



# FAIR Data Maturity Model

RDA-GEDE Expert meeting in Paris

28th October 2019

# Context



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**
- FAIR data **checklist**
- RDA recommendation

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

# Objectives

## FAIR data maturity model

● What are to be evaluated to determine FAIRness?

● Identify the indicators that can serve as core criteria

● Propose guidelines and a checklist

● Test the core criteria

Enable the development of automated tools  
for evaluation

Update the core criteria based on feedback

# Scope

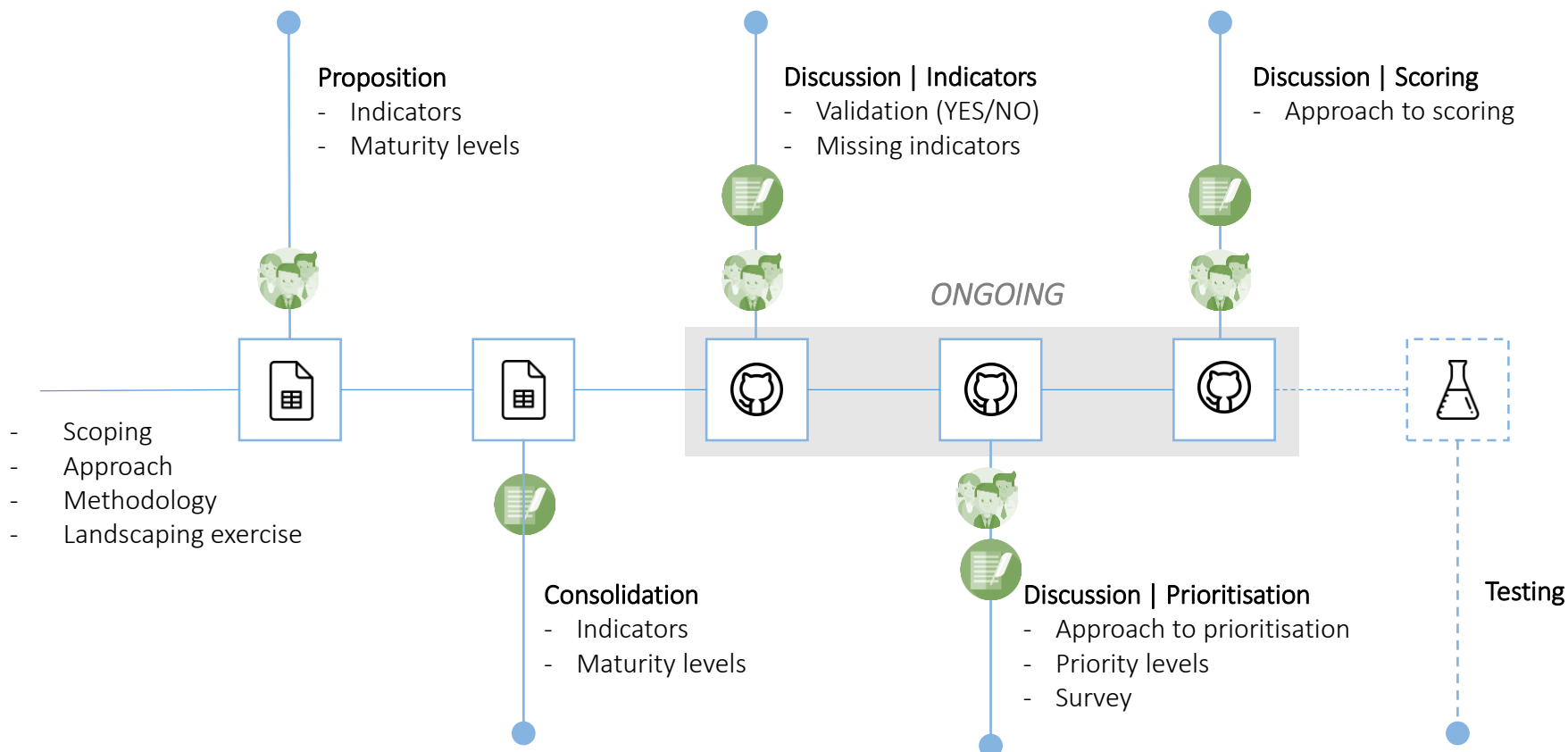
BUT the Working Group does NOT have the purpose to ...

- ⊗ **develop yet-another-evaluation-method:** the core criteria are intended to provide a common 'language' across evaluation approaches, not to be applied directly to datasets.
- ⊗ **define how the core criteria need to be evaluated.** The exact way to evaluate data based on the core criteria is up to the owners of the evaluation approaches, taking into account the requirements of their community
- ⊗ **revise and re-design the FAIR principles**



State of play

# State of play



# State of play

- ❗ Goal is to finalise indicators and priorities
- ❗ Indicators and priorities will be further used in their current state
- ❗ Indicators and priorities will be re-evaluated after the testing phase



# Development

## *First Phase*



# Overview | Indicators & levels

- Under discussion
- Provisionally agreed

F	<ul style="list-style-type: none"> <li>● F1 (Meta)data are assigned globally unique and persistent identifiers</li> <li>● F2 Data are described with rich metadata</li> <li>● F3 Metadata clearly and explicitly include the identifier of the data they describe</li> <li>● F4 (Meta)data are registered or indexed in a searchable resource</li> </ul>
A	<ul style="list-style-type: none"> <li>● A1 (Meta)data are retrievable by their identifier using a standardised communication protocol</li> <li>● A1.1 The protocol is open, free and universally implementable</li> <li>● A1.2 The protocol allows for an authentication and authorisation where necessary</li> <li>● A2 Metadata are accessible, even when the data are no longer available</li> </ul>
I	<ul style="list-style-type: none"> <li>● I1 (Meta)data use a formal, accessible, shared and broadly applicable language for knowledge representation</li> <li>● I2 (Meta)data use vocabularies that follow the FAIR principles</li> <li>● I3 (Meta)data include qualified references to other (meta)data</li> </ul>
R	<ul style="list-style-type: none"> <li>● R1 (Meta)data are richly described with a plurality of accurate and relevant attributes</li> <li>● R1.1 (Meta)data are released with a clear and accessible data usage license</li> <li>● R1.2 (Meta)data are associated with detailed provenance</li> <li>● R1.3 (Meta)data meet domain-relevant community standards</li> </ul>



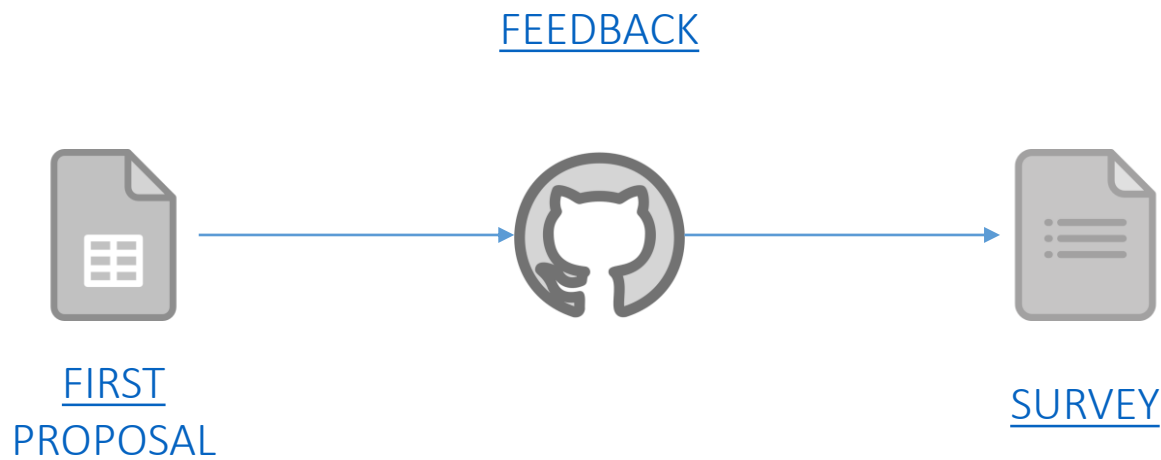
# Development

## *Second Phase*

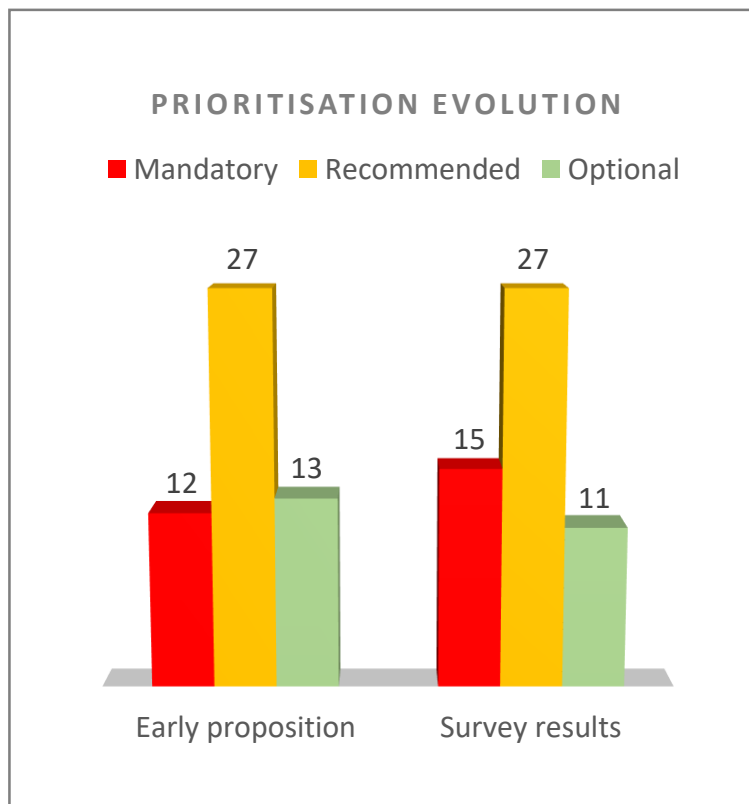
# Development | Weighting

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

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



# Development | Weighting



30  
participants

Notable results\*

- Metadata for discovery > **recommended** (F2)
- Metadata for reuse > **mandatory** (R1)
- (Machine-understandable) knowledge representation > **mandatory** for metadata & **recommended** for data (I1)
- All references to data > **optional** (I3)

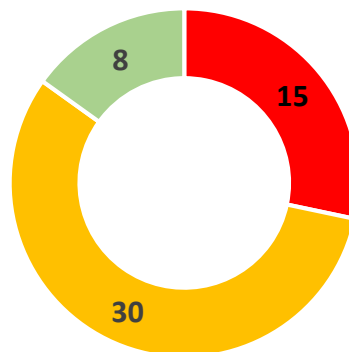
\* Results can be accessed [here](#)



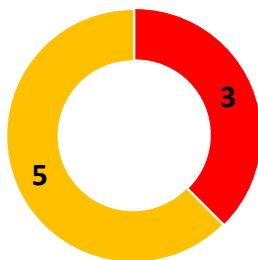
# Development | Weighting Stats

Distribution of the weight of the indicators

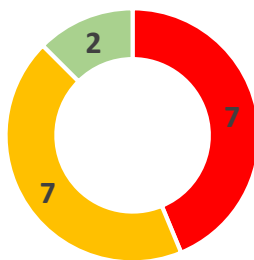
- Mandatory
- Recommended
- Optional



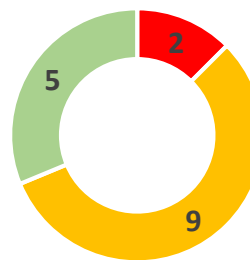
FAIR PRINCIPLES



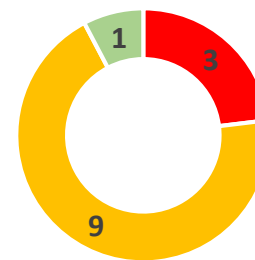
FINDABLE



ACCESSIBLE



INTEROPERABLE



REUSABLE



# Discussion

# Indicators | Discussions

Four discussion topics to choose from, two were discussed in Helsinki;

1

Identifier to point to data or a landing page

2

Machine-processable data versus human access

3

Linking and referencing

4

Two-speed FAIRness



slido.com #8935



# Development

## *Next steps*



# Development | Scoring

As presented during Workshop #3

## 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

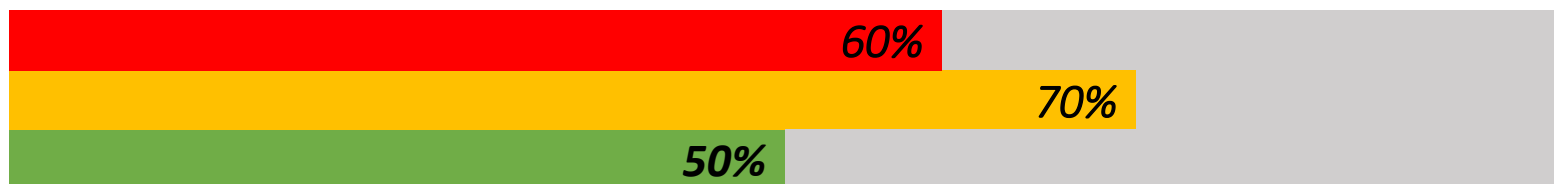


# Development | Scoring\*

Triple overall FAIR score and levels for FAIR areas

*May be too crude and could be misused*

## OVERALL FAIRness



- Mandatory
- Recommended
- Optional

\***Proposal** discussed on [GitHub](#)



# Development | Scoring

FAIRness per area

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

- None of the indicators are satisfied
- ◐ Half of the indicators are satisfied
- All indicators are satisfied



# Development | Guidelines, checklist and next steps



## GUIDELINES

- Context
- Framework
- Indicators
  - *Description*
  - *Examples*
  - *Prioritisation*
- Utility and utilization
- Integration with other initiatives
- Continuity



## CHECKLIST

**Summary of the guidelines;** focus on the key elements considered to be FAIR compliant and improve reusability

# Testing the set of indicators

As presented during workshop #3, **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

- We encourage you to share any feedback in the [GitHub](#)
  - Indicators
  - Prioritisation
  - [Scoring](#)
  - Next steps

## WORKSHOP #6

4 December 2019

09.00 - 10.30 CET | *Morning session*

17.00 - 18.30 CET | *Afternoon session*

# 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 – **Indicators prioritisation survey results**

[https://drive.google.com/open?id=11hyAYCKz\\_NV0Ob9-vlPqjN9LCarOFmc3](https://drive.google.com/open?id=11hyAYCKz_NV0Ob9-vlPqjN9LCarOFmc3)

- RDA FAIR data maturity model WG – **Mailing list**

[fair\\_maturity@rda-groups.org](mailto:fair_maturity@rda-groups.org)





Thank you!