Workshop report

FAIR data maturity model Working Group

Online meeting #3 - 18 June 2019

Project	RDA FAIR data maturity model working group	Date & Time	18 June 2019 07:00—08:30 UTC 18 June 2019 15:00—16:30 UTC
Type	Online meeting	Location	Online GoToMeeting
Meeting Chairs	Edith Herczog Keith Russell	Issue date	24/06/2019

Objectives

The objective of the third workshop was to discuss the consolidation of the contributions of the WG to the collaborative document. Consolidated indicators and maturity levels were brought before the WG in order to foster a discussion about their accuracy and coverage. In addition, this meeting served to present the current status and the next steps.

Agenda

1. Welcome, objectives of the meeting
2. Roundtable
3. State of play
4. Development | First phase
5. Development | Second phase
6. Action items and next steps

Useful links

[RDA FAIR data maturity model WG](#)
[RDA FAIR data maturity model Case Statement](#)
[Workshop #3 material (Presentation, chat logs and report)](#)
[RDA FAIR data maturity model GitHub](#)
[RDA FAIR data maturity model Collaborative document](#)

Participants

The workshop was attended by 20 participants in the morning session and 26 participants attended the session in the afternoon. Below you can find the names of participants. Please note that this list may not be complete.

Name		Affiliation	Presence 0700 UTC	Presence 1500 UTC
Alejandra Gonzalez-Beltran	UK	Oxford e-Research Centre		•
Andrea Perego	BE	European Commission, Joint Research Centre (JRC)		•
Angus Whyte	UK	Digital Curation Center		•
Anne Cambon-Thomsen	FR	CNRS, University of Toulouse		•
Athanasios Karalopoulos	BE	European Commission DG RTD	•	
Barbara Sierman	NL	KB National Library of the Netherlands	•	
Brian Matthews	UK	STFC		•
Chris De Loof	BE	Belspo		•
Christine Laaboudi	LU	EU Open Data Portal		•
Christophe Bahim	BE	PwC, Editor team	•	•
Ebtisam Aharbi	UK	PhD, University of Manchester		•
Edith Herczog	BE	Chair, Vision & values SPRL		•
Frans Huigen	NL	DANS	•	
Ge Peng	US	North Carolina State University/NCEI		•
Gilles Ohanessian	FR	French National Centre for Scientific Research	CNRS		•
Hannah Calckins	US	Children's Hostipal of Philadelphia		•
Harri Hirvonsalo	FI	CSC - IT Center for Science		•
Ilona Von Stein	NL	DANS	•	
Jacquelijn Ringersma	NL	Wageningen University and Research		•
Jerry de Vries	NL	DANS	•	
Jonathan Petters	US	Virginia Tech		•
Keith Jeffery	UK	Keith G Jeffery Consultants	•	
Keith Russell	AU	Chair, ARDC	•	
Konstantinos Repanas	BE	European Commission DG RTD	•	•
Laurence Mabile	FR	Toulouse University	•	
Lesley Wyborn	AU	ARDC, NCI, AuScope, AGU FAIR Project	•	
Maggie Hellström	SE	Lund University		•

Makx Dekkers	ES	Independent Consultant, Editor team	•	•
Mark Wilkinson	ES	Universidad Politécnica de Madrid		•
Marie-Christine Jacquemot	FR	Data Librarian	•	
Matthew Viljoen	NL	Infrastructure Operations Manager, EGI Foundation		•
Melanie Imming	NL	Imming Impact		•
Mohamed Yahia	FR	Inist-CNRS / Datacite		•
Mustapha Mokrane	NL	DANS		•
Nick Juty	UK	University of Manchester, ELIXIR-UK		•
Nicolas Loozen	BE	PwC, Editor team	•	•
Oya Beyan	DE	EOSC FAIR WG & FAIRplus CMMI	•	
Rob Hooft	NL	Dutch Techcentre for Life Sciences	•	
Romain David	FR	INRA	•	
Ronald Cornet	NL	University of Amsterdam	•	
Simon Lambert	UK	UKRI-STFC / H2020 FREYA	•	
Zsuzsanna Szeredi	BE	Vision & Values	•	
			20	*26*
			42	
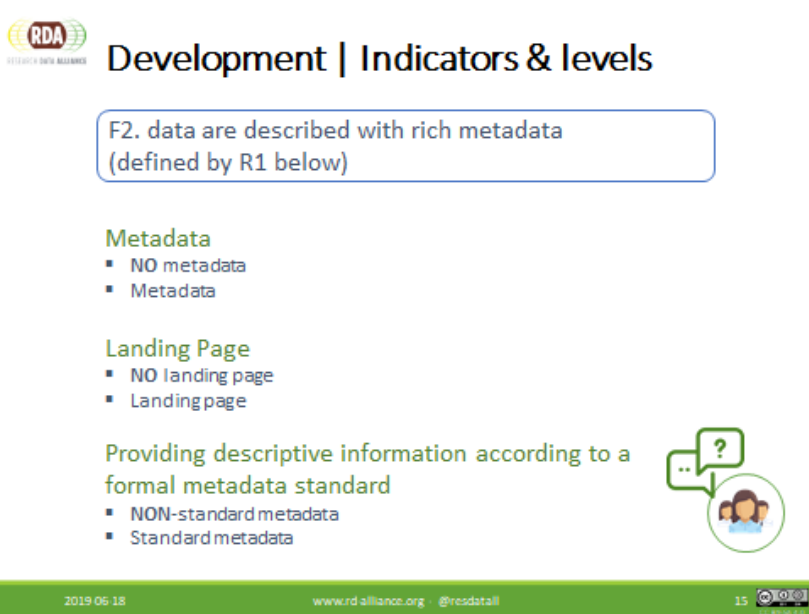
Content

The workshop was designed in order to be as interactive as possible: to do so the attendees were given the opportunity to speak up whenever necessary. The meetings were fruitful and enabled lively discussions.

The report below includes the results of both sessions.

1. The Chairs opened the workshop, welcomed the participants and addressed the agenda. The approach to the Working Group was again presented:

   a. Challenges rising from the different interpretations of FAIRness
   b. Bringing together the relevant stakeholders to discuss and build on existing expertise and different approaches
   c. Intended results

2. The Chairs and the editor team introduced themselves, following; the participants were kindly invited to share where they are from and what their roles are in their organisation via the chat window.

3. The editorial team reported on the current state of development: what part of the methodology is being addressed at the moment and where we stand.

4. Later, the editorial team walked the Working Group through the consolidated indicators and maturity levels per principle. After each principle, feedback was sought from the Working Group.

F1. (meta)data are assigned a globally unique and eternally persistent identifier	
Proposed indicators	
Discussion about the Principle	- What could be the definition of 'Unique' and 'Persistent'? - The main requirement for identifiers are that they are persistent and unambiguous. - It was proposed to change 'Global' to 'Universal'. - Resolvability of the persistent identifier should be part of A1 rather than F1 as resolution of identifiers is not about Findability. - 'Resolve' should be renamed 'Resolvability'. - Sometimes some identifiers do not resolve, and if they do 'what they resolve to' may not be the same in all cases; for example, sometimes an identifier points to a digital proxy, such as landing page. - The indicators need to be duplicated, as both the data and the metadata are in scope for the principle. - The ultimate goal is to identify an object in a given state.
Proposed resolution(s)	1. *The 'Resolve' indicator will be moved over to A1* 2. *Indicators will be rephrased and further described*
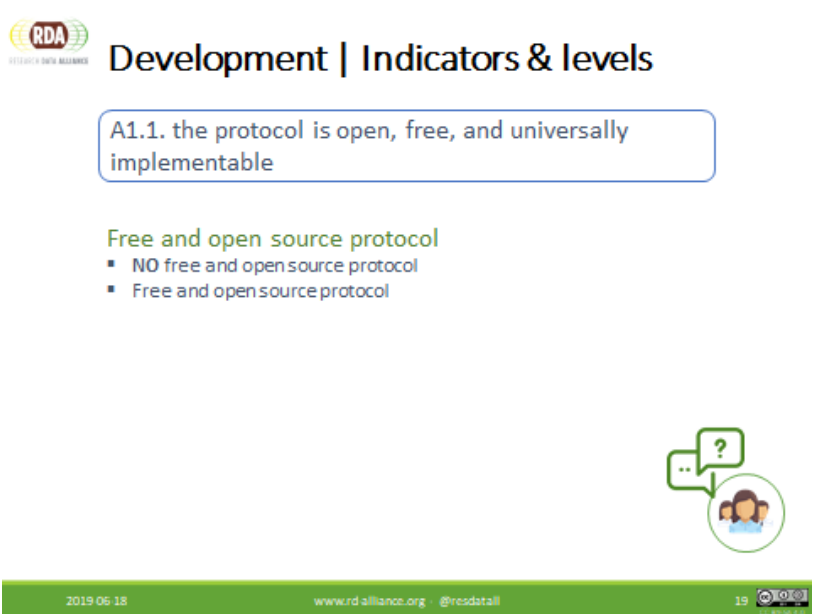
F2. data are described with rich metadata (defined by R1 below)

Proposed indicators	
Discussion about the Principle	- The landing page is targeted to human users, and cannot be interpreted automatically as there is no guarantee that it contains structured data. It also introduces an extra step between the identifier and the asset. However, DOIs point often to a landing page. - F2 is about the richness of metadata for discoverability. Rich metadata for reuse are covered under the principles for the area Reusable. - The contextualisation is an important part of F2. - 'Rich' metadata is a property of degree rather than being a binary YES/NO. There is a possibility that a digital object satisfies all the indicators proposed yet the metadata could not be qualified as 'Rich'. - Additional indicators for 'Rich' metadata may have to be defined by the context and the domain. However, a set of core indicators for 'Rich' metadata should be universal and cross-domain.
Proposed resolution(s)	1. *Start a separate discussion on requirements for metadata* 2. *Seek liaison with the RDA Metadata IG.*

F3. metadata clearly and explicitly include the identifier of the data it describes	
Proposed indicators	
Discussion about the Principle	- The structure of identifiers may also be considered.
Proposed resolution(s)	*1. The indicator will be rephrased and further described*

F4. (meta)data are registered or indexed in a searchable resource	
Proposed indicators	**Development | Indicators & levels** F4. (meta)data are registered or indexed in a searchable resource **Harvested by search engine** ▪ NOT harvested by a search engine ▪ Harvested by a search engine **Providing metadata to specific portals** ▪ Metadata NOT indexed in specific portals ▪ Metadata indexed in specific portals **Institution repositories** ▪ (meta)data NOT present in institution repositories ▪ Presence of the (meta)data in institution repositories 2019 06 18 www.rd alliance.org · @resdatall 17
Discussion about the Principle	- Clarifications are needed for the portals (e.g. domain specific aggregator?) and for the repositories. - Not all the proposed indicators are needed. - Possibility to include some notions of what constitutes a trusted (or managed) repository. - Something could be indexed, but it doesn't mean that someone can do anything with the information that is associated with the index entry. - The aspect of universal accessibility of the search system needs to be considered.
Proposed resolution(s)	*No further action will be taken*

A1. (meta)data are retrievable by their identifier using a standardised communications protocol	
Proposed indicators	
Discussion about the Principle	- The protocol has to be transparent (i.e. people should be able to find out what it is). - Manual access via the researcher needs to be clarified and broadened (e.g. through a data access committee, data stewards, etc.). - 'Standardised' should be part of the indicators.
Proposed resolution(s)	*1. The indicators will be rephrased and further described*

A1.1. the protocol is open, free, and universally implementable	
Proposed indicators	**Development | Indicators & levels** A1.1. the protocol is open, free, and universally implementable Free and open source protocol ▪ NO free and open source protocol ▪ Free and open source protocol 2019 06 18 www.rd alliance.org · @resdatall 19
Discussion about the Principle	- The indicator should be split into two indicators. One about a free protocol and the other one about an open source protocol. - An example is the Digital Object Interface Protocol (DOIP). - Access only by telephone is a barrier to machine actionability. - There may never be a machine-actionable protocol for access to some types of data, in which case a transparent phone-based protocol must be acceptable.
Proposed resolution(s)	1. *The current indicator will be split into two different indicators: Free and Open Source.*

A1.2. the protocol allows for an authentication and authorization procedure, where necessary	
Proposed indicators	
Discussion about the Principle	- The 'where necessary' part is important here; the indicator would be optional for cases where the data is openly available. - The text does not read correctly; it's not the protocol that is authenticated but the protocol should allow authentication. - There should also be an indicator about 'authorisation'.
Proposed resolution(s)	1. *The indicator will be rephrased* 2. *A new indicator will be proposed*

A2. metadata are accessible, even when the data are no longer available	
Proposed indicators	
Discussion about the Principle	- It is important to consider who is the owner of the (meta)data (e.g. if there is a curation partner, it is another story). - If data is lost, chances are big that the metadata will be lost too; however, this would be contrary to FAIR principles.
Proposed resolution(s)	*No further action will be taken*

I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation	
Proposed indicators	
Discussion about the Principle	- There is a slight difference between machine-readable and machine-understandable. The latter is needed for autonomous processing. That point could include a subset of indicators for actionability, readability, understandability, interpretability. - What is an acceptable language for knowledge representation? - A formal syntax (e.g. XML) should be one of the levels of the 'format' indicator. - Community-standard format and proprietary format might overlap. If the indicator is maintained, there is still a difference for data and metadata. - It might be better to distinguish 'syntax' versus 'semantic'. - 'Open' should replace 'Community-standard' as opposed to 'Proprietary'. - In some disciplines, researchers only use proprietary formats. If the indicator remains, it means that some of the digital objects would never become FAIR. There may be a preference for open formats, but it is simply not always available.

Proposed resolution(s)	1. *Start a separate discussion thread to determine a reasonable set of indicators related to knowledge representation languages.*

I2. (meta)data use vocabularies that follow FAIR principles	
Proposed indicators	
Discussion about the Principle	- It is not explicit which of the FAIR principles should be tested, and how, to determine if (and how much) a (meta)data vocabulary complies with FAIR principles. - It is not explicit how to measure the FAIR compliance of vocabularies. - Apply the maturity indicators to the vocabulary itself
Proposed resolution(s)	1. *Consider a set of indicators to test FAIRness of vocabularies in the context of the wider discussion on knowledge representation proposed under I1.*

I3. (meta)data include qualified references to other (meta)data	
Proposed indicators	
Discussion about the Principle	- The qualification cannot be tested logically. For that reason, it should be considered in the context of the discussion on 'rich metadata'. - Since vocabularies and terms change should this not have a temporal range for validity? - General qualification is meant as unspecific links (like Web links). Contributor is already a specific qualification.
Proposed resolution(s)	1. *The indicators will be rephrased.*

R1. meta(data) are richly described with a plurality of accurate and relevant attributes	
Proposed indicators	
Discussion about the Principle	- The RDA Metadata IG working on relevant attributes to include in metadata. This WG could liaise with them and further build the indicators related to metadata on their work. - 'Richly described with a plurality of accurate and relevant attributes' is relative - so how can it be tested? - Richness needs to be defined by the community taking into account the needs of the potential (re)users. - The question concerns whether there are sufficient attributes (rich metadata) presumably with formal syntax and declared semantics to allow re-use.
Proposed resolution(s)	1. *Further indicators could be proposed based on cooperation with the RDA Metadata IG*

R1.1. (meta)data are released with a clear and accessible data usage licence	
Proposed indicators	
Discussion about the Principle	- The licence should be easily located in the metadata.
Proposed resolution(s)	*No further action will be taken*

R1.2. (meta)data are associated with detailed provenance	
Proposed indicators	
Discussion about the Principle	- Context information is needed (domain-specific). - Provenance is critical; the (re)user needs to know who the author is and how to reach him/her. Provenance includes information on how the dataset was generated (calibration, methodology etc.), source and lineage, versioning, project and/or activity in the framework of which the data was produced. - The end-user (or a software agent acting on their behalf) needs to know if there is machine-readable or machine-understandable provenance information since this is essential for contextualisation (relevance, quality) of the asset. - R1.2 is very important for the long term.
Proposed resolution(s)	1. *Consider the metadata needed for provenance in the separate discussion thread on 'Rich metadata'*

R1.3. (meta)data meet domain-relevant community standards	
Proposed indicators	
Discussion about the Principle	- It was proposed that the indicator could refer to a schema rather than a template.
Proposed resolution(s)	*1. The indicator will be rephrased*

X. Choices beyond FAIR	
Proposed indicators	
Discussion about the Principle	- Versioning and keywords should go under 'rich metadata'. - Consideration of multilingual vocabularies should be part of the discussion on knowledge representation.
Proposed resolution(s)	*No further action will be taken.*

5. The editorial team presented the second phase of the development of core assessment criteria to evaluate and compare FAIRness. It consists of looking at the level scale of indicators and possibly weigh them. Then, different aspects were presented:
    a. FAIRness report for a resource under evaluation (Indicators classified per importance)
    b. FAIRness score per principle [to which the indicator pertain]
    c. FAIRness score for the FAIR areas
    d. FAIRness score across the FAIR areas, possibly?
    e. Documentation of the results

Feedback with regard to the second development phase
We need to look at the capability of the resource in terms of Findability, Accessibility, Interoperability and Reusability. Instead of measuring FAIRness per se, we should measure the 'contractual capabilities' of the resource (i.e. what does it offer?)

General remarks

During the workshop, the participants brought up a few generic topics. They are listed are together with the actions taken to address them, if any.

Topic	Status
It is important to note that depending on the format of the assessment - manual versus automated - the nature of the indicators may vary. A further comment is 'how well a machine or a human can assess' a digital object.	To be further discussed.
Quality should not be discussed as part of the indicators. According to GO-FAIR "quality issues are not addressed by the FAIR principles. The data's reliability lies in the eye of the beholder and depends on the intended application." Quality needs to be considered as part of the data curation process or Data Management Plan.	No further action.
In a new iteration, GO FAIR is trying to explain what generic metadata is according to the different communities	GO-FAIR to inform the Working Group about the results of their further work on this.
The indicators also need to be considered in the context of roles and responsibilities. For example, a	To be considered in the context of linking indicators to the roles of the

researcher will need to consider different indicators than a data curator or repository owner.	persons/organisations responsible for satisfying the requirement.

Further discussion

During discussions, two main issues were identified that need further discussion:

1. Which indicators should be related to 'rich metadata'? There are different requirements for metadata for discovery (for a user to be able to find relevant data) and for reuse (for a reuser to be able to understand how the data was produced and how it can be reused). This discussion may benefit from a liaison and cooperation with the RDA Metadata IG.
    A GitHub issue has been opened to gather opinions and proposals for this topic.

2. What should be expected from knowledge representation systems in terms of syntax and semantics? How can knowledge representation systems (code lists, controlled vocabularies, ontologies) help or hinder FAIRness?
    A GitHub issue has been opened to gather opinions and proposals for this topic.

Any other discussion item can be contributed by opening an issue on GitHub at https://github.com/RDA-FAIR/FAIR-data-maturity-model-WG/issues.

Follow-up Action Plan

Provide feedback to the proposals presented at the meeting of today on the GitHub, if at all possible, **by the 30th June**

Contribute more indicators and maturity levels on Google Sheet, **until the 31st of August**
- Analysis of all the FAIR principles
    - FAIR – Findable   [Link]
    - FAIR – Accessible   [Link]
    - FAIR – Interoperable   [Link]
    - FAIR – Reusable   [Link]

Share ideas about consolidation and weighting of indicators and maturity levels on the GitHub

---

The next and **fourth workshop** will take place online

0700 UTC on the **12th of September 2019 (local times)**
1500 UTC on the **12th of September 2019 (local times)**

Link: **https://global.gotomeeting.com/join/560494093**