our-work/research-research>)

Konstantinos Repanas and Carlos Casorrán introduced Open Science and FAIR data. The COVID-19 outbreak helped to realise the importance of data sharing. The Commission hopes that the trend will continue and will continue to invest substantially in their next framework programme, Horizon Europe, to enable FAIR data and respond to societal challenges. The Commission will propose requirements and recommendations to increase open access and open data requirements. They agreed with other funders that DMPs are the cornerstone of proper and responsible data management. DMPs should be living documents.

Through its COVID-19 data platform the Commission showed its vision for federated data sharing. The platform, as a use-case of the European Open Science Cloud will promote early sharing of data. The speakers stressed the importance of having data as open as possible and as FAIR as possible. The Commission will not force researchers to open up their data, but rather subscribe the notion of “as open as possible, as closed as necessary”, making open data the default, unless good reasons exist to keep data under restricted access.

Currently, publications remain a first-class research output, with data gaining prominence and relevance. Software, however, has historically received less attention. The Commission wants to highlight the importance of software, algorithms, protocols, workflows, tools, etc. These types of outputs should also comply with the FAIR principles, and beneficiaries will be strongly encouraged to ensure this.

The Commission continues to strongly support Open Science and the policy regarding FAIR and Open data will be mainstreamed throughout the programme. FAIR data is a cornerstone of their Open Science policy.

The editors introduced the survey on FAIR assessments which tries to improve the understanding of benefits and challenges of the FAIR assessments from the perspective of the funders and communities. The objective are as follows:

1. Formulate conclusions and recommendations on the level of policy, (i.e. better understanding of the perspectives of both sides)
2. Finding out how the research community and the funders' community might want to use the model and what changes they would want to see

The editors will anonymise, compile and analyse the results, draw conclusions and propose recommendations for further actions to improve understanding. The editors called for volunteers.

Next to that, the editors introduced a list of topics likely to be investigated for the next versions of the FAIR data maturity model. The topics were proposed for a vote and below are the most voted for topics (by order of importance):

- Metadata practices: (i) Role of generic platforms in improving domain-specific metadata, (ii) Metadata at several levels, (iii) Consensus on minimum level for 'Rich' metadata and (iv) Shared understanding of knowledge representation
- Data granularity (collection, dataset, data item)
- Approaches toward evaluation of FAIR assessment tools and services, taking into account community aspects
- PID practices across communities (identifiers for metadata, data, separately, combined)

The idea, of course, is to tackle all of the topics proposed but in a specific order.

The chairs wrapped up the call by disclosing the agenda for the remainder of the year. First, the WG will identify topics that need clarification and consensus and second it will address the governance and maintenance practical aspects. In 2021, the maintenance and preparation for the first 1st revision will take place (i.e. the topics will be developed and consensus-driven solutions to optimize the model and move away from a fit-for-all to a tailored solution will be proposed).

The chairs reminded that the survey will run until late November and that the date for the next webinar - late October / Early November - will be communicated soon.