



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

FAIR

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**



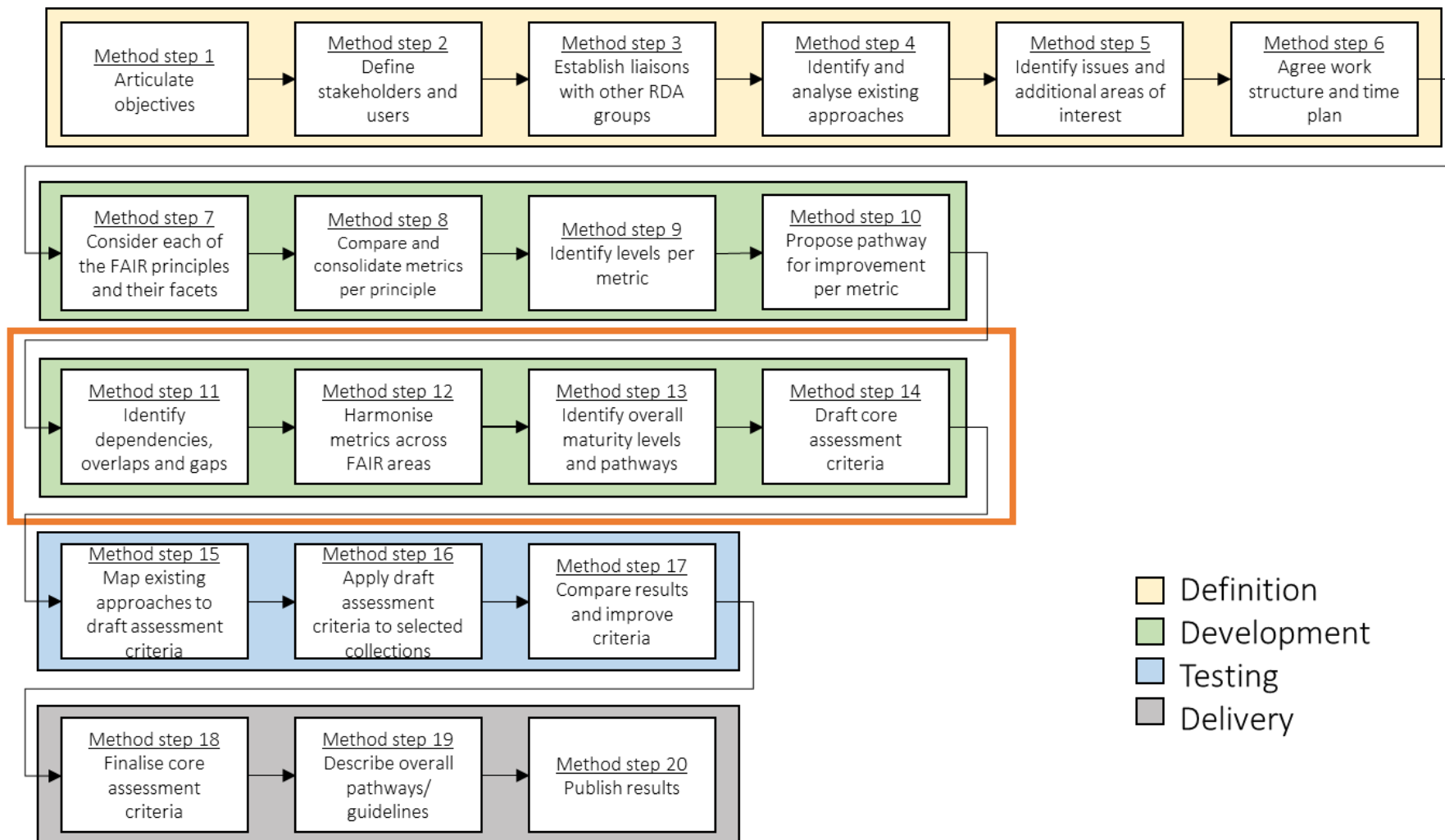
NOT re-design
the FAIR Principles

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

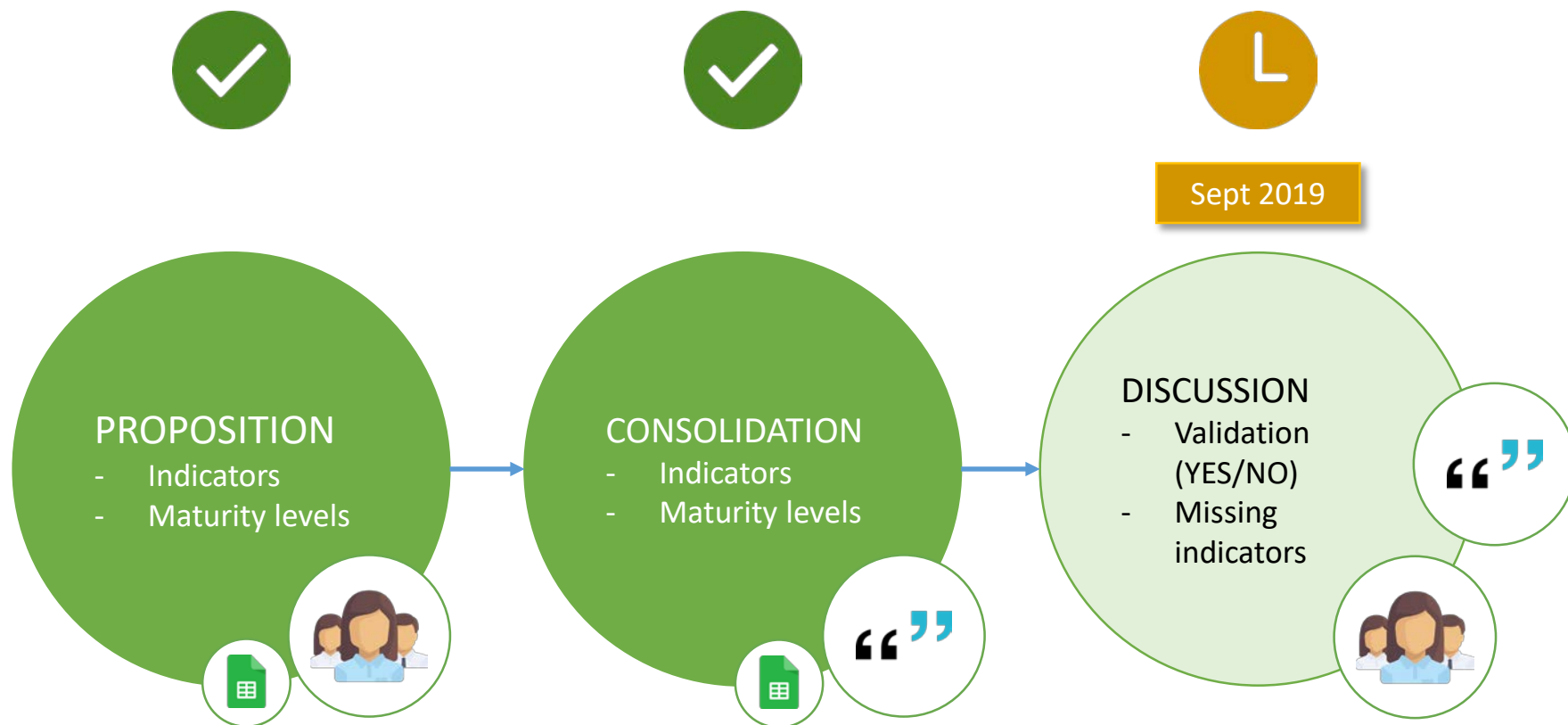


Process

Overview of the methodology

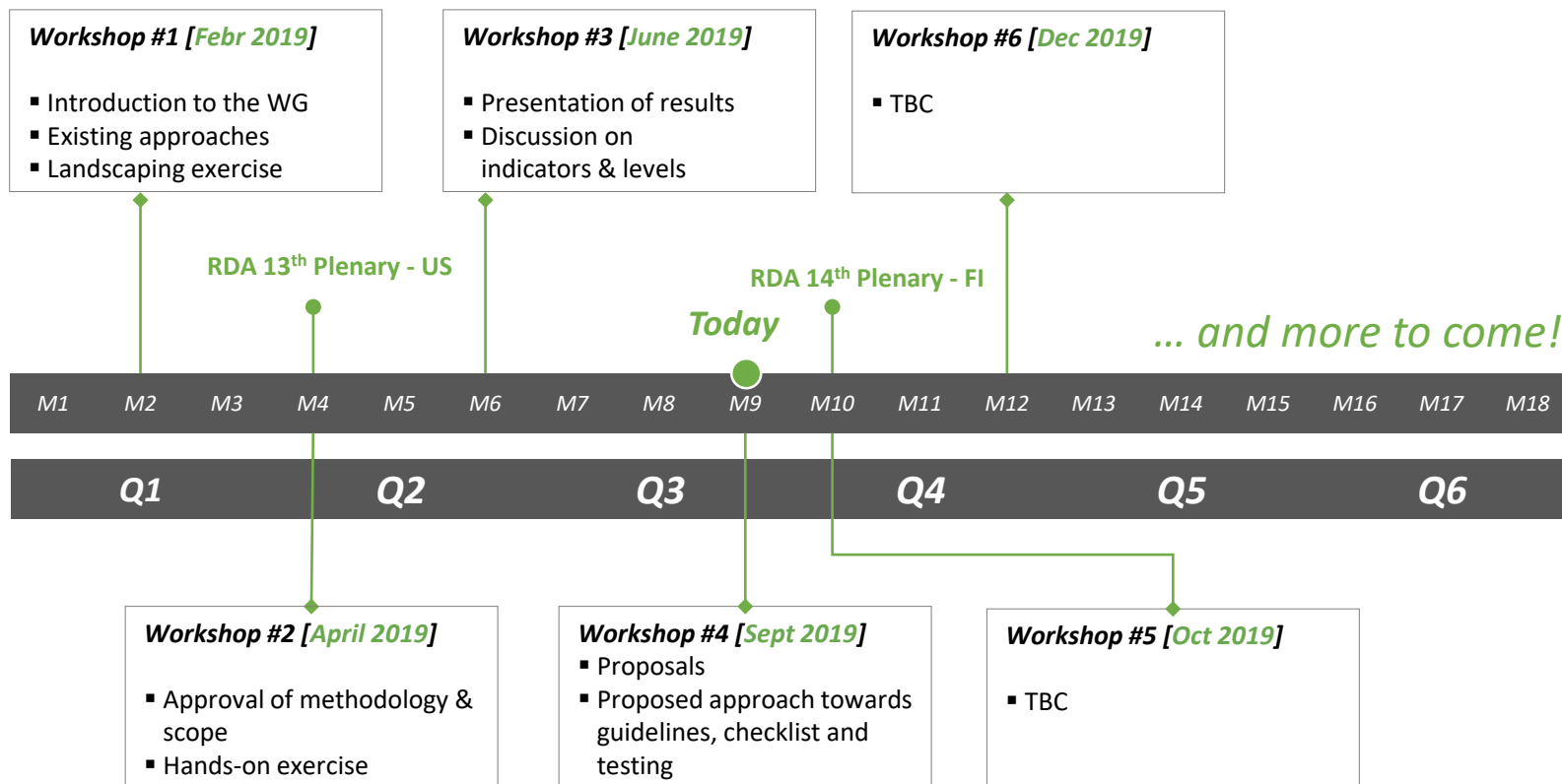


Development | Where are we at?



* The indicators and levels later presented are derived from the contributions on the [Gsheet](#) and [GitHub](#)

Timeline





Where are we
at?

Overview | Findable

- Under discussion
- Provisionally agreed

F

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

Development | Levels



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 accross indicator per levels

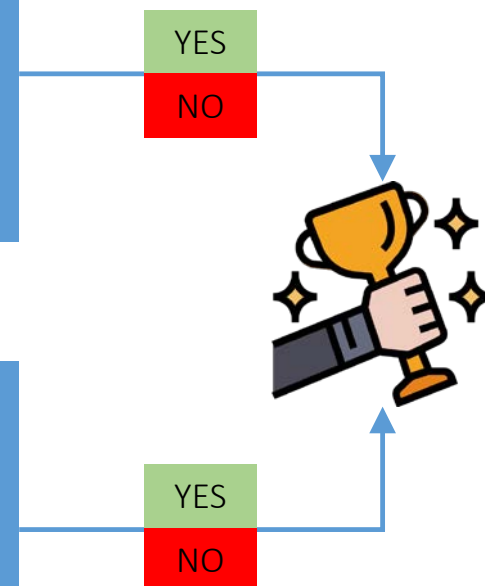
Multiple indicators with consolidated levels – *whenever possible*

- 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)



Development | Weighting

Weighting the indicators, developed as part of the WG, following the [key words for use](#) in RFC2119

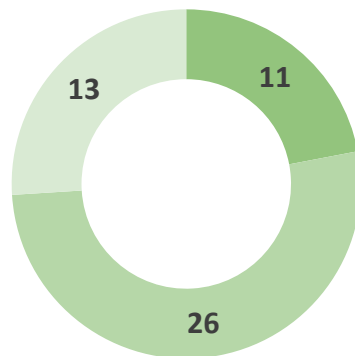
- Mandatory/Essential: indicator **MUST** be satisfied for FAIRness
- Recommended/Important: indicator **SHOULD** be satisfied, if at all possible, to increase FAIRness
- Optional/Useful: indicator **MAY** be satisfied, but not necessarily so

	PRINCIPLE	INDICATOR_ID	INDICATORS	PRIORITY
F	F1	F1-01M	Metadata is identified by a persistent identifier	Recommended
	F1	F1-02M	Metadata is identified by a universally unique identifier	Recommended
	F1	F1-01D	Data is identified by a persistent identifier	Mandatory
	F1	F1-02D	Data is identified by a universally unique identifier	Mandatory
	F2	F2-01M	Sufficient metadata is provided to allow discovery, following domain/discipline-specific metadata standard	Recommended
	F2	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	Recommended
	F3	F3-01M	Metadata includes the identifier for the data	Mandatory
	F4	F4-01M	Metadata or landing page is harvested by general search engine	Recommended
	F4	F4-02M	Metadata is harvested by or submitted to domain/discipline-specific portal	Recommended
	F4	F4-03M	Metadata is indexed in institutional repository	Recommended

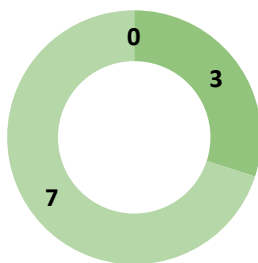
Development | Weighting Stats

Distribution of the weight of the indicators

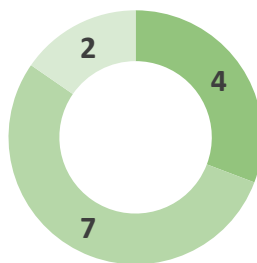
- Mandatory
- Recommended
- Optional



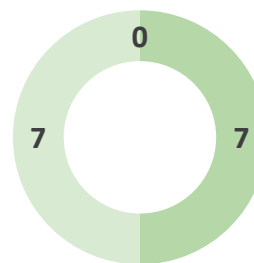
FAIR PRINCIPLES



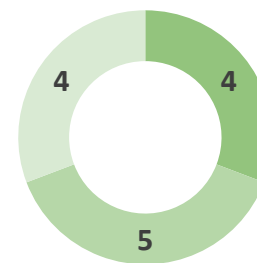
FINDABLE



ACCESSIBLE



INTEROPERABLE



REUSABLE

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)

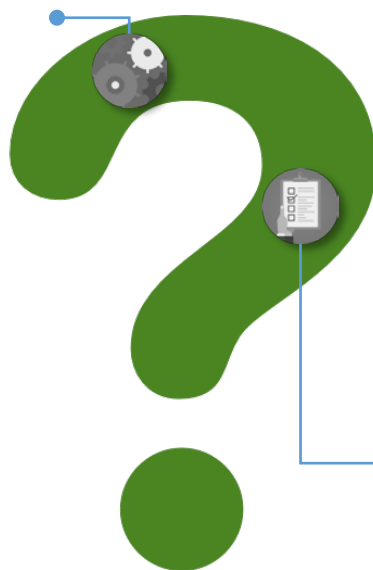
4 FAIRness implies machine readability for metadata and data – **as opposed** to the evaluation

Lessons learned

- Varied levels of maturity, don't want to scareaway
- 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 assessability 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**

<https://www.rd-alliance.org/group/fair-data-maturity-model-wg/case-statement/fair-data-maturity-model-wg-case-statement>

➤ RDA FAIR data maturity model WG – **GitHub**

<https://github.com/RDA-FAIR/FAIR-data-maturity-model-WG>

➤ 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 – **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...