FAIR Data Maturity Model

14th RDA Plenary session - Workshop #5

23rd October 2019
## Agenda

<table>
<thead>
<tr>
<th>Time</th>
<th>Session Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>14:30 – 14:40</td>
<td>Welcome, objectives of the meeting</td>
</tr>
<tr>
<td>14:40 – 14:45</td>
<td>Roundtable</td>
</tr>
<tr>
<td>14:45 – 14:50</td>
<td>State of play</td>
</tr>
<tr>
<td>14:50 – 14:55</td>
<td>Development – First phase</td>
</tr>
<tr>
<td>14:55 – 15:00</td>
<td>Development – Second phase</td>
</tr>
<tr>
<td>15:00 – 15:45</td>
<td>Panel discussion</td>
</tr>
<tr>
<td>15:45– 15:50</td>
<td>Development – Next steps</td>
</tr>
<tr>
<td>15:55– 16:00</td>
<td>Next steps and Q&amp;A</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](https://www.rd-alliance.org) | [GitHub](https://github.com)
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
Scope

**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

By show of hands

➤ Which region?
➤ Your role
   ➤ Researcher
   ➤ Librarian
   ➤ Infrastructure manager
   ➤ Policy developer
   ➤ Research funder
➤ Introducing the editorial team

If you are dialing in, please type your name and affiliation in the chat window
Join at slido.com
Ask questions and vote in live polls

MEETING CODE #8935
State of play
State of play

- Scoping
  - Approach
  - Methodology
  - Landscaping exercise

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

ONGOING

Testing

Editorial team
Working group
State of play

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
Development
First Phase
Overview | Indicators & levels

- **F1** (Meta)data are assigned globally unique and persistent identifiers
- **F2** Data are described with rich metadata
- **F3** Metadata clearly and explicitly include the identifier of the data they describe
- **F4** (Meta)data are registered or indexed in a searchable resource

- **A1** (Meta)data are retrievable by their identifier using a standardised communication protocol
  - **A1.1** The protocol is open, free and universally implementable
  - **A1.2** The protocol allows for an authentication and authorisation where necessary
- **A2** Metadata are accessible, even when the data are no longer available

- **I1** (Meta)data use a formal, accessible, shared and broadly applicable language for knowledge representation
- **I2** (Meta)data use vocabularies that follow the FAIR principles
- **I3** (Meta)data include qualified references to other (meta)data

- **R1** (Meta)data are richly described with a plurality of accurate and relevant attributes
  - **R1.1** (Meta)data are released with a clear and accessible data usage license
  - **R1.2** (Meta)data are associated with detailed provenance
  - **R1.3** (Meta)data meet domain-relevant community standards

Under discussion
Provisionally agreed
Overview | Indicators & levels

Indicators for Findability

- [F1-01M] Metadata is identified by a persistent identifier
- [F1-01D] Data is identified by a persistent identifier
- [F1-02M] Metadata is identified by a universally unique identifier
- [F1-02D] Data is identified by a universally unique identifier
- [F2-01M] Sufficient metadata is provided to allow discovery, following domain/discipline-specific metadata standard
- [F2-02M] Metadata is provided for the discovery-related elements defined by the RDA Metadata IG, as much as possible and relevant, if no domain/discipline-specific metadata standard is available
- [F3-01M] Metadata includes the identifier for the data
- [F4-01M] Metadata is offered/published/exposed in such a way that it can be harvested and indexed

* The full list of indicators can be found on the following GSheet
Development
Second Phase
Weighting the indicators, developed as part of the WG, following the key words for use in RFC2119

- **Mandatory**: indicator **MUST** be satisfied for FAIRness (Essential)
- **Recommended**: indicator **SHOULD** be satisfied, if at all possible (Important)
- **Optional**: indicator **MAY** be satisfied, but not necessarily so (Useful)
PRIORITY EVOLUTION

- Mandatory
- Recommended
- Optional

Early proposition
- 12
- 13
- 8

Survey results
- 27
- 28

Notable results*

- Metadata for discovery > recommended (F2)
- Metadata for reuse > mandatory (R1)
- (Machine-understandable) knowledge representation > mandatory for metadata & recommended for data (I1)
- All references to data > optional (I3)

* Results can be accessed [here](https://slido.com/#8935)
Distribution of the weight of the indicators

**FAIR PRINCIPLES**

- Mandatory
- Recommended
- Optional

**FINDABLE**
- Mandatory: 5
- Recommended: 7
- Optional: 3

**ACCESSIBLE**
- Mandatory: 2
- Recommended: 3
- Optional: 2

**INTEROPERABLE**
- Mandatory: 2
- Recommended: 9
- Optional: 5

**REUSABLE**
- Mandatory: 1
- Recommended: 9
- Optional: 3
Indicators | Discussions

Four discussion topics to choose from, one to be addressed today:

1. Identifier to point to data or a landing page

2. Machine-processable data versus human access

3. Linking and referencing

4. Two-speed FAIRness

slido.com #8935
### Indicators | Discussions

**1** Identifier to point to data or landing page?

<table>
<thead>
<tr>
<th>A</th>
<th>OPPOSED VIEWS</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>FAIRness requires separate unique and persistent identifiers for metadata and data</td>
<td>OR</td>
<td>Data published with an identifier (e.g. DOI) pointing to a landing page with embedded metadata and a URL to access the data can also be considered FAIR</td>
</tr>
</tbody>
</table>

---

*slido.com #8935*
# Machine-processable data versus human access

<table>
<thead>
<tr>
<th>A</th>
<th>OPPOSED VIEWS</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>FAIRness requires data to be machine-processable</td>
<td><strong>OR</strong></td>
<td>Human access to data can also be considered FAIR</td>
</tr>
</tbody>
</table>
### Indicators | Discussions

#### 3 Linking and referencing

<table>
<thead>
<tr>
<th>A</th>
<th>OPPOSED VIEWS</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>FAIRness requires rich linkages between metadata and other metadata, between metadata and other data between data and other data</td>
<td>OR</td>
<td>Metadata and data that have no links to other metadata and other data can also be considered FAIR</td>
</tr>
</tbody>
</table>
### Indicators | Discussions

#### 4 Two-speed FAIRness

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>FAIRness should aim for the same level of FAIRness for all domains in the long term</td>
<td>Different communities need to be able to define their own target FAIRness levels</td>
</tr>
</tbody>
</table>

OR

**slido.com #8935**
Development

Next steps
Core assessment criteria to evaluate and compare FAIRness

- FAIRness report for a resource under evaluation
  - Indicators classified per importance
- FAIRness score per principle [to which the indicator pertain]
- FAIRness score for the FAIR areas
- FAIRness score across the FAIR areas, possibly?
- Documentation of the results
Development | Scoring*

Triple overall FAIR score and levels for FAIR areas

*Proposal discussed on GitHub

May be too crude and could be misused

OVERALL FAIRness

- **60%** Mandatory
- **70%** Recommended
- **50%** Optional

---

*Proposal discussed on GitHub*
## Development | Scoring

### FAIRness per area

<table>
<thead>
<tr>
<th>Level</th>
<th>Mandatory</th>
<th>Recommended</th>
<th>Optional</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 0</td>
<td>○</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level 1</td>
<td>●</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level 2</td>
<td>●</td>
<td>●</td>
<td>○</td>
</tr>
<tr>
<td>Level 3</td>
<td>●</td>
<td>●</td>
<td></td>
</tr>
<tr>
<td>Level 4</td>
<td>●</td>
<td>●</td>
<td>○</td>
</tr>
<tr>
<td>Level 5</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
</tbody>
</table>

○ None of the indicators are satisfied

○ Half of the indicators are satisfied

● All indicators are satisfied

---

[slido.com #8935](https://slido.com/s/8935)
Development | Guidelines, checklist and next steps

GUIDELINES

- Context
- Framework
- Indicators
  - Description
  - Examples
  - Prioritisation
- Utility and utilization
- Integration with other initiatives
- Continuity

CHECKLIST

Summary of the guidelines; focus on the key elements considered to be FAIR compliant and improve reusability
Testing the set of indicators

As presented during workshop #3, we identified two levels of testing:

1\textsuperscript{st} Level

- Test whether the indicators are aligned with the current methodologies to measure FAIRness
  
i) Indicator(s) not present in the methodology but in the core set of assessment criteria
  
ii) Indicator(s) present in the methodology but not present in the core set of assessment criteria

2\textsuperscript{nd} Level

- Owner of methodologies to test the core set of assessment criteria (i.e. Indicators with their methodology and a given dataset)

In scope for the WG

In scope for future work
Next steps
Next steps

- We encourage you to share any feedback in the GitHub
  - Indicators
  - Prioritisation
  - Scoring
  - Next steps

WORKSHOP #6
4 December 2019
09.00 - 10.30 CET | Morning session
17.00 - 18.30 CET | Afternoon session
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/1mkjElFrTBPBH0QViODexNur0xNghJqau0zkL4w8RRAw/edit

➤ RDA FAIR data maturity model WG – **Indicators prioritisation survey results**
https://drive.google.com/open?id=1hyAYCKz_NVOb9-vlPqjN9LCarOFmc3

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