FAIR Data Maturity Model

Workshop #7
13th February 2020
## Agenda

<table>
<thead>
<tr>
<th>Time</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>5’</td>
<td>Welcome, objectives of the meeting</td>
</tr>
<tr>
<td>5’</td>
<td>Roundtable</td>
</tr>
<tr>
<td>5’</td>
<td>State of play</td>
</tr>
<tr>
<td>5’</td>
<td>Testing phase overview</td>
</tr>
<tr>
<td>25’</td>
<td>Testing insights – feedback</td>
</tr>
<tr>
<td>25’</td>
<td>Testing insights – general discussion</td>
</tr>
<tr>
<td>15’</td>
<td>Potential scoring mechanisms</td>
</tr>
<tr>
<td>5’</td>
<td>Action items and next steps</td>
</tr>
</tbody>
</table>
Context

The principles are **NOT** strict
- **Ambiguity**
- Wide range of interpretations of FAIRness

Different **FAIR Assessment** Frameworks
- Different metrics
- No comparison of results
- No benchmark

**SOLUTION** is to bring together **stakeholders** to build on **existing approaches** and **expertise**
- Set of **core assessment criteria** for FAIRness
- FAIR **data maturity model & toolset**
- FAIR **data checklist**
- RDA recommendation

Join the **RDA Working Group**: [RDA WG web page](#) | [GitHub](#)
Objectives

What are to be evaluated to determine FAIRness?

- Identify the indicators that can serve as core criteria
- Propose guidelines and a checklist
- Test the core criteria
- Enable the development of automated tools for evaluation
- Update the core criteria based on feedback

FAIR data maturity model
BUT the Working Group does **NOT** have the purpose to ...

- **develop yet-another-evaluation-method**: the core criteria are intended to provide a common ‘language’ across evaluation approaches, not to be applied directly to datasets.

- **define how the core criteria need to be evaluated**: The exact way to evaluate data based on the core criteria is up to the owners of the evaluation approaches, taking into account the requirements of their community.

- **revise and re-design the FAIR principles**
Roundtable

In the chat window, please type...

- Your name
- Your affiliation
- Your role
  - Researcher
  - Librarian
  - Service provider
  - Policy maker
  - Funder

- Introducing the editorial team
State of play
State of play

1. Definition
   - DONE

2. Development
   - i) First phase
     - DONE
   - ii) Second phase
     - DONE
   - DONE

3. Testing
   - ONGOING

4. Delivery
   - ON HOLD

* Any comments are still welcomed with regards to the output produced during the first phase | GitHub
State of play

Proposition
- Indicators
- Maturity levels

Discussion | Indicators
- Validation (YES/NO)
- Missing indicators

Discussion | Prioritisation
- Approach to prioritisation
- Priority levels
- Survey

Discussion | Scoring
- Approach to scoring

Consolidation
- Indicators
- Maturity levels

Testing

- Scoping
- Approach
- Methodology
- Landscaping exercise

Editorial team
Working group

2020-02-13
Goal is to **finalise** indicators and priorities

Indicators and priorities will be further **used in their current state**

Indicators and priorities **will be re-evaluated** after the testing phase
Continuity

2020

January  | February  | March  | April  | May  | June  

Testing phase

FAIR data maturity model maintenance (Guidelines, checklist & indicators)

RDA recommendation

End of RDA WG  | Workshop  | Deliverable
Testing phase
Testing phase | Overview

- **December 2019**
  - Pilot testing
  - Early results

- **January 2020**
  - Testing phase
  - 1st level of testing (i.e. comparing indicators against methodologies)

- **February 2020**
  - Aggregating feedback
  - Request for changes
  - General issues

- **March 2020**
  - Second run of tests
  - Feedback integration in the FAIR data maturity model
Thanks to all testers for their contribution

- 13 volunteers having different affiliations
- Various range of disciplines and entities
- Different approaches to the scoring

<table>
<thead>
<tr>
<th>Discipline / Domain</th>
<th>Affiliation / Tool</th>
<th>Tester</th>
<th>Entity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Earth Science</td>
<td>NCEI of NOAA</td>
<td>Ge Peng</td>
<td>Dataset</td>
</tr>
<tr>
<td>Engineering &amp; Technical sciences</td>
<td>4TU.ResearchData</td>
<td>Egbert Gramsbergen, Paula Martinez-Lavanchy, Madeleine de Smaele, Manta Teperek</td>
<td>Dataset</td>
</tr>
<tr>
<td>Humanities, Spatial, Health, etc.</td>
<td>FAIRsFAIR</td>
<td>Anusuriya Devaraju</td>
<td>Methodology</td>
</tr>
<tr>
<td>Human-Environment Observatories (OHMs)</td>
<td>DRIIHM infrastructure</td>
<td>Romain David &amp; Emilie Lerigoleur</td>
<td>Methodology</td>
</tr>
<tr>
<td>Biology</td>
<td>ODAM information system</td>
<td>Romain David &amp; Daniel Jacob</td>
<td>Dataset</td>
</tr>
<tr>
<td>Agronomic &amp; Biomedical</td>
<td>Agroportal</td>
<td>Romain David, Clément Jonquet &amp; Emma Amdouni</td>
<td>Ontology</td>
</tr>
<tr>
<td>ALL disciplines / domains</td>
<td>ARDC FAIR self-assessment tool</td>
<td>Kerry Levett &amp; Nichola Burton</td>
<td>Methodology</td>
</tr>
<tr>
<td>Humanities &amp; Social Sciences</td>
<td>DRI</td>
<td>Kathryn Cassidy &amp; Natalie Harrower</td>
<td>Dataset</td>
</tr>
<tr>
<td>Astronomy</td>
<td>CDS</td>
<td>Françoise Genova</td>
<td>Dataset</td>
</tr>
</tbody>
</table>
Comments on indicators

- There are (too) many indicators. However, others note that this level of granularity is useful, because it helps to think about all the aspects.
- Testing the indicators provided suggestions for improving existing evaluation approaches or existing standards.
- Some indicators depend on data characteristics (public, embargoed, private), domain practice and type of repository (domain-specific vs. domain-agnostic), ‘perimeter scale’ (infrastructure vs. dataset).
- Distinguishing indicators for metadata separate from indicators for data does not work for resources with embedded metadata.
- Overlap between indicators (e.g. across principles F1/A1 and F2/R1).
- Some indicators are conditional, e.g. the ones on authentication, authorisation, references and consent – if not applicable, they should not ‘count’.
- Several indicators require compliance with community standards, but who defines them?
- If data is an ontology, a different set of indicators or priorities may be needed.
FAIR principles are aspirational and ambitious, aiming at full machine-understandability, but current practices are not well aligned at this point in time.

Identification: there are various comments, some favour identification of metadata over identification of data, others data over metadata, others see both as equally essential, but there is also a comment that having separate identifiers is not common practice.

There seems to be a role for landing pages and other human-readable documentation in providing information, in addition to structured metadata.

Requests for adding maturity levels > scoring.

Data comes in different granularities: whole dataset or part of dataset or individual data items (e.g. observations, concepts).

Different perspectives on metadata and how it relates to data:
- repository level / collection level / dataset / data item level metadata
- separate metadata records or embedded metadata.
• There may be a need to ‘profile’ indicators for specific cases, i.e. selecting a subset of indicators, adapting priorities, following discipline-specific guidelines.

• Noteworthy that testers are often stricter than the priorities that the WG has defined, e.g. making essential:
  o machine-understandable community standards
  o standard, open-source protocols
  o machine-understandable knowledge representation
  o standard vocabularies
  o standard reuse licences

• One tester proposes to do away with all priorities entirely.
### Testing insights | Feedback

<table>
<thead>
<tr>
<th>Comments on indicators</th>
<th>General issues</th>
<th>Specific issues</th>
<th>Information needs</th>
</tr>
</thead>
</table>

- Need for better explanation of terms, in particular ones that are vague or subjective, e.g. ‘sufficient’

- Need for better definition of terms used in FAIR principles, e.g. knowledge representation, FAIR-compliant vocabularies – to take into account [https://doi.org/10.1162/dint_r_00024](https://doi.org/10.1162/dint_r_00024)

- Need for information on best practices that may be applied to increase FAIRness, e.g. identification, protocols, licences

- These issues to be addressed in the Guidelines, using suggestions and examples provide by testers
Testing insights | General discussion

1. Even Open Data requires clarity on access conditions and reuse licence

2. Requirement for persistent identification of both metadata and data, but this is not always possible

3. Metadata standards may not include elements for all FAIR requirements (e.g. provenance, licences); role for documentation and landing pages?

4. Knowledge representation is about (1) controlled vocabularies used and (2) data models/formats
Scoring mechanism
Scoring mechanisms | Overview

5-level scale per indicator

- Five levels of compliance
- Per indicator – aggregated per FAIR area
- Non applicable or consideration/implementation as options
- Useful for giving credit for evolution and helping people to improve

FAIRNESS per area

- Measurement based on priorities
- Per indicator – aggregated per FAIR area
- Score determined based on the compliance to priorities
- Provides a ‘measure of FAIRness’

Overall FAIRNESS

- Measurement based on priorities
- Per indicator – overall score
- Aggregated score
- Provides a quick view of how priorities are met -- but does not give detailed view
Action items & Next steps
## GUIDELINES

### INTRODUCTION
- Introduction
- Objectives
- Use of the document

### FRAMEWORK
- Indicators
- Maturity levels
- Prioritization
- Indicators description

### IMPLEMENTATION
- How to evaluate
Guidelines | further development

Working Group to share **remarks** and **suggestions** about the guidelines

**Testing phase** will bring out comments and suggestions for change and for additional guidance

**Stable version of the guidelines** to be published for the next RDA plenary

https://docs.google.com/document/d/1pDGGL3-BbBJu18KlfZUI3AizKLHXGXdlI_mPtpEWmeg/
Action item and next steps

Working Group members are invited to:

- Share feedback, comments & suggestions – on the [Guidelines](#)
- Discuss proposals for changes in priorities on GitHub (issues will be created)
- Contribute to GitHub discussion on [scoring](#)

We’re also looking for [volunteers for further testing](#); please contact us!

---

**WORKSHOP #8**

15th RDA PLENARY IN MELBOURNE  
19 March 2020

11.30 – 13h00 (GTM+11) | Breakout 4
Resources

- RDA FAIR data maturity model WG
  https://www.rd-alliance.org/groups/fair-data-maturity-model-wg

- RDA FAIR data maturity model WG – Case Statement

- RDA FAIR data maturity model WG – GitHub

- RDA FAIR data maturity model WG – Collaborative document
  https://docs.google.com/spreadsheets/d/1gvMfbw46oV1idztsr586aG6-teSn2cPWe_RJZG0U4Hg/edit#gid=0

- RDA FAIR data maturity model WG – Indicators prioritisation
  https://docs.google.com/spreadsheets/d/1mkjElFrTBPBH0QViODeXur0xNGhJqau0zkL4w8RRaw/edit

- RDA FAIR data maturity model WG – Indicators prioritisation survey results
  https://drive.google.com/open?id=11hyAYCKz_NV0Ob9-vlPqjN9LCarOFmc3

- RDA FAIR data maturity model WG – Guidelines
  https://docs.google.com/document/d/1pDGGL3-BbBJu18KlfZUI3AizKLHXGXdi_mPtpEWmeg/

- RDA FAIR data maturity model WG – Mailing list
  fair_maturity@rda-groups.org
Thank you!