



FAIR Data Maturity Model WG

presented by Edit Herczog Co-chair

Réunion RDA France 2019

le vendredi 13 septembre 2019.

Who we are

- › WG started the WG in January 2019
- › Co chairs:
 - › Keith Russel from Australia
 - › Edit Herczog from Europe
 - › Co chair from USA (under discussion)
- › TAB member:
 - › Jane Wyngaard from South Africa
- › Editorial team: EC special support
 - › Makx Dekkers and the PWC team
- › 129 members: 61 Female, 68 male
- › We had our 4 workshops and the P13 session
- › We have a session in P14 in Helsinki

We aim to keep the WG 18 months timeline: It would allow to use our recommendation in 2021

Minimum CORE criteria

➤ **WHAT**

➤ **NOT HOW**

We do not reintend the wheel. We build on what we have



WG

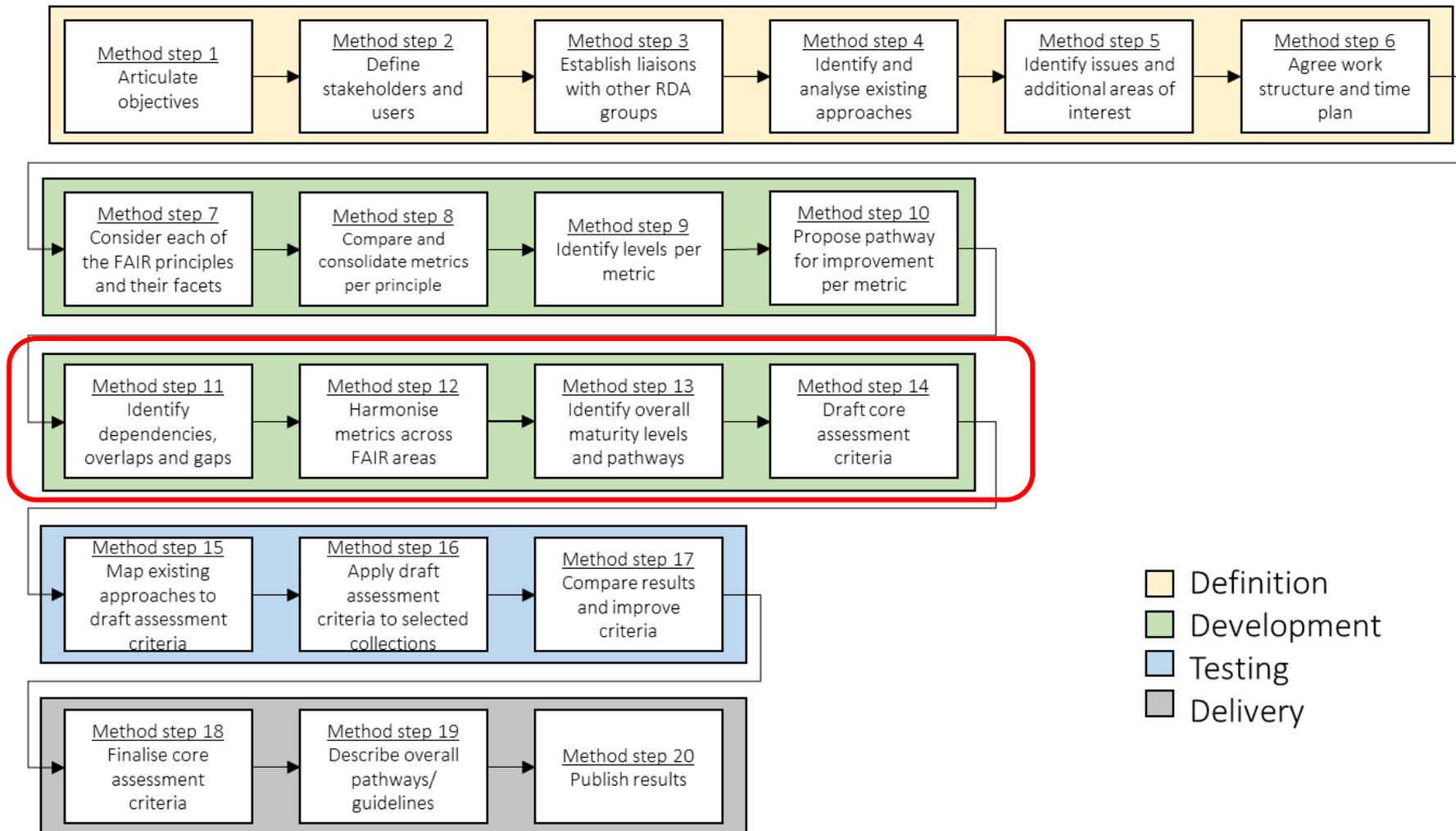
methodology,
timeline &
scope

Proposed development methodology

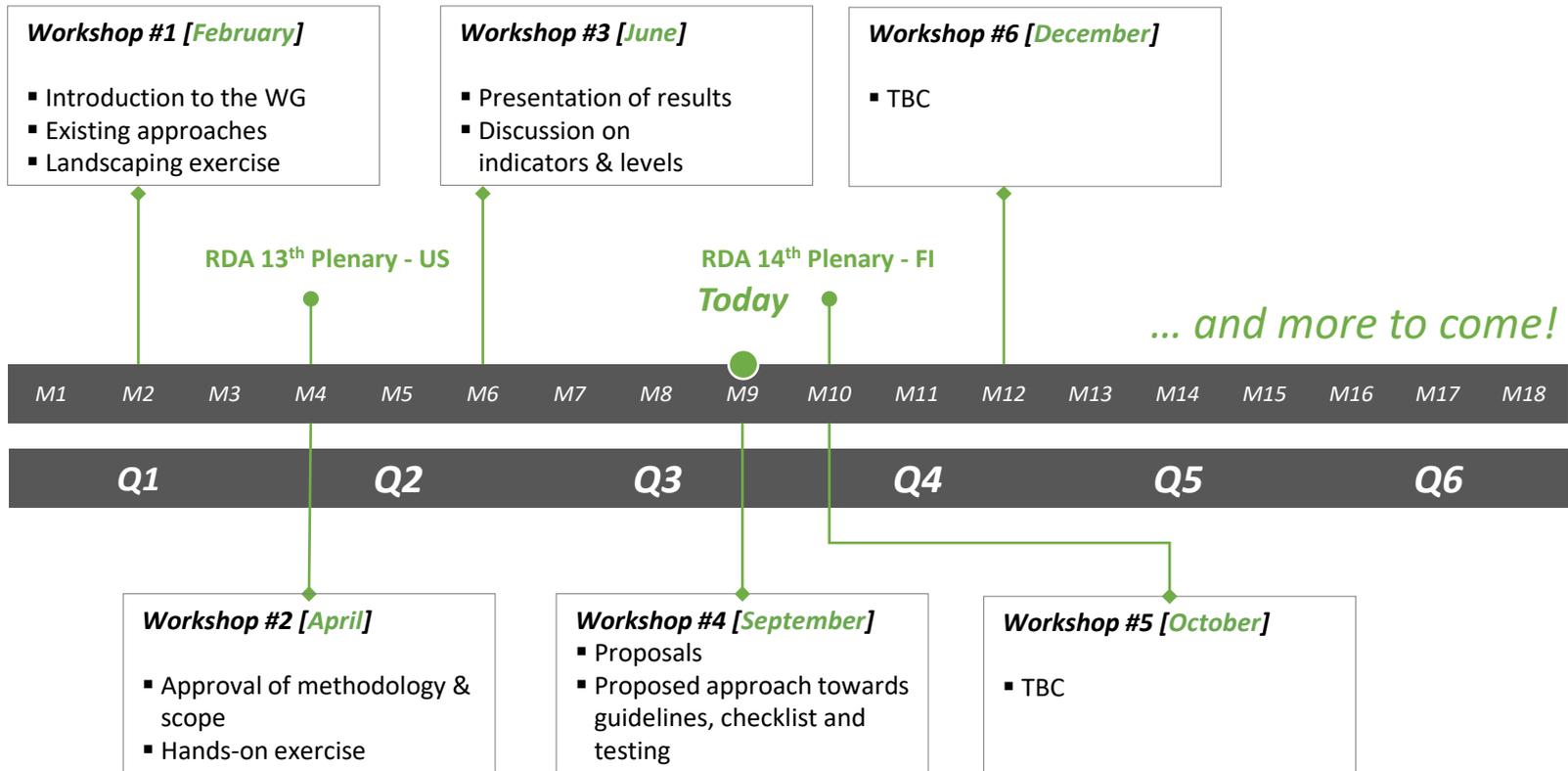
Bottom-up approach comprising 4 phases

- Definition
- Development
 - Assessment of the four FAIR principles in four 'strands'
 - Fifth 'strand': beyond the FAIR principles
- Testing
- Delivery

Overview of the methodology



Timeline



State of play

1. Definition	DONE
2. Development	ONGOING
i) First phase	CLOSING*
ii) Second phase	ONGOING
3. Testing	TO BE COMMENCED
4. Delivery	ON HOLD

* Any comments are still welcomed with regards to the output produced during the first phase | [GitHub](#)



Definitions

Done

› So far, **11** approaches are on the radar

Approaches considered

- › ANDS-NECTAR-RDS-FAIR data assessment tool
- › DANS-Fairdat
- › DANS-FAIR enough?
- › The CSIRO 5-star Data Rating Tool
- › FAIR Metrics questionnaire
- › Checklist for Evaluation of Dataset Fitness for Use
- › RDA-SHARC Evaluation
- › FAIR evaluator

Approach partially considered*

- › Data Stewardship Wizard

Approaches not considered*

- › Big Data Readiness
- › Support Your data: A Research Data Management Guide for Researchers

*Methodologies analysed but partially/not included in the results because of questions that could not be classified

Results of preliminary analysis - 3

› Early observations

123 questions

5 types of option

4 scoring approaches

- › On average, six questions per facet
 - › Overlaps and different terminologies used
 - › Some facets are underused [e.g. A1, A1.1, A1.2, A2]
 - › Some facets are overused [e.g. F1, F2]
- › Different options
 - › YES/NO
 - › TRUE/FALSE
 - › URL
 - › Multiple choice
 - › Free text
- › Different scoring mechanisms
 - › Stars
 - › Grade
 - › Loading bar
 - › None

Proposed scope

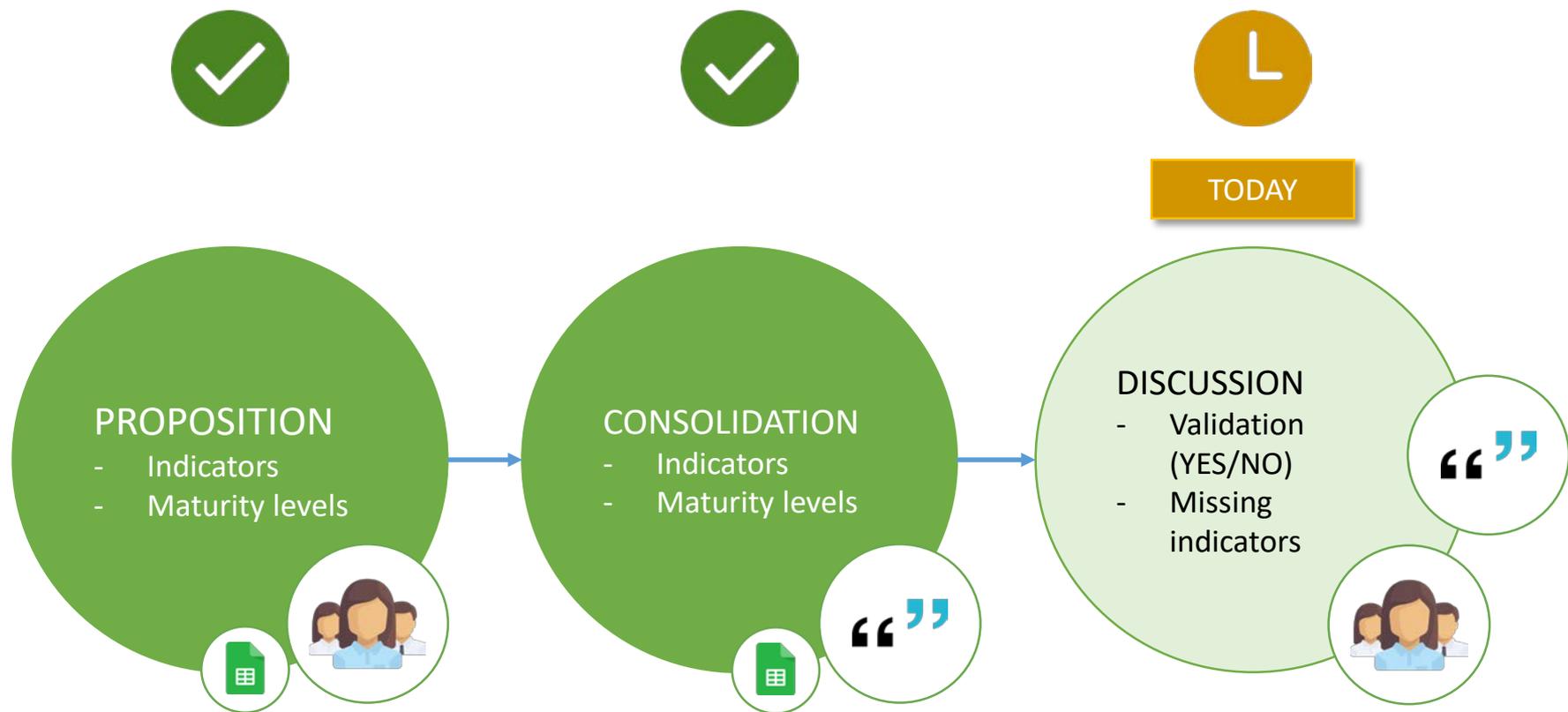
	Proposed resolutions
ENTITY	<u>Dataset</u> and <u>data-related aspects</u> (e.g. algorithms, tools and workflows)
NATURE	<u>Generic</u> assessment (i.e. cross-disciplines)
FORMAT	<u>Manual</u> assessment
TIME	<u>Periodically</u> throughout the lifecycle of the data
RESPONDENT	People with <u>data literacy</u> (e.g. researchers, data librarians, data stewards)
AUDIENCE	Researchers, data stewards, data professionals, data service owners, organisations involved in research data and policy makers



Development

First Phase

Development | First phase



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

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

- L 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

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 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)

Development | Weighting

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

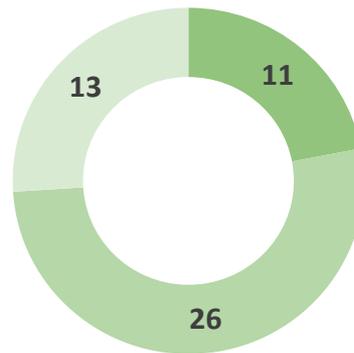
- Mandatory: indicator **MUST** be satisfied for FAIRness
- Recommended: indicator **SHOULD** be satisfied, if at all possible, to increase FAIRness
- Optional: indicator **MAY** be satisfied, but not necessarily so

	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

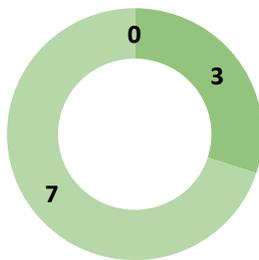
Development | Weighting Stats

Distribution of the weight of the indicators

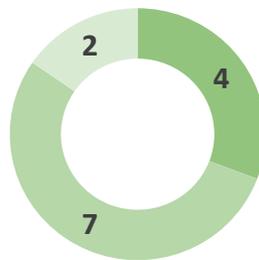
- Mandatory
- Recommended
- Optional



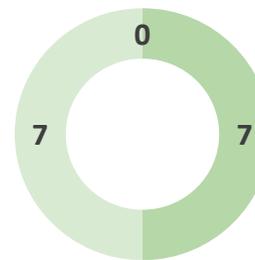
FAIR PRINCIPLES



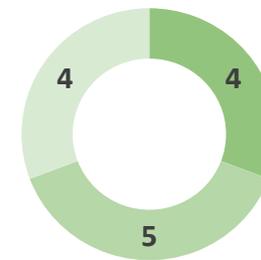
FINDABLE



ACCESSIBLE



INTEROPERABLE



REUSABLE

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



Development

Next steps

Development | Scoring

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

	Mandatory	Recommended	Optional
Level 0	○		
Level 1	●		
Level 2	●	◐	
Level 3	●	●	
Level 4	●	●	◐
Level 5	●	●	●

Level 0 – The resource did not comply with all the mandatory indicators

Level 1 – The resource did comply with all the mandatory indicators, and less than half of the recommended indicators

Level 2 – The resource did comply with all the mandatory indicators and at least half of the recommended indicators

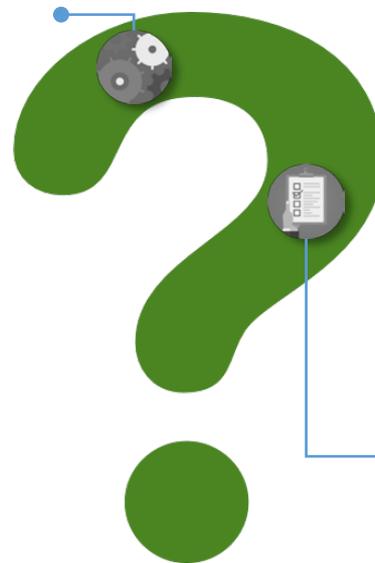
Level 3 – The resource did comply with all the mandatory and recommended indicators, and less than half of the optional indicators

Level 4 – The resource did comply with all the mandatory and recommended indicators and at least half of the optional indicators

Level 5 – The resource did comply with all the mandatory, recommended and optional indicators

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.)



Actions items & next steps

Testing the set of indicators

We identified two levels of testing;

1st 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

2nd Level

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

Be one of us!

- › Fourth Workshop 12th of Sept, 9 -10.30 CET
- › RDA 14th Plenary Helsinki, 23rd of October

- › **Sign to RDA WG today**
- › <https://www.rd-alliance.org/groups/fair-data-maturity-model-wg>



Thank you!