FAIR Data Maturity Model

Workshop #2 [RDA 13th Plenary session]
3rd April 2019
Agenda

› Welcome, objectives of the meeting 5’
› Round table 5’
› Approval of methodology and scope 10’
› Report on discussions on GitHub 10’
› Hands-on exercise 45’
› Action items and next steps 15’
Case statement of the WG

Challenge

› Ambiguity and wide range of interpretations of FAIRness
› Lack of a common set of core assessment criteria and a minimum set of shared guidelines

Approach

› Bring together stakeholders
› Build on existing approaches and expertise

Intended results

› RDA Recommendation of core assessment criteria
› Generic and expandable self-assessment model
› Self-assessment toolset
› FAIR data checklist
Case statement of the WG

Target audiences

› Researchers, data stewards, other data professionals
› Data service owners, e.g. infrastructure, repositories
› Organisations that manage research data
› Policymakers

Connections

› RDA Disciplinary Framework Interest Group
› RDA Domain Repositories Interest Group
› Other RDA groups

Scope of the assessment

› Datasets
› Data-related aspects (e.g. algorithms, tools, workflows)
Roundtable

› Which region?
› Your role
  › Researcher
  › Librarian
  › Infrastructure manager
  › Policy developer
  › Research funder
› Introducing the editorial team
› First adopters: EOSC & European Commission
› Owners of FAIR assessment methodologies

If you are dialling in, please type your name and affiliation in the chat window.
WG methodology, timeline & scope
Proposed development methodology

Bottom-up approach comprising 4 phases

- Definition
- Development
  - Assessment of the four FAIR principles in four ‘strands’
  - Fifth ‘strand’: beyond the FAIR principles
- Testing
- Delivery
Overview of the methodology

1. Articulate objectives
2. Define stakeholders and users
3. Establish liaisons with other RDA groups
4. Identify and analyse existing approaches
5. Identify issues and additional areas of interest
6. Agree work structure and time plan
7. Consider each of the FAIR principles and their facets
8. Compare and consolidate metrics per principle
9. Identify levels per metric
10. Propose pathway for improvement per metric
11. Identify dependencies, overlaps and gaps
12. Harmonise metrics across FAIR areas
13. Identify overall maturity levels and pathways
14. Draft core assessment criteria
15. Map existing approaches to draft assessment criteria
16. Apply draft assessment criteria to selected collections
17. Compare results and improve criteria
18. Finalise core assessment criteria
19. Describe overall pathways/guidelines
20. Publish results

Legend:
- Definition
- Development
- Testing
- Delivery
Proposed timeline

**Workshop #1 [February]**
- Introduction to the WG
- Existing approaches
- Landscaping exercise

**Workshop #2 [April]**
- Approval of methodology & scope
- Hands-on exercise

**Workshop #3 [June]**
- Presentation of results
- Discussion

**Workshop #4 [September]**
- Proposals
- Proposed approach towards guidelines, checklist and testing

<table>
<thead>
<tr>
<th>M1</th>
<th>M2</th>
<th>M3</th>
<th>M4</th>
<th>M5</th>
<th>M6</th>
<th>M7</th>
<th>M8</th>
<th>M9</th>
<th>M10</th>
<th>M11</th>
<th>M12</th>
<th>M13</th>
<th>M14</th>
<th>M15</th>
<th>M16</th>
<th>M17</th>
<th>M18</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>Q2</td>
<td>Q3</td>
<td>Q4</td>
<td>Q5</td>
<td>Q6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Today
## Proposed scope

### Proposed resolutions

<table>
<thead>
<tr>
<th>ENTITY</th>
<th>Dataset and <strong>data-related aspects</strong> (e.g. algorithms, tools and workflows)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NATURE</td>
<td><strong>Generic</strong> assessment (i.e. cross-disciplines)</td>
</tr>
<tr>
<td>FORMAT</td>
<td><strong>Manual</strong> assessment</td>
</tr>
<tr>
<td>TIME</td>
<td><strong>Periodically</strong> throughout the lifecycle of the data</td>
</tr>
<tr>
<td>RESPONDENT</td>
<td>People with <strong>data literacy</strong> (e.g. researchers, data librarians, data stewards)</td>
</tr>
<tr>
<td>AUDIENCE</td>
<td>Researchers, data stewards, data professionals, data service owners, organisations involved in research data and policy makers</td>
</tr>
</tbody>
</table>
Overview of discussions on GitHub

Findable: What does it mean?
- Human Findable
- Machine Findable
- Meaning of ‘rich metadata’

‘Flows’ beyond the FAIR assessment
- Data flow
- Data flow legal issues
- People flow
- Financial flow
- Hardware infrastructure
Hands-on Exercise
Scope of the exercise

Method step 1: Articulate objectives

Method step 2: Define stakeholders and users

Method step 3: Establish liaisons with other RDA groups

Method step 4: Identify and analyse existing approaches

Method step 5: Identify issues and additional areas of interest

Method step 6: Agree work structure and time plan

Method step 7: Consider each of the FAIR principles and their facets

Method step 8: Compare and consolidate metrics per principle

Method step 9: Identify levels per metric

Method step 10: Propose pathway for improvement per metric

Method step 11: Identify dependencies, overlaps and gaps

Method step 12: Harmonise metrics across FAIR areas

Method step 13: Identify overall maturity levels and pathways

Method step 14: Draft core assessment criteria

Method step 15: Map existing approaches to draft assessment criteria

Method step 16: Apply draft assessment criteria to selected collections

Method step 17: Compare results and improve criteria

Method step 18: Finalise core assessment criteria

Method step 19: Describe overall pathways/guidelines

Method step 20: Publish results

Scope for today

Definition
Development
Testing
Delivery
Scope of the exercise

- Analysis of **one** of the FAIR principles
  - Findable
  - Accessible
  - Interoperable
  - Reusable
    - R1
      - R1.1 (Meta)data are released with a clear and accessible data usage licence
    - R1.2
    - R1.3

- Comparison and consolidation of the metric
- Identification of levels per metric
- Pathways for increasing maturity per metric
Deep dive into R1.1

"... R1.1 is about legal interoperability. What usage rights do you attach to your data? This should be described clearly. Ambiguity could severely limit the reuse of your data by organisations that struggle to comply with licensing restrictions. Clarity of licensing status will become more important with automated searches involving more licensing considerations. The conditions under which the data can be used should be clear to machines and humans.

GO-FAIR definition of R1.1
How do existing methodologies assess R1.1

**R1.1 (meta)data is released with a clear and accessible data usage licence**

<table>
<thead>
<tr>
<th></th>
<th>Which of the following best describes the license/usage rights attached to the data?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>No licence</td>
</tr>
<tr>
<td></td>
<td>Non standard text based licence</td>
</tr>
<tr>
<td></td>
<td>Non standard machine readable licence (e.g. clearly indicating under what conditions the data may be used)</td>
</tr>
<tr>
<td></td>
<td>Standard text based licence</td>
</tr>
<tr>
<td></td>
<td>Standard machine-readable licence (e.g. Creative Commons)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Does the user license have any user restrictions for accessing the data?</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Does the dataset have a user license?</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Does the dataset have a usage licence?</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Licensed - conditions for re-use are available and clearly expressed</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>License described in text</td>
</tr>
<tr>
<td></td>
<td>Link to a standard license (e.g. Creative Commons)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Please provide the IRI for your usage license regarding the content returned from RESOURCE ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>No</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>The existence of a license document, for BOTH (independently) the data and its associated metadata, and the ability to retrieve those documents</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>Somewhat</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Terms of usage (licenses, other conditions of reuse, data protection, ethical issues)</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>Somewhat</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
</tr>
</tbody>
</table>

from the [landscaping exercise](#)
Elements of a maturity model

<table>
<thead>
<tr>
<th>Questions</th>
<th>Options</th>
<th>Scoring Mechanism</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overlaps</td>
<td>Binary (Y/N)</td>
<td>Stars</td>
</tr>
<tr>
<td>Principles overused/underused</td>
<td>Multiple choice</td>
<td>Grade</td>
</tr>
<tr>
<td>Beyond the principles</td>
<td>Free text</td>
<td>Loading bar</td>
</tr>
</tbody>
</table>

**INDICATORS***
- Not standardising questions
- Defining indicators based on questions

**METRICS**
Definition of metrics to measure the indicators

**MATURITY**
Identification of the maturity levels

* an indicator can be seen as a component of a Principle (e.g. F1, R1)
Derived indicators

R1.1 (Meta)data are released with a clear and accessible data usage licence

**INDICATORS**

**User licence**
- Is there any?

**Nature of the licence**
- (non) standard text
- (non) standard machine readable

**Terms of usage**
- Restrictions
- Conditions of reuse
- Data protection
- Ethical issues

Is there any other indicator?
Measuring maturity
Alternative #1

Five options for R1.1 [metadata/data]

- Level 0: no licence
- Level 1: non standard licence in a human-readable format allowing access
- Level 2: standard licence in a human-readable format allowing access
- Level 3: standard open licence in a human-readable format allowing reuse
- Level 4: standard open licence in a machine-readable format allowing reuse
- Level 5: standard open licence in a machine-readable format with clear criteria allowing reuse

Each option is defining a maturity level

FAIRNess compliance for R1.1
Measuring maturity
Alternative #2

Two options per indicator for R1.1 [metadata/data]

- (1) Is there a **licence**? [YES/NO]
- (2) Is the licence a **standard** licence? [YES/NO]
- (3) Is the licence **machine** readable? [YES/NO]
- (4) Does the licence permit **access**? [YES/NO]
- (5) Does the licence permit **reuse**? [YES/NO]

The options aggregated correspond to a maturity level

FAIRNess compliance for R1.1
Measuring maturity
Your feedback

Do you agree with the proposed alternatives?
Which alternative is more suitable?
Do you think of other alternatives?
...

Method step 9
Pathways for increasing maturity

Method step 10

- From no licence to any licence
  - From a human-readable licence to a machine-readable licence
  - From licence allowing access to licence allowing reuse
Actions items & next steps
Discussion

- **Nature** of RDA recommendations & outputs
- How to keep you involved?
Action items

▷ Call for volunteers
▷ Development of the core assessment criteria on [GitHub](https://github.com)
  ▷ Analysis of all the FAIR principles
  ▶ FAIR – Findable [Link](https://example.com)
  ▶ FAIR – Accessible [Link](https://example.com)
  ▶ FAIR – Interoperable [Link](https://example.com)
  ▶ FAIR – Reusable [Link](https://example.com)
  ▷ Comparison and consolidation of the metrics per principle
  ▷ Identification of levels per metric
  ▷ Pathways of improvement per metric
▷ Online workshop #3
  ▷ at 09:00 CEST on the 18 June 2019
  ▷ at 17:00 CEST on the 18 June 2019
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 – **Mailing list**
  fair_maturity@rda-groups.org
Thank you!