

# Workshop report

## *FAIR data maturity model Working Group*

Online meeting #9 — 20 & 21 May 2020

Project	RDA FAIR data maturity model working group	Date & Time	20 May 2020 15:00 — 16:30 UTC 21 May 2020 07:00 — 08:30 UTC
Type	Online meeting	Location	Google Meet
Meeting Chairs	Shelley Stall Edit Herczog	Issue date	2020-06-03

## Objectives

The 9th and last workshop of the FAIR data maturity model Working Group – in its current form – had a threefold objective. First, the editorial team reported back to the members on the comments received during the [public review period](#). The comments were clustered and categorised. Two courses of action were identified; i) comments that led to amendments in the document and ii) comments that were beyond the scope of the WG. Secondly, early adopters of the – to be – RDA recommendations were invited to share their experiences with regards to the adoption of the FAIR data maturity model. Thirdly, the last WG meeting was an occasion to present the high-level work plan for the next six months, and more specifically to announce the kick-off of the maintenance phase.

## Agenda

1. Welcome, objectives of the meeting
2. Roundtable
3. State of play
4. Public review period
5. Early adopters – Experience sharing
6. RDA recommendation
7. Disseminate the FDMM to communities
8. Thanks and resources
9. Maintenance work plan
10. Initiation of a feedback loop

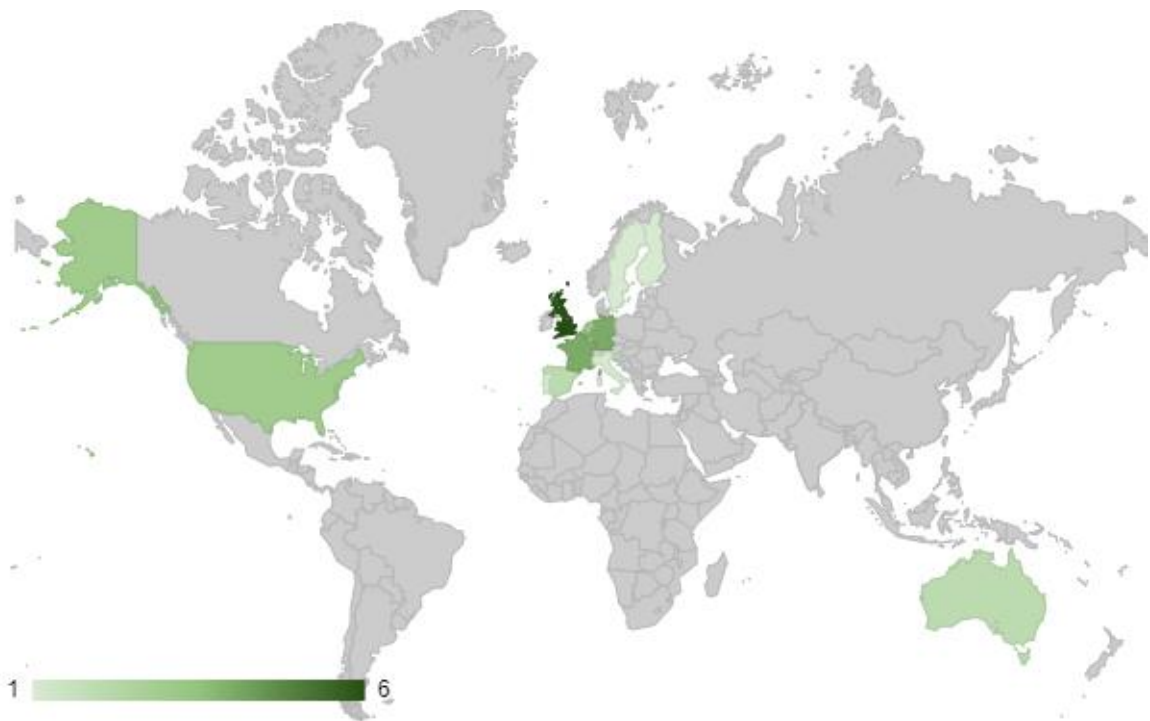
# Participants

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

<b>Name</b>	<b>Country code</b>	<b>Affiliation</b>
Alicia Fátima Gómez Sánchez	ES	INAECU
Anusuriya Devaraju	DE	PANGAEA / University of Bremen
Carlos Casorrán	BE	European Commission DG RTD
Christian Pichot	FR	INRAE
Christophe Bahim	BE	PwC, Editor team
Cristina Justino	PT	University of Aveiro Portugal
Daniele Bailo	IT	INGV - Istituto Nazionale di Geofisica e Vulcanologia
Emna Amdouni	FR	LIRMM
Françoise Genova	FR	Strasbourg Astronomical Data Center
Ge Peng	US	North Carolina State University / NCEI
Hervé L'Hours	GB	UK Data Archive
Janez Štebe	SI	University of Ljubljana, Social science data archives
Keith Jeffery	GB	Keith G Jeffery Consultants
Keith Russell	AU	Chair, ARDC
Kostas Repanas	BE	European Commission DG RTD
Leyla Garcia	DE	ZBMED
Maggie Hellström	SE	Lund University
Makx Dekkers	ES	Editor team
Mari Elisa Kuusniemi	FI	University of Helsinki
Milan Ojsteršek	SI	University of Maribor
Mustapha Mokrane	NL	DANS
Nick Juty	GB	University of Manchester, ELIXIR-UK
Nichola Burton	AU	Australian Research Data Commons
Nicolas Loozen	BE	PwC, Editor team
Oya Deniz Beyan	DE	EOSC FAIR WG & FAIRplus CMMI
Patricia Herterich	GB	Digital Curation Centre
Rob Hooft	NL	Dutch Techcentre for Life Sciences

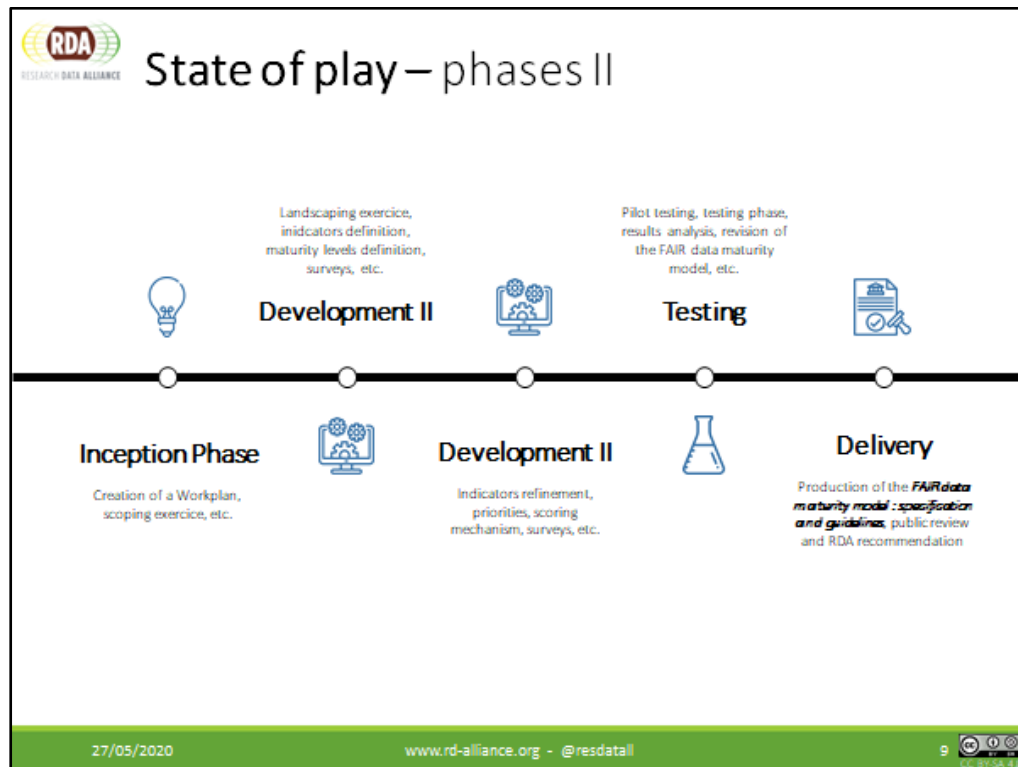
Romain David	FR	INRA
Sandor Brockhauser	DE	EuXFEL / PaNOSC
Sarah Jones	GB	Digital Curation Centre
Shelley Stall	US	Chair, American Geophysical Union
Sophie Hou	US	U.S. Geological Survey
Susanna-Assunta Sansone	GB	University of Oxford
Yan Grange	NL	ASTRON

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



# Content<sup>1</sup>

1. The Chairs opened the workshop, welcomed the participants and addressed the agenda. The approach to the Working Group was again presented. As usual, the Chairs insisted that despite all the challenges arising when designing indicators, the purpose of the WG was **NOT** to re-design the FAIR principles. As there are currently different interpretations of what the FAIR principles entail, the primordial goal is to build a common understanding.
2. The Chairs and the editor team introduced themselves, following, the participants were kindly invited to say where they come from and what are their roles in their organisation via the chat window.
3. The editorial team reported on the current state of development: what steps have been taken and what steps remain to be taken. The last two steps to be undertaken are the finalisation of the *RDA FAIR data maturity model: specification and guideline* following the feedback received during the public review period and the submission of the document for endorsement as RDA recommendation.



4. The public review period which ended on May 13th generated more than 3600 pages views and comments from 14 different persons.

<sup>1</sup> Please note that some of the slides are displayed for information purposes. The full presentation can be accessed via the RDA FAIR data maturity model WG web page.

Due to the arrival of several comments on the eve of the closing day, some of the comments have not yet been answered. Despite not having had an answer, these comments have been taken into consideration and classified according to the scheme presented above.

As outlined in the introduction, the editorial team reported on the comments received during the [public review period](#).

Comments that led to amendments in the document	
Category	Sub-category
Addition of a section implementation to the section Framework	<ul style="list-style-type: none"> <li>• Clarification on the scenarios that have not been validated by the WG (e.g. in the context of data-related algorithms, tools, workflows, protocols and other data-related services)</li> <li>• Clarification of the requirement in the FAIR principles for persistent identification of both metadata and data</li> <li>• Clarification that not all of the indicators are equally important or relevant for every community</li> </ul>
Clarification in the section on Evaluation methods	<ul style="list-style-type: none"> <li>• Different scenarios depending on when evaluation is done: before of after data has been created</li> <li>• Necessity to highlight the importance of the responsible and careful implementation of the indicators to minimize unintended consequences</li> </ul>
Additions to the section Future maintenance	<ul style="list-style-type: none"> <li>• Taking into account community perspective</li> <li>• Developing use cases</li> <li>• Including examples of supporting technologies</li> </ul>
Editorial changes, minor rephrasing, correction of examples	
Comments that were identified as beyond the scope of the FAIR data maturity model	
<ul style="list-style-type: none"> <li>• Issues beyond the FAIR principles (e.g. versioning, DMP, data deletion, quality)</li> <li>• Change requests as for the indicators (e.g. because overlaps, should be merged, etc.)</li> <li>• Change requests as for the priorities (the WG already reached a consensus)</li> <li>• Requests for clarifications concerning elements such as "what is a metadata file", "what is a data file", etc.</li> </ul>	

5. As part of the recommendation process for RDA output, the FAIR data maturity model: specification and guidelines document was proposed for a 1-month public review period. In parallel, the editorial team gave the names of two early adopters of the – to be – recommendation to the RDA secretariat. The early adopters were questioned about their adoption of the FAIR data maturity model by the RDA secretariat. These early adopters were invited to present their experiences with the adoption so far.

Ge Peng, scientist at the Earth and Science Institute between NOAA and NC State University reported on how they have used the indicators developed as part of this WG for US environment data. The NCEI has various data sources (national and international) and the volume of data is expected to be multiplied by six in the next ten years. Additionally, the NCEI provides an authoritative environment to data and provides information to government agencies and the public. Consequently, the NCEI has to demonstrate compliance with federal requirements.

Adopting the FAIR data maturity model served two purposes. The first being to baseline the FAIRness of the managed data and the second to identify the potential gaps and define the path forward for NCEI data sharing practices.

The current way to evaluate their resources looks at various elements (e.g. data online, granules metadata, Data Stewardship Maturity Matrix, etc.) and classifies the resources according to their compliance to the enunciated elements in three tiers. Most of their resources can be classified as top tier (i.e. gold).

They have used the version 0.04 of the FAIR data maturity model indicators to evaluate the FAIRness of some datasets. Levels of compliance were associated with each and every one of the indicators. Average results were calculated per priority but also similarly per FAIR area. Overall, the resources evaluated did score good, but Reusability was lagging behind (mainly due to the fact that reuse licences are not captured in the metadata record). From that exercise, some good practices were identified (e.g. presence of a DOI, well curated metadata, catalogued metadata).

NCEI & NOAA are working together to include data usage licenses and are looking into adapting the creative common licence for their needs. Furthermore, they want to extend the scope of the evaluation to additional datasets.

Following the presentation, some questions were asked by the participants.

**Q1:** *How do you define FAIR vocabularies?* For the metadata, it is done by keywords. Utilisation of community standards, master directory for sensors and platforms as well as standard names (notation climate compliant). Most of the standards are accessible, findable, open and reusable.

**Q2:** *Did the evaluation process make a distinction between FAIR for human and FAIR for machines?* For machines, the evaluation is binary; either you comply or you don't. The upside is that you have a way to better automate your evaluation. On the opposite side, a human can integrate knowledge (e.g. if you are in the middle of an implementation, you will be able to capture that information and share it with the user, which gives a better indication of where you stand)

The audience debated the concept of PID redirecting to a landing page, which ultimately prevents machine to machine interoperation. Furthermore, for some, both metadata and data should have separate PIDs, for others not. Similarly, someone observed that the FAIR principles suggest identifiers for metadata and data, but having both set as essential indicators makes that mandatory when many communities may just have one PID.

Anusuriya Devaraju and Hervé L'Hours shared their impression and experience with the adoption on behalf of FAIRsFAIR. FAIRsFAIR is fostering data practices in Europe. It is one of the projects working towards EOSC with a very clear focus on FAIR data.

FAIRsFAIR involvement in the RDA WG is mainly through FAIR certification. FAIRsFAIR is evaluating repositories to enable FAIR data (i.e. FAIR certification of repositories). Once the data is published in repositories, there is a continuous support through reuse.

CoreTrustSeal certification defines sixteen repository requirements against which an applicant will perform a self-evaluation and there is also a peer review model that comes afterwards. FAIRsFAIR is offering support to this CoreTrustSeal certification from a FAIR perspective by mapping object characteristics to where repositories can enable FAIR. To identify how repositories enable FAIR (and to provide evidence for this), there is a need to go beyond the object data and metadata and look at repository process metadata and other business information.

The long-term goal is to integrate object evaluation into the outcomes of the repository assessment / certification. As there is a clear attempt to separate repository assessment and object assessment, the goal is to develop a CoreTrustSeal + FAIR model (i.e. unified approach) to unite these two perspectives.

FAIRsFAIR adapted the recommendation in the context of two use cases; i) developing a self-assessment tool to educate and raise awareness on making data FAIR before depositing in repositories and ii) enabling a periodic programmatic assessment of datasets already published in data repositories.

They suggest to anybody who is planning to define criteria to evaluate FAIRness to start from the RDA recommendation. The recommendation currently describes the 'WHAT' aspect (i.e. definition of indicators) but also ventured to define the 'HOW' aspect (i.e. assessment details). Anusuriya Devaraju made the observation that the 'HOW' aspect is largely dependent on the context (e.g. data type, restricted data, etc.). She suggested to include potential technologies and services, but also document possibilities for further development.

Anusuriya Devaraju suggested some other improvement points (e.g. definition of FAIR compliant vocabularies) but also remarked that there is no essential indicator for Interoperability, which necessitates further work. It was remarked by a member of the audience that the fact of not having essential indicators pleads for the fact that the priorities are suggestions. Furthermore, the fact that communities are not ready for Interoperability sheds light on other communities which can be less ready than others for some aspects marked as essential.

Another suggestion was to define a plan to allow future adopters to feed information back into the working group. Additionally, a few comments were made with regards to the implementation of the indicators. (1) different practices of identifying and locating 'objects', (2) Indicators and priorities may be changed / extended depending on community practices, users (evaluators) and at which stage of the data cycle the assessment is performed and (3) the need of a transparent approach to communicate the FAIR assessments results to different stakeholders in a meaningful way.

Following the presentation, some questions were asked by the participants. Related to the last observation "how to communicate meaningful results", one participant mentioned that it is the reason for this community exercise – with the guidelines – to capture results and represent that quality information consistently to users.

6. Once the amendments are made, the immediate next step is: submission of final version to RDA Council for endorsement of the FAIR data maturity model: specification and guidelines as an RDA recommendation

7. Provided that the proposition for recommendation will be endorsed, the editorial team invited all the participants to actively promote and distribute the recommendation to their communities. The end of development milestone has been reached and now comes the implementation. Feedback loops will be initiated to allow adopters to be assisted during their implementation of the recommendation. Simultaneously, the feedback received will be aggregated and transformed into 'change requests', which are to be addressed when the recommendation will be revised.
8. As the RDA Working Group is coming to an end, at least in its current form, the editorial team took a moment to acknowledge the members whose contributions were significant to the development of the FAIR data maturity model.



The slide features the RDA logo in the top left corner. The main title is 'Acknowledgment'. Below the title, it says 'SPECIAL THANKS to the following members for their unfailing commitment and critical contribution to this WG'. A list of 20 names is presented in two columns, each preceded by a right-pointing chevron. At the bottom, there is a green bar containing the date '27/05/2020', the website 'www.rd-alliance.org - @resdata', the slide number '43', and a Creative Commons license icon.

**RDA**  
RESEARCH DATA ALLIANCE

## Acknowledgment

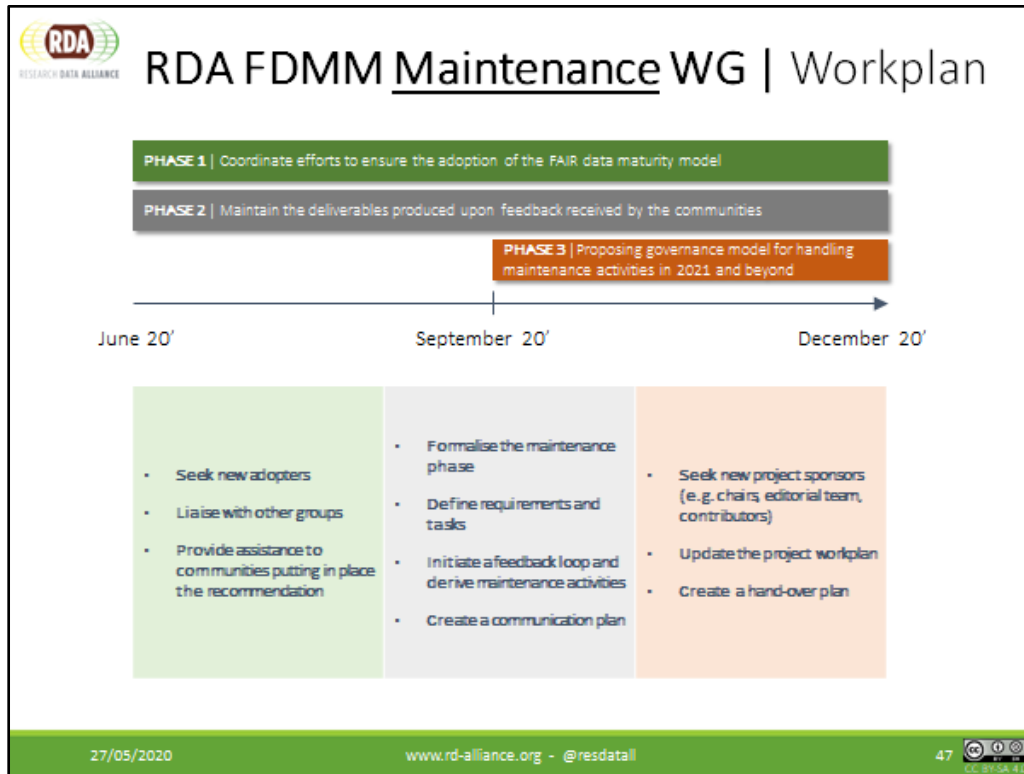
**SPECIAL THANKS** to the following members for their unfailing commitment and critical contribution to this WG

- > Alejandra González Beltrán
- > Alicia Fátima Gómez Sánchez
- > Andras Holl
- > Anusuriya Devaraju
- > Barbara Sierman
- > Carole Globle
- > Françoise Genova
- > Ge Peng
- > Gerry Coen
- > Helen Parkinson
- > Hervé L'hours
- > Keith Jeffery
- > Kerry Levett
- > Kevin Long
- > Jean-Eudes Hollebecq
- > Jolanda Strubel
- > Jonathan Petters
- > Leyla Garcia
- > Marco Molinaro
- > Maggie Hellström
- > Mark Wilkinson
- > Marta Teperek
- > Michel Dumontier
- > Nichola Burton
- > Nick Juty
- > Mustapha Mokrane
- > Oya Deniz Bayan
- > Peter McQuilton
- > Rob Hooft
- > Romain David
- > Susanna Sansone
- > Yann Le Franc

27/05/2020      www.rd-alliance.org - @resdata      43      CC BY-NC-ND

9. As it has been previously decided, the Working Group will be turned into a Maintenance Working Group with the purpose to support and maintain the adoption of the recommendation. Three main phases have been identified as illustrated on the figure below.





## Follow-up action plan

Working Group members are invited to:

- Reach out to their communities as for the publication of the FAIR data maturity model: specification and guidelines (i.e. RDA recommendation)
- Continuously provide feedback to the editorial team and pass on information with regards to the use of the FAIR data maturity model: specification and guidelines (i.e. RDA recommendation)

The editorial team will look into a release calendar and change management schedule. The next workshop will take place online and at the end of September. Lastly, the agenda and connection details will be shared soon through the usual channels.