



RESEARCH DATA ALLIANCE

# FAIR Data Maturity Model

Workshop #4

12th September 2019

# Agenda

1. Welcome, objectives of the meeting
2. Roundtable
3. State of play
4. Development | First phase  
Presentation of the work conducted **for approval**
5. Development | Second phase  
Presentation of an approach & **discussion**
6. Testing  
Presentation of an approach & **discussion**
7. 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

# 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](#)

# 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





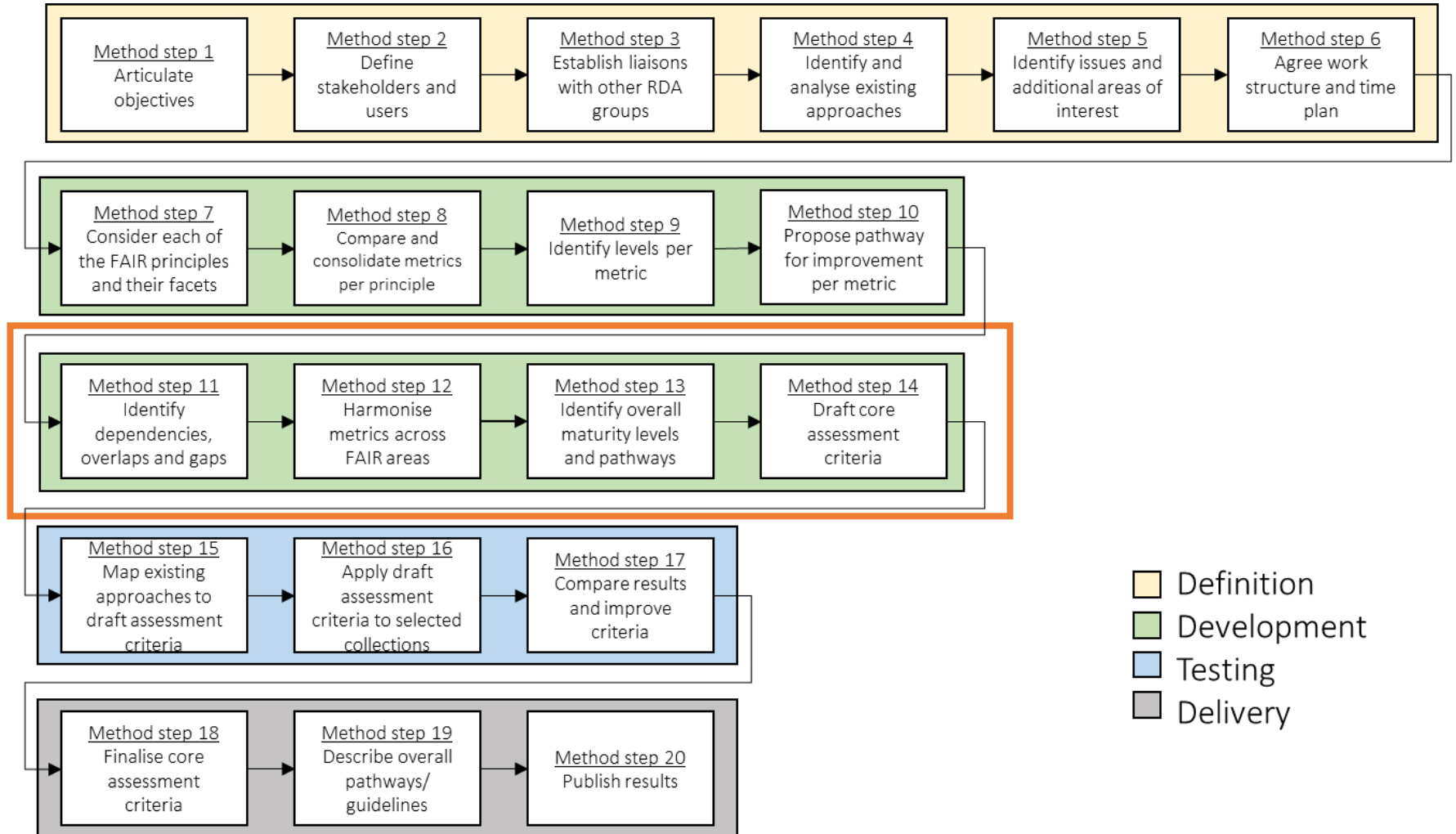
# State of play

# State of play

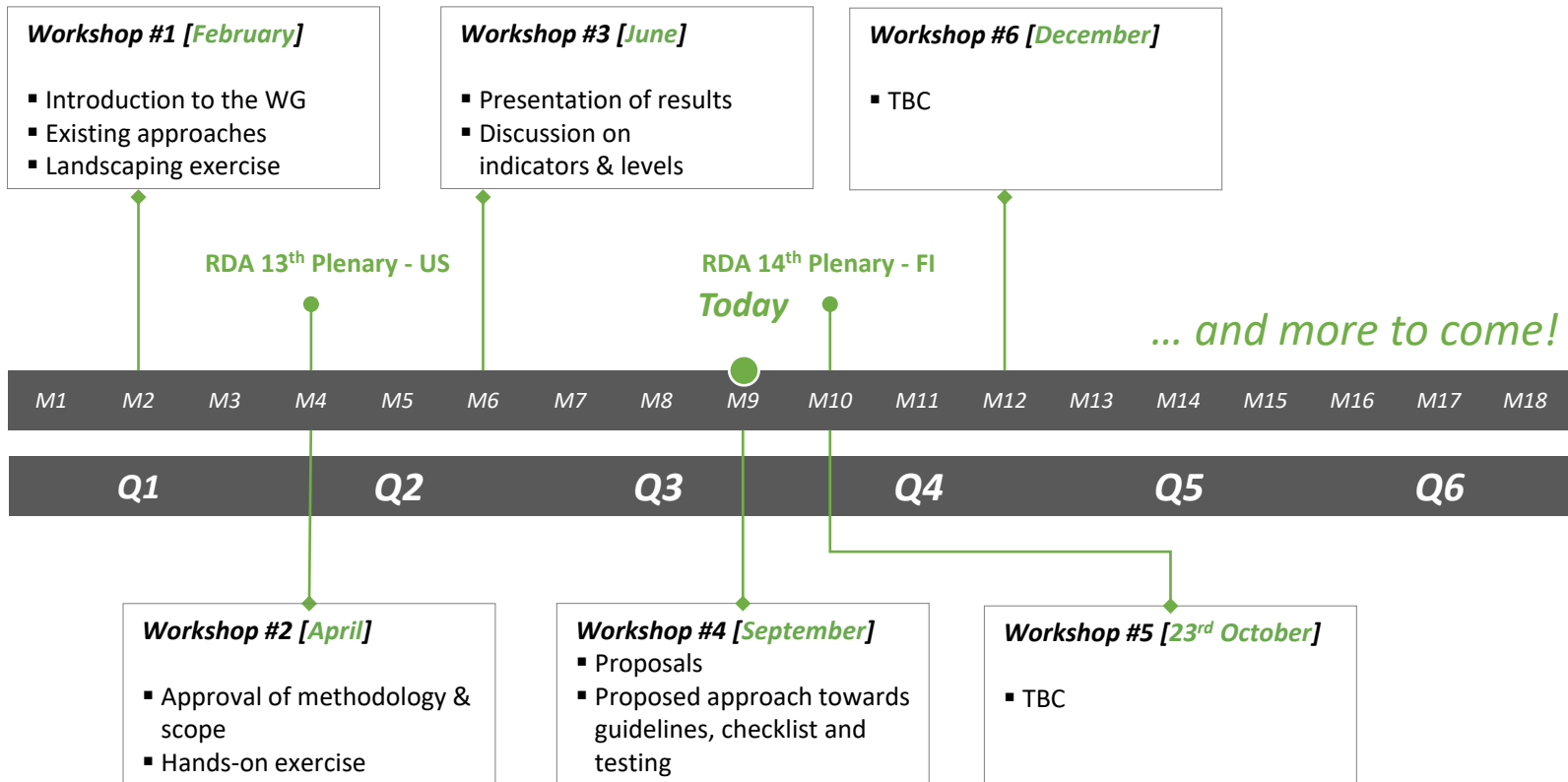
1. Definition	DONE
2. Development	ONGOING
i) First phase	CLOSING*
ii) Second phase	ONGOING
3. Testing	TO BE COMMENCED
4. Delivery	ON HOLD

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

# Overview of the methodology



# Timeline



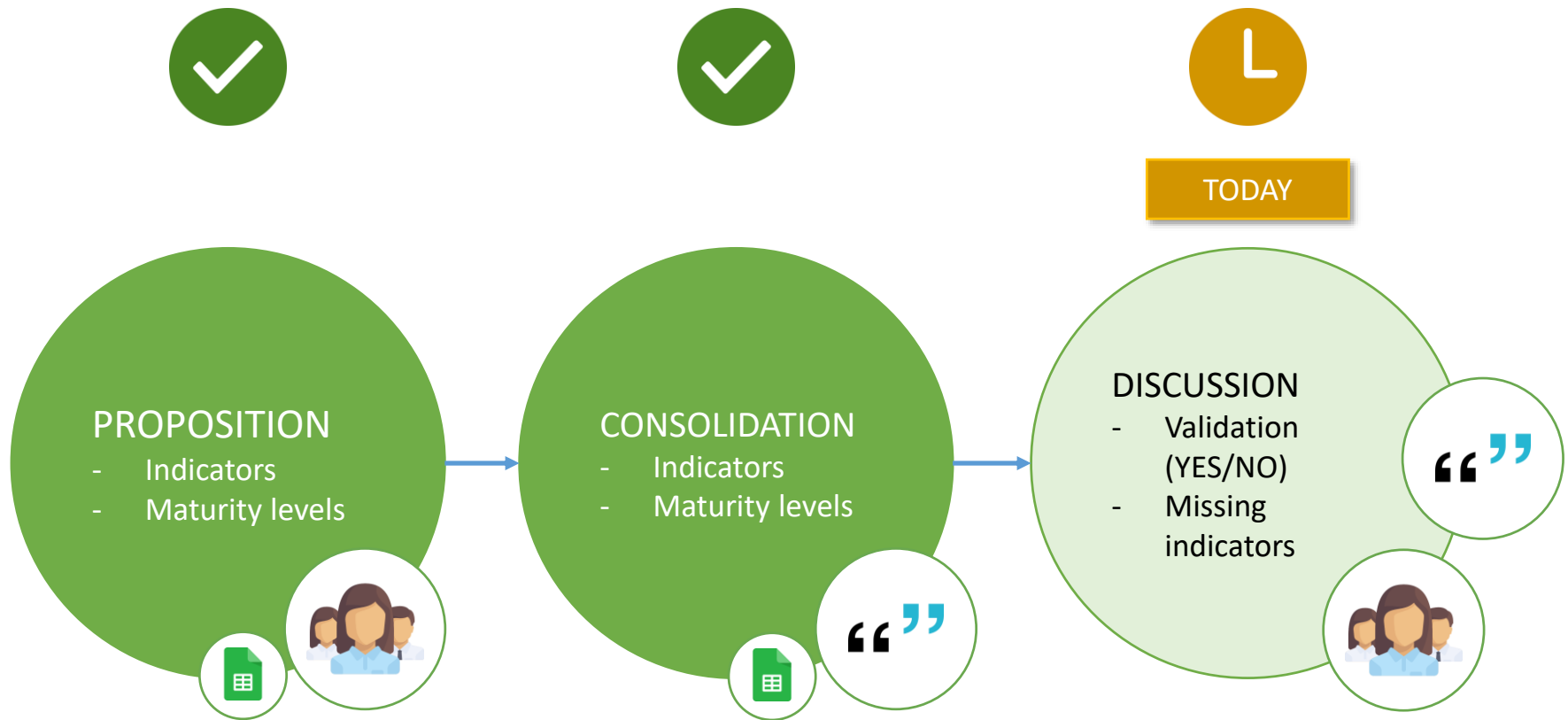




# Development

## *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

# Overview | Indicators & levels

- Under discussion
- Provisionally agreed

F

- F1 (Meta)data are assigned globally unique and persistent identifiers
- F2 Data are described with rich metadata
- F3 Metadata clearly and explicitly include the identifier of the data they describe
- F4 (Meta)data are registered or indexed in a searchable resource

A

- A1 (Meta)data are retrievable by their identifier using a standardised communication protocol
- A1.1 The protocol is open, free and universally implementable
- A1.2 The protocol allows for an authentication and authorisation where necessary
- A2 Metadata are accessible, even when the data are no longer available

I

- I1 (Meta)data use a formal, accessible, shared and broadly applicable language for knowledge representation
- I2 (Meta)data use vocabularies that follow the FAIR principles
- I3 (Meta)data include qualified references to other (meta)data

R

- R1 (Meta)data are richly described with a plurality of accurate and relevant attributes
- R1.1 (Meta)data are released with a clear and accessible data usage license
- R1.2 (Meta)data are associated with detailed provenance
- R1.3 (Meta)data meet domain-relevant community standards

## 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 is offered/published/exposed in such a way that it can be harvested and indexed

## A

### A1 (Meta)data are retrievable by their identifier using a standardised communication protocol

- A1-01M Metadata includes information about access conditions
- A1-01D Data can be accessed manually (i.e. with human intervention)
- A1-02D Data can be accessed automatically (i.e. by a computer program)
- A1-02M Metadata identifier resolves to a metadata record
- A1-03D Data identifier resolves to a digital object
- A1-03M Metadata is accessed through standardised protocol
- A1-04D Data is accessible through standardised protocol

#### A1.1 The protocol is open, free and universally implementable

- A1.1-01M Metadata is accessible through a free access protocol
- A1.1-01D Data is accessible through a free access protocol
- A1.1-02M Metadata is accessible through an open-source access protocol
- A1.1-02D Data is accessible through an open-source access protocol
- A1.1-03D Actions to be taken by a reuser to get access to the data are well documented

## A

### A1.2 The protocol allows for an authentication and authorisation where necessary

- A1.2-01D Data is accessible through an access protocol that supports authentication
- A1.2-02D Data is accessible through an access protocol that supports authorisation
- A1.2-01M Metadata includes information relevant for access control

### A2 Metadata are accessible, even when the data are no longer available

- A2-01M Metadata is guaranteed to remain available after data is no longer available

## I1 (Meta)data use a formal, accessible, shared and broadly applicable language for knowledge representation

- I1-01M Metadata uses knowledge representation expressed in standardised format
- I1-01D Data uses knowledge representation expressed in standardised format
- I1-02M Metadata uses machine-understandable knowledge representation
- I1-02D Data uses machine-understandable knowledge representation
- I1-03M Metadata uses self-describing knowledge representation
- I1-03D Data uses self-describing knowledge representation

## I2 (Meta)data use vocabularies that follow the FAIR principles

- I2-01M Metadata uses standard vocabularies
- I2-01D Data uses standard vocabularies
- I2-02M Metadata uses FAIR-compliant vocabularies
- I2-02D Data uses FAIR-compliant vocabularies



# Overview | Interoperable

- Under discussion
- Provisionally agreed

## I3 (Meta)data include qualified references to other (meta)data

- I3-01M Metadata includes references to other metadata
- I3-01D Data includes references to other data
- I3-02M Metadata includes references to other data
- I3-02D Data includes sufficiently qualified references to other data

## R1 (Meta)data are richly described with a plurality of accurate and relevant attributes

- R1-01M Sufficient metadata is provided to allow reuse, following domain/discipline-specific metadata standard
- R1-02M Metadata is provided for the reuse-related elements defined by the RDA Metadata IG, as much as possible and relevant, if no domain/discipline-specific metadata standard is available

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

- R1.1-01M Metadata includes information about the licence under which the data can be reused
- R1.1-02M Metadata refers to a standard reuse licence
- R1.1-03M Metadata includes licence information in the appropriate element of the metadata standard used
- R1.1-04M Metadata refers to a machine-understandable reuse licence
- R1.1-06M Metadata includes information about consent for reuse (e.g. for personal data)

### R1.2 (Meta)data are associated with detailed provenance

- R1.2-01M Metadata includes provenance information according to community-specific guidelines
- R1.2-02M Metadata includes provenance information according to a cross-domain language

### R1.3 (Meta)data meet domain-relevant community standards

- R1.3-01M Metadata complies with a community standard
- R1.3-01D Data complies with a community standard
- R1.3-02M Metadata is expressed in compliance with a machine-understandable community standard
- R1.3-02D Data is expressed in compliance with a machine-understandable community standard



# Development

## *Second Phase*

# 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]

YES  
NO



## Option 2

FAIRness across indicator per levels

Multiple indicators with consolidated levels – *whenever possible*

- Level 2
- Level 1
- Level 0

YES  
NO



For example:

A1-01D+A1-02D:

level 2 – Data can be accessed automatically

level 1 – Data can be accessed manually

level 0 – Data cannot be accessed automatically or manually

# 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	F	F1-01M	Metadata is identified by a persistent identifier	Recommended
	F1	F	F1-01D	Data is identified by a persistent identifier	Mandatory
	F1	F	F1-02M	Metadata is identified by a universally unique identifier	Recommended
	F1	F	F1-02D	Data is identified by a universally unique identifier	Mandatory
	F2	F	F2-01M	Sufficient metadata is provided to allow discovery, following domain/discipline-specific metadata standard	Recommended
	F2	F	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	F	F3-01M	Metadata includes the identifier for the data	Mandatory
	F4	F	F4-01M	Metadata is offered/published/exposed in such a way that it can be harvested and indexed	Recommended

# Development | Weighting Accessibility

A	A1	A	A1-01M	Metadata includes information about access conditions	Optional
	A1	A	A1-01D	Data can be accessed manually (i.e. with human intervention)	Recommended
	A1	A	A1-02D	Data can be accessed automatically (i.e. by a computer program)	Recommended
	A1	A	A1-02M	Metadata identifier resolves to a metadata record	Optional
	A1	A	A1-03D	Data identifier resolves to a digital object	Mandatory
	A1	A	A1-03M	Metadata is accessed through standardised protocol	Recommended
	A1	A	A1-04D	Data is accessible through standardised protocol	Recommended
	A1.1	A	A1.1-01M	Metadata is accessible through a free access protocol	Mandatory
	A1.1	A	A1.1-01D	Data is accessible through a free access protocol	Mandatory
	A1.1	A	A1.1-02M	Metadata is accessible through an open-source access protocol	Recommended
	A1.1	A	A1.1-02D	Data is accessible through an open-source access protocol	Recommended
	A1.1	A	A1.1-03D	Actions to be taken by a reuser to get access to the data are well documented	Recommended
	A1.2	A	A1.2-01M	Metadata includes information relevant for access control	Mandatory
	A1.2	A	A1.2-01D	Data is accessible through an access protocol that supports authentication	Recommended
	A1.2	A	A1.2-02D	Data is accessible through an access protocol that supports authorisation	Recommended
	A2	A	A2-01M	Metadata is guaranteed to remain available after data is no longer available	Mandatory

# Development | Weighting Interoperability

I	I1	I	I1-01M	Metadata uses knowledge representation expressed in standardised format	Recommended
	I1	I	I1-01D	Data uses knowledge representation expressed in standardised format	Recommended
	I1	I	I1-02M	Metadata uses machine-understandable knowledge representation	Optional
	I1	I	I1-02D	Data uses machine-understandable knowledge representation	Optional
	I1	I	I1-03M	Metadata uses self-describing knowledge representation	Optional
	I1	I	I1-03D	Data uses self-describing knowledge representation	Optional
	I2	I	I2-01M	Metadata uses standard vocabularies	Recommended
	I2	I	I2-01D	Data uses standard vocabularies	Recommended
	I2	I	I2-02M	Metadata uses FAIR-compliant vocabularies	Optional
	I2	I	I2-02D	Data uses FAIR-compliant vocabularies	Optional
	I3	I	I3-01M	Metadata includes references to other metadata	Recommended
	I3	I	I3-01D	Data includes references to other data	Recommended
	I3	I	I3-02M	Metadata includes references to other data	Recommended
	I3	I	I3-02D	Data includes sufficiently qualified references to other data	Optional
	I3	I	I3-03M	Metadata includes sufficiently qualified references to other metadata	Recommended
	I3	I	I3-04M	Metadata include sufficiently qualified references to other data	Optional

# Development | Weighting Reusability

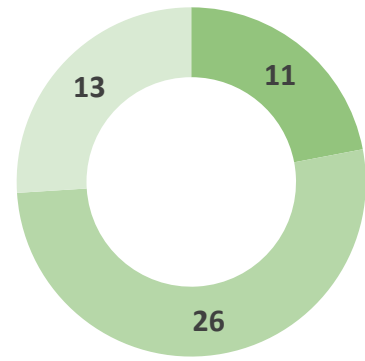
R	R1	R	R1-01M	Sufficient metadata is provided to allow reuse, following domain/discipline-specific metadata standard	Recommended
	R1	R	R1-02M	Metadata is provided for the reuse-related elements defined by the RDA Metadata IG, as much as possible and relevant, if no domain/discipline-specific metadata standard is available	Recommended
	R1.1	R	R1.1-01M	Metadata includes information about the licence under which the data can be reused	Mandatory
	R1.1	R	R1.1-02M	Metadata refers to a standard reuse licence	Recommended
	R1.1	R	R1.1-03M	Metadata includes licence information in the appropriate element of the metadata standard used	Mandatory
	R1.1	R	R1.1-04M	Metadata refers to a machine-understandable reuse licence	Optional
	R1.1	R	R1.1-06M	Metadata includes information about consent for reuse (e.g. for personal data)	Recommended
	R1.2	R	R1.2-01M	Metadata includes provenance information according to community-specific guidelines	Recommended
	R1.2	R	R1.2-02M	Metadata includes provenance information according to a cross-domain language	Optional
	R1.3	R	R1.3-01M	Metadata complies with a community standard	Mandatory
	R1.3	R	R1.3-01D	Data complies with a community standard	Mandatory
	R1.3	R	R1.3-02M	Metadata is expressed in compliance with a machine-understandable community standard	Optional
	R1.3	R	R1.3-02D	Data is expressed in compliance with a machine-understandable community standard	Optional



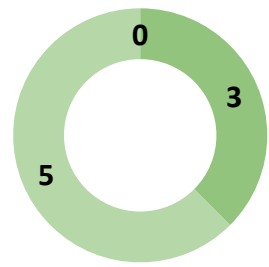
# Development | Weighting Stats

Distribution of the weight of the indicators

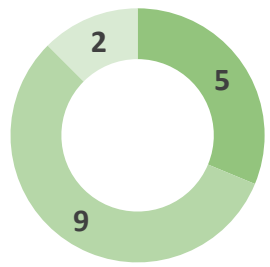
- Mandatory
- Recommended
- Optional



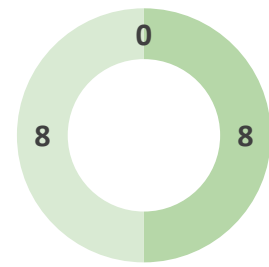
FAIR PRINCIPLES



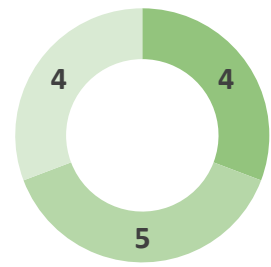
FINDABLE



ACCESSIBLE



INTEROPERABLE



REUSABLE

# Discussion items

## 1 GENERAL ISSUES / SCOPE

- What is the objective of evaluation: “How FAIR is this data?” versus “How can the FAIRness of this data be improved?” [\[GitHub Link\]](#)
- Does FAIR allow reuse based on human action or only machine-processable reuse? [\[GitHub Link\]](#)
- Can FAIR allow both direct as well as indirect identification: example of DOI. [\[GitHub Link\]](#)
- Allow for innovation, do not require existing standards to be applied in all cases. [\[GitHub Link\]](#)
- Should only FAIR-compliant vocabularies be used? How to manage the recursion? [\[GitHub Link\]](#)
- Metadata issues will be discussed at the joint meeting on the 25<sup>th</sup> of October | RDA Plenary in Helsinki

## 2 CATEGORISATION OF INDICATORS

- Mandatory/essential, recommended/important, optional/useful? [\[GitHub Link\]](#)
- Do we need both *Test if relevant* and *Mandatory if applicable* or only one of them? [\[GitHub Link\]](#)
- Is it useful or even possible to define a minimum set of indicators for FAIRness? [\[GitHub Link\]](#)



# 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

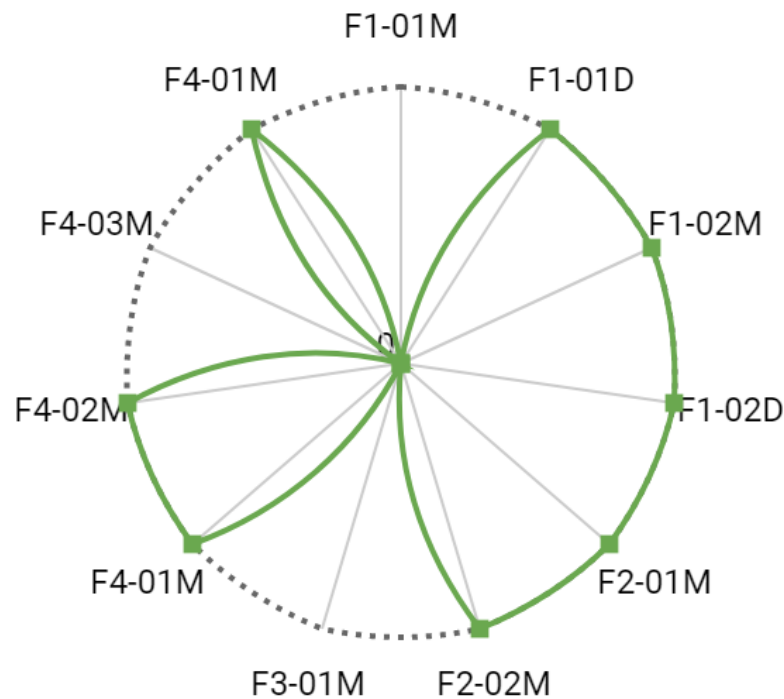
**Level 3** – The resource did comply with all the mandatory and recommended indicators, and less than half of the optional indicators

**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

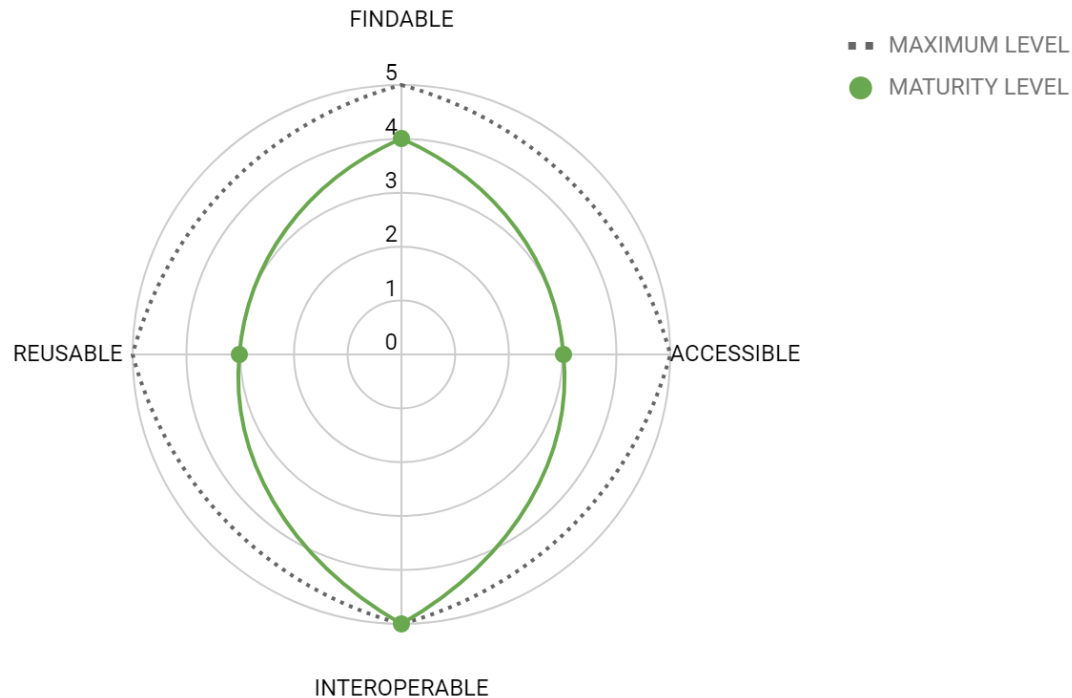
# Development | Scoring visualisation

## Indicators for Findability



■ Compliance with the indicator

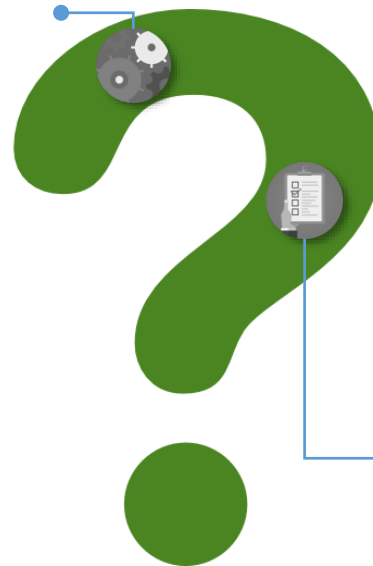
# Development | Scoring visualisation



FAIR Principles	MAXIMUM LEVEL	MATURITY LEVEL	AMBITION
FINDABLE	5	4	+1
ACCESSIBLE	5	3	+2
INTEROPERABLE	5	5	+0
REUSABLE	5	3	+2

# Development | Tool set and checklist

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



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





# Testing

# Testing the set of indicators

We identified two levels of testing;

## 1<sup>st</sup> 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

In scope for the WG

## 2<sup>nd</sup> Level

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

In scope for future work



Next steps

# Next steps

- Provide feedback to the proposals presented at the meeting of today on the [GitHub](#), if at all possible, **by the 11<sup>th</sup> October**
- Share feedback about consolidation and weighting of indicators and maturity levels on the [GitHub](#)
- Share feedback about the structure for tool set and data checklist on the [GitHub](#)

## WORKSHOP #5

RDA 14th Plenary session in Helsinki (FI)

**23<sup>rd</sup> October 2019**

Breakout 2 – **14.30 - 16.00** EEST

# 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](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!