FAIR Data Maturity Model

Workshop #3
18th June 2019
Agenda

1. Welcome, objectives of the meeting
2. Roundtable
3. State of play
4. Development | First phase
   Presentation of the work conducted & discussion
5. Development | Second phase
   Presentation of an approach & discussion
6. Action items and next steps
Welcome, objectives of the meeting

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 | GitHub
Roundtable

Please type your name and affiliation in the chat window

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

1. Definition  DONE
2. Development  ONGOING
   i) First phase  ONGOING
   ii) Second phase  TO BE COMMENCED
3. Testing  ON HOLD
4. Delivery  ON HOLD
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
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 on indicators & levels

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

Q1 | Q2 | Q3 | Q4 | Q5 | Q6

RDA 13th Plenary - US
Today
RDA 14th Plenary - FI
... and more to come!
Development
First Phase
**STEP 1 - INDICATORS**

- **PI 1:** Data identified by a globally unique identifier
- **PI 2:** Persistence guarantee of the Identifier scheme used
- **PI 3:** What data outputs are given an identifier
- **PI 4:** UUPID assigned automatically
- **PI 5:** UUPID Assignment manual guarantee of continuity
- **PI 6:** Universally unique and eternally persistent identifier
- **PI 7:** Universally unique and eternally persistent identifier resolves
- **PI 8:** [https://github.com/FAIRMetrics/Metrics/blob/master/MaturityIndicators/Gen2Gen2_M_F1A.md](https://github.com/FAIRMetrics/Metrics/blob/master/MaturityIndicators/Gen2Gen2_M_F1A.md)
- **PI 9:** [https://github.com/FAIRMetrics/Metrics/blob/master/MaturityIndicators/Gen3Gen2_M_F18.md](https://github.com/FAIRMetrics/Metrics/blob/master/MaturityIndicators/Gen3Gen2_M_F18.md)
- **PI 10:** (Proposed indicator - expect to be treated - suggestion: Please remove this text)

**STEP 2 - LEVELS**

**STEP 3 - CONSOLIDATION**

---

* The indicators and levels later presented are derived from the contributions on the GSheet.*
Development | First phase

**PROPOSITION**
- Indicators
- Maturity levels

**CONSOLIDATION**
- Indicators
- Maturity levels

**DISCUSSION**
- Validation (YES/NO)
- Missing indicators

TODAY & COMING MONTHS
Development | Bottom-up approach

✅ Looking at all ‘atomic’ indicators and their ‘binary’ maturity levels [Slide 20 Workshop #2]

- Indicator #1
  - YES
  - NO

- Indicator #2
  - YES
  - NO

 الذهب Looking at deriving a set of levels across indicators for a principle [Slide 19 Workshop #2]

Combination of Indicator #1 and Indicator #2
- Level 0
- Level 1
- Level 2
More concrete contribution for F & A

Most contribution are about metadata

Complex versus simple principles [e.g. 11 indicators for F1 compared to 1 indicator for I3]
F1. (meta)data are assigned a globally unique and eternally persistent identifier

Globally unique
- NO globally unique identifier
- Globally unique identifier

Persistent
- NO persistent identifier
- Persistent identifier

Resolve
- Identifier does NOT resolve
- Identifier resolve
F2. data are described with rich metadata (defined by R1 below)

**Metadata**
- NO metadata
- Metadata

**Landing Page**
- NO landing page
- Landing page

Providing descriptive information according to a formal metadata standard
- NON-standard metadata
- Standard metadata
F3. metadata clearly and explicitly include the identifier of the data it describes

Presence of globally unique and eternally persistent identifier in the metadata

- **NO** globally unique and eternally persistent identifier
- Globally unique and eternally persistent identifier (e.g. DOI)
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
A1. (meta)data are retrievable by their identifier using a standardised communications protocol

Access conditions
- NO access conditions
- Access conditions

Manual access
- Data retrievable via the researcher
- Data retrievable via a repository

Automated access
- Data retrievable via human interaction
- Data retrievable using a standard client software
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
A1.2. the protocol allows for an authentication and authorization procedure, where necessary

Protocol authentication

- **NO** protocol authentication
- Protocol authentication
A2. metadata are accessible, even when the data are no longer available

Metadata persistence policy / guarantee

- NO persistence policy / guarantee
- Persistence policy / guarantee
(meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation

**Format**
- Proprietary format
- Community standard format

**Machine-readable [Knowledge representation]**
- **NOT** machine readable
- Machine readable

**Self-describing**
- **NOT** self-describing
- Self-describing
I2. (meta)data use vocabularies that follow FAIR principles

Standard vocabularies
- **NO** standard vocabularies
- Standard vocabularies

FAIR compliant vocabularies
- **NO** FAIR compliant vocabularies
- FAIR compliant vocabularies
I3. (meta)data include qualified references to other (meta)data

References to other metadata
- **NO** references to other metadata
- References to other metadata (e.g. ORCID for author)

Sufficient level of qualification for other metadata
- General qualification (e.g. contributor)
- Specific qualification (e.g. author, reviewer, publisher, funder)
R1. meta(data) are richly described with a plurality of accurate and relevant attributes

Use of guidelines for relevant attributes accompanying metadata

- NO guidelines
- Guidelines
R1.1. (meta)data are released with a clear and accessible data usage licence

User licence
- **NO** user licence
- Presence of a user licence

Nature of the licence
- Local licence
- Standard licence (e.g. CC)

Machine-readable licence
- **NON** machine-readable licence
- Machine-readable licence

Consent for reuse
- **NO** information about the consent
- Explicit documented consent
R1.2. (meta)data are associated with detailed provenance

Authorship included
- **NO** author
- Reference to an author

Technical parameters [e.g. software’s and instruments]
- **NO** information about the technical parameters
- Information about the technical parameters
R1.3. (meta)data meet domain-relevant community standards

Presence of a template for metadata following a community standard

- NO template
- Presence of a template
X. Choices beyond FAIR

- Versioning of the identifier [F1]
- Versioning of the dataset [F1]
- Keywords for rich metadata [F2]
- Quality (i.e. referential and functional integrity) [F2]
- Access control [A1.2]
- Minimal metadata [R1.3]

Should they have **indicators**? If so, under which **principle**?
Core assessment criteria to evaluate and compare FAIRness

- One to three indicators per FAIR principle [e.g. F1, A1.1]
- Bottom-up perspective

**FAIRness on a two | three level scale for the indicator**

1. Not FAIR
2. FAIR

**Weighing indicators?**

1. Required [Must be FAIR]
2. Recommended [Should be FAIR, if possible]
3. Optional [May be FAIR but not essentially so]
Development | Second phase

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
Next steps
Next steps

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

Online Workshop #4
  ▶ at 09:00 CEST on the 12 September 2019
  ▶ at 17:00 CEST on the 12 September 2019
Resources

- RDA 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/1gvMfbw46oV1idztsr586aG6teSn2cPWe_RJZG0U4Hg/edit#gid=0](https://docs.google.com/spreadsheets/d/1gvMfbw46oV1idztsr586aG6teSn2cPWe_RJZG0U4Hg/edit#gid=0)

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