Agreeing on core criteria for FAIR

the work of the RDA FAIR Data Maturity Model Working Group

11th September 2019
Agenda

1. Background to the WG
2. Process and where are at
3. Lessons learned so far
4. How to get involved
Background
Aim of the Working Group

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
- RDA recommendation
- FAIR data **checklist**

Join the **RDA** Working Group: [RDA WG web page](https://www.rd-alliance.org) | [GitHub](https://github.com)
Process
Overview of the methodology

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

Legend:
- Definition
- Development
- Testing
- Delivery
* The indicators and levels later presented are derived from the contributions on the [Gsheet](https://gheet) and [GitHub](https://github)
Timeline

**Workshop #1 [Febr 2019]**
- Introduction to the WG
- Existing approaches
- Landscaping exercise

**Workshop #3 [June 2019]**
- Presentation of results
- Discussion on indicators & levels

**Workshop #6 [Dec 2019]**
- TBC

**Workshop #2 [April 2019]**
- Approval of methodology & scope
- Hands-on exercise

**Workshop #4 [Sept 2019]**
- Proposals
- Proposed approach towards guidelines, checklist and testing

**Workshop #5 [Oct 2019]**
- TBC

... and more to come!
Where are we at?
Overview | Findable

F1 (Meta)data are assigned globally unique and persistent identifiers
- F1-01M Metadata is identified by a persistent identifier
- F1-02M Metadata is identified by a universally unique identifier
- F1-01D Data is identified by a persistent identifier
- F1-02D Data is identified by a universally unique identifier

F2 Data are described with rich metadata
- 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 Metadata clearly and explicitly include the identifier of the data they describe
- F3-01M Metadata includes the identifier for the data

F4 (Meta)data are registered or indexed in a searchable resource
- F4-01M Metadata or landing page is harvested by general search engine
- F4-02M Metadata is harvested by or submitted to domain/discipline-specific portal
- F4-03M Metadata is indexed in institutional repository
Option 1
FAIRness on a two level scale for the indicator
F1-01M – Metadata is identified by a persistent identifier
  - No persistent identifier [Not FAIR]
  - Persistent identifier [FAIR]

Option 2
FAIRness across indicator per levels
Multiple indicators with consolidated levels – whenever possible
  - Level 0
  - Level 1
  - Level 2

A1: (Meta)data are retrievable by their identifier using a standardised communication protocol
Two separate indicators can become levels for the principle, as demonstrated below
  - Level 1 – Metadata identifier resolves to a metadata record (A1-02M)
  - Level 2 – Metadata is accessed through a standardised protocol (A1-03M)
## Weighting

Weighting the indicators, developed as part of the WG, following the key words for use in RFC2119

- Mandatory/Essential: indicator MUST be satisfied for FAIRness
- Recommended/Important: indicator SHOULD be satisfied, if at all possible, to increase FAIRness
- Optional/Useful: indicator MAY be satisfied, but not necessarily so

<table>
<thead>
<tr>
<th>PRINCIPLE</th>
<th>INDICATOR ID</th>
<th>INDICATORS</th>
<th>PRIORITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1</td>
<td>F1-01M</td>
<td>Metadata is identified by a persistent identifier</td>
<td>Recommended</td>
</tr>
<tr>
<td>F1</td>
<td>F1-02M</td>
<td>Metadata is identified by a universally unique identifier</td>
<td>Recommended</td>
</tr>
<tr>
<td>F1</td>
<td>F1-01D</td>
<td>Data is identified by a persistent identifier</td>
<td>Mandatory</td>
</tr>
<tr>
<td>F1</td>
<td>F1-02D</td>
<td>Data is identified by a universally unique identifier</td>
<td>Mandatory</td>
</tr>
<tr>
<td>F2</td>
<td>F2-01M</td>
<td>Sufficient metadata is provided to allow discovery, following domain/discipline-specific metadata standard</td>
<td>Recommended</td>
</tr>
<tr>
<td>F2</td>
<td>F2-02M</td>
<td>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</td>
<td>Recommended</td>
</tr>
<tr>
<td>F3</td>
<td>F3-01M</td>
<td>Metadata includes the identifier for the data</td>
<td>Mandatory</td>
</tr>
<tr>
<td>F4</td>
<td>F4-01M</td>
<td>Metadata or landing page is harvested by general search engine</td>
<td>Recommended</td>
</tr>
<tr>
<td>F4</td>
<td>F4-02M</td>
<td>Metadata is harvested by or submitted to domain/discipline-specific portal</td>
<td>Recommended</td>
</tr>
<tr>
<td>F4</td>
<td>F4-03M</td>
<td>Metadata is indexed in institutional repository</td>
<td>Recommended</td>
</tr>
</tbody>
</table>
Development | Weighting Stats

Distribution of the weight of the indicators

**FAIR PRINCIPLES**

- **FINDABLE**
  - Mandatory: 0
  - Recommended: 3
  - Optional: 7

- **ACCESSIBLE**
  - Mandatory: 2
  - Recommended: 4
  - Optional: 7

- **INTEROPERABLE**
  - Mandatory: 0
  - Recommended: 7
  - Optional: 7

- **REUSABLE**
  - Mandatory: 4
  - Recommended: 4
  - Optional: 5
Current discussion items

1. DOI **without** explicit persistent identifiers for metadata or data
   - Indirect versus direct identification
   - What could be the priority levels of F1 indicators

2. **NO** common understanding for ‘Rich metadata’ F2 and ‘plurality of attributes’ R1
   - Rely on the output of the Metadata for FAIR data joint meeting
   - Minimum set common across fields of research | broader set required by the community (e.g. FAIRsharing)

3. ‘Knowledge representation’ I1 is too vague
   - Up to the evaluator to interpret
   - Agreed set of definitions per community
   - All indicators for I1 optional
   - More precise definitions of terms for I1 and I2 (e.g. Glossary)

4. FAIRness implies machine readability for metadata and data – **as opposed** to the evaluation
Lessons learned

- Varied levels of maturity, don’t want to scare away
- Some principles harder to test than others
- Apply both to data and metadata sometimes hard
- Underlying standards and vocabularies to test against are not in place
- Machine accessibility of FAIRness of data is hard
Development | Tool set and checklist

- Implement the indicators
- Automatic evaluation (e.g. FAIR Sharing registry, other registries, etc.)
- What to assess?

- Mandatory indicators
- Textual information
- Responsibility of the indicators
- Audiences (e.g. data stewards, data repositories, etc.)
Getting involved
Be involved

- Join the RDA FAIR Maturity Model Working Group

- Provide feedback to the proposals on GitHub, if at all possible, by the 30th September

- Share feedback about consolidation and weighting of indicators and maturity levels on GitHub

- Share feedback about the structure for tool set and data checklist on GitHub

- Join the virtual workshop tomorrow morning at 11am (EDT)

WORKSHOP #5
RDA 14th Plenary session in Helsinki (FI)
23rd October 2019
Breakout 2 – 14.30 - 16.00 EEST (7.30 EDT)
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/1mkjElFrTBPBHOQViODexNur0xNGhJqau0zkL4w8RRAw/edit

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

- We want to get the whole community along, not just the leaders
- How do you allow for varying maturity across communities
- FAIR is a scale
- Purpose of the tool: assessing or growing?
- Make it easy for the researcher
- Machine readable data down the track
- Machine assessable down the track
- Getting socially agreed norms/standards is more time consuming than technical agreement...