FAIR Data Maturity Model

14th RDA Plenary session - Workshop #5

23rd October 2019

Agenda

14:30 – 14:40	Welcome, objectives of the meeting
14:40 – 14:45	Roundtable
14:45 – 14:50	State of play
14:50 – 14:55	Development – First phase
14:55 – 15:00	Development – Second phase
15:00 – 15:45	Panel discussion| Consensus
15:45– 15:50	Development – Next steps
15:55– 16:00	Next steps and Q&A

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

SOLUTION is to bring together **stakeholders** to build on **existing approaches** and **expertise**
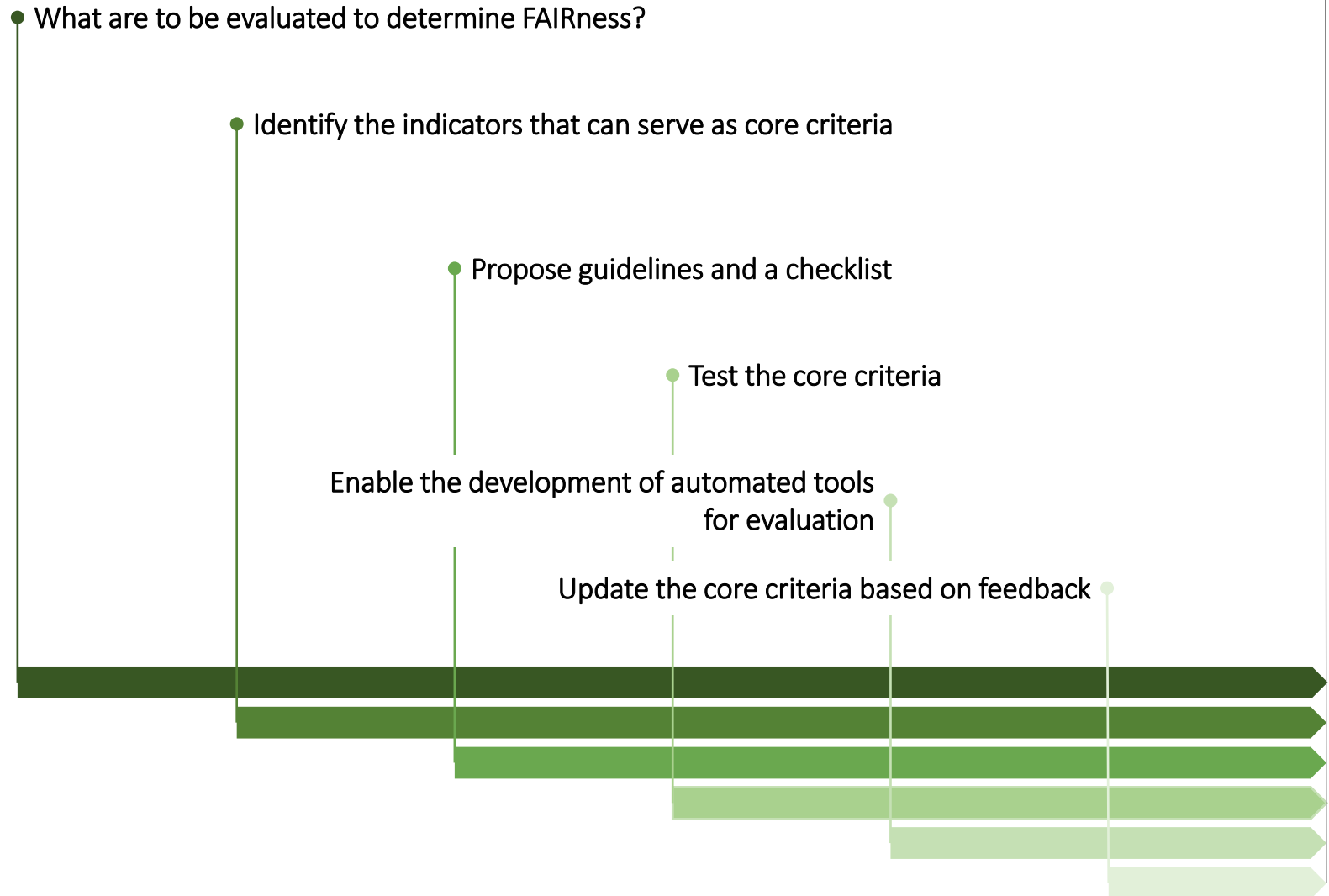
- Set of **core assessment criteria** for FAIRness
- FAIR **data maturity model & toolset**
- FAIR data **checklist**
- RDA recommendation

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 <u>NOT</u> 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**

Roundtable

By show of hands

> Which region?

> Your role
>> Researcher
>> Librarian
>> Infrastructure manager
>> Policy developer
>> Research funder

> Introducing the editorial team

If you are dialing in, please type your name and affiliation in the chat window

Join at slido.com

Ask questions and vote in live polls

MEETING CODE **#8935**
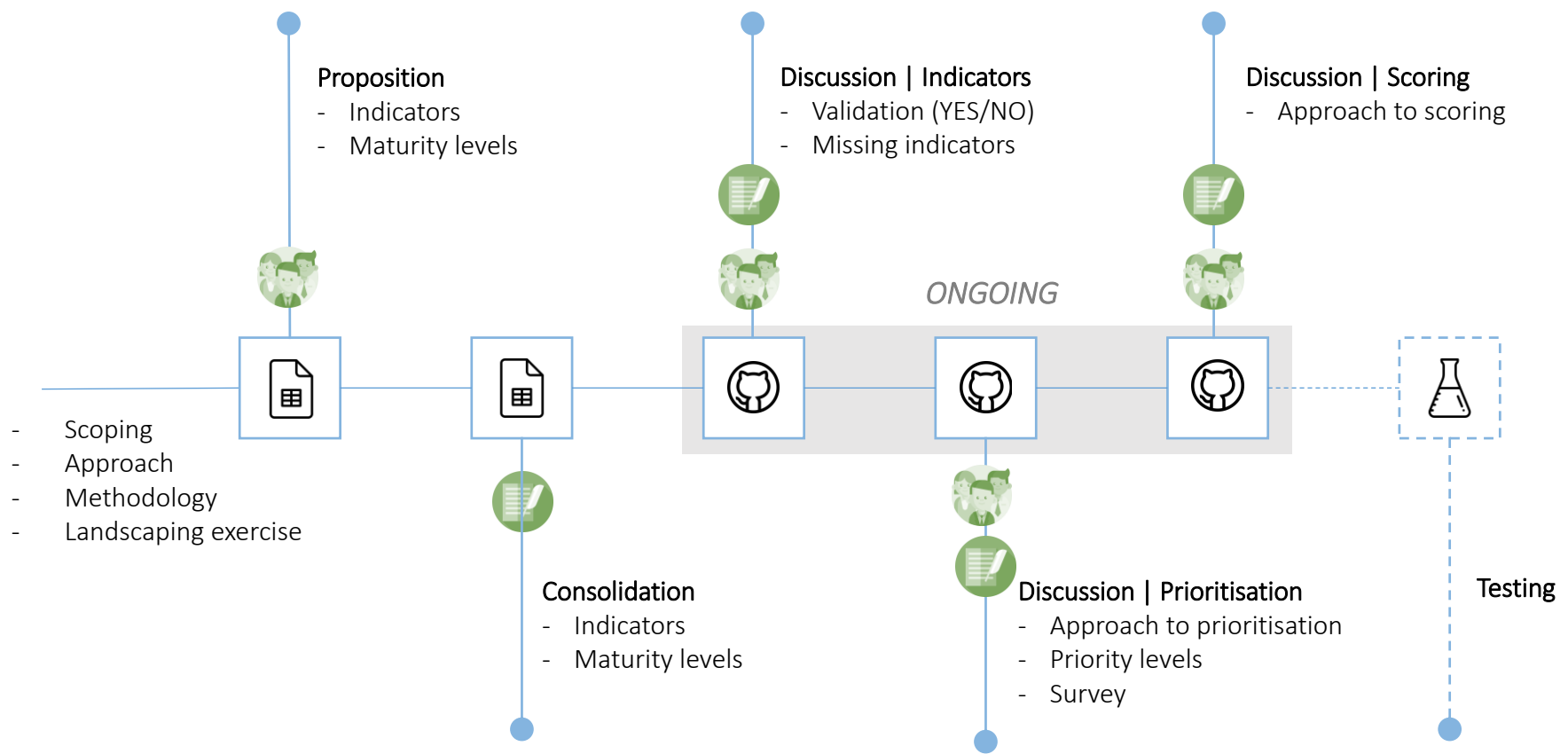
State of play

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

- 🟢 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

Overview | Indicators & levels

FAIR PRINCIPLES

REUSABLE · FINDABLE · INTEROPERABLE · ACCESSIBLE
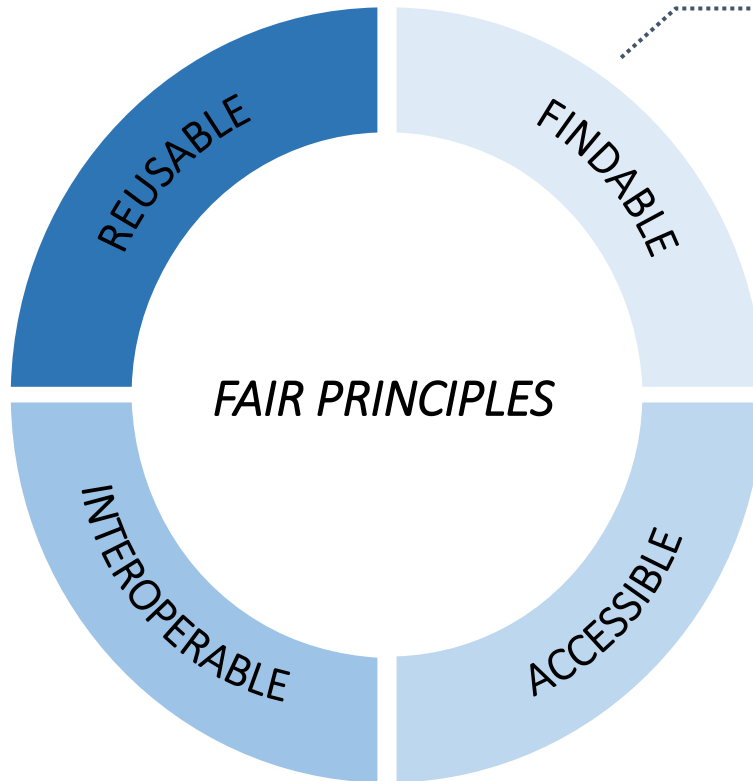
Indicators for Findability

- [F1-01M] Metadata is identified by a persistent identifier
- [F1-01D] Data is identified by a persistent identifier
- [F1-02M] Metadata is identified by a universally unique identifier
- [F1-02D] Data is identified by a universally unique identifier
- [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
- [F3-01M] Metadata includes the identifier for the data
- [F4-01M] Metadata is offered/published/exposed in such a way that it can be harvested and indexed

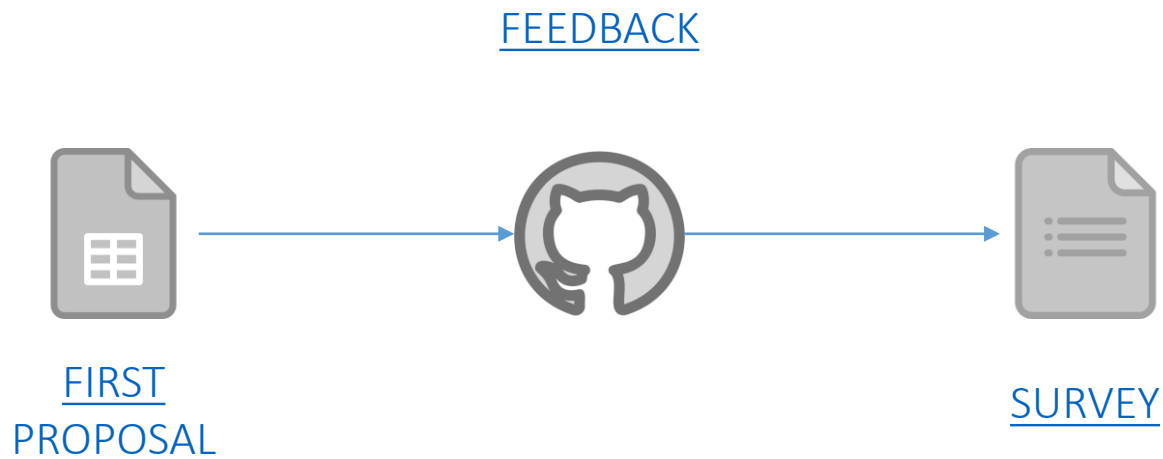* The full list of indicators can be found on the following GSheet
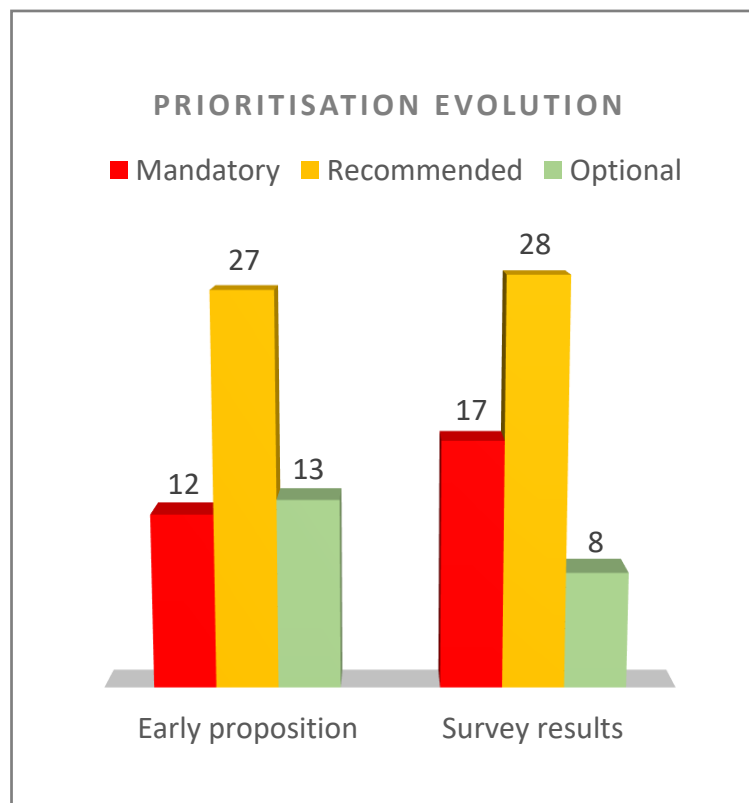
Development
Second Phase

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)

FEEDBACK

FIRST
PROPOSAL

SURVEY

Development | Weighting

PRIORITISATION EVOLUTION

■ Mandatory ■ Recommended ■ Optional

Early proposition: Mandatory 12, Recommended 27, Optional 13
Survey results: Mandatory 17, Recommended 28, Optional 8

| 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)

slido.com #8935

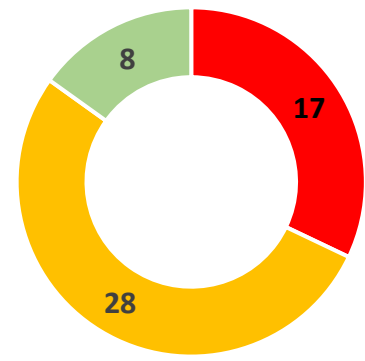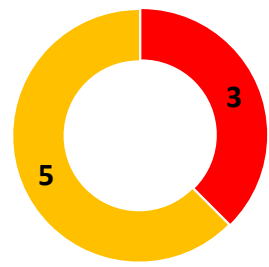* Results can be accessed here

Development | Weighting Stats

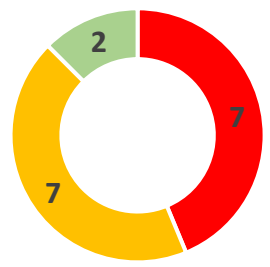Distribution of the weight of the indicators

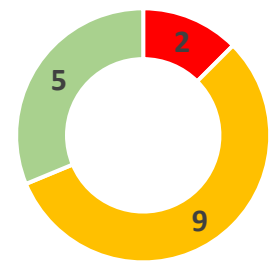Legend:
- 🔴 Mandatory
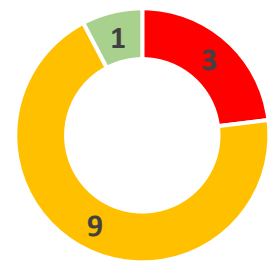- 🟡 Recommended
- 🟢 Optional

FAIR PRINCIPLES

FINDABLE ACCESSIBLE INTEROPERABLE REUSABLE

Discussion

Indicators | Discussions

Four discussion topics to choose from, <u>one to be addressed today</u>;

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

Indicators | Discussions

1 Identifier to point to data or landing page?

A	OPPOSED VIEWS	B
FAIRness requires separate unique and persistent identifiers for metadata and data	*OR*	Data published with an identifier (e.g. DOI) pointing to a landing page with embedded metadata and a URL to access the data can also be considered FAIR

slido.com #8935

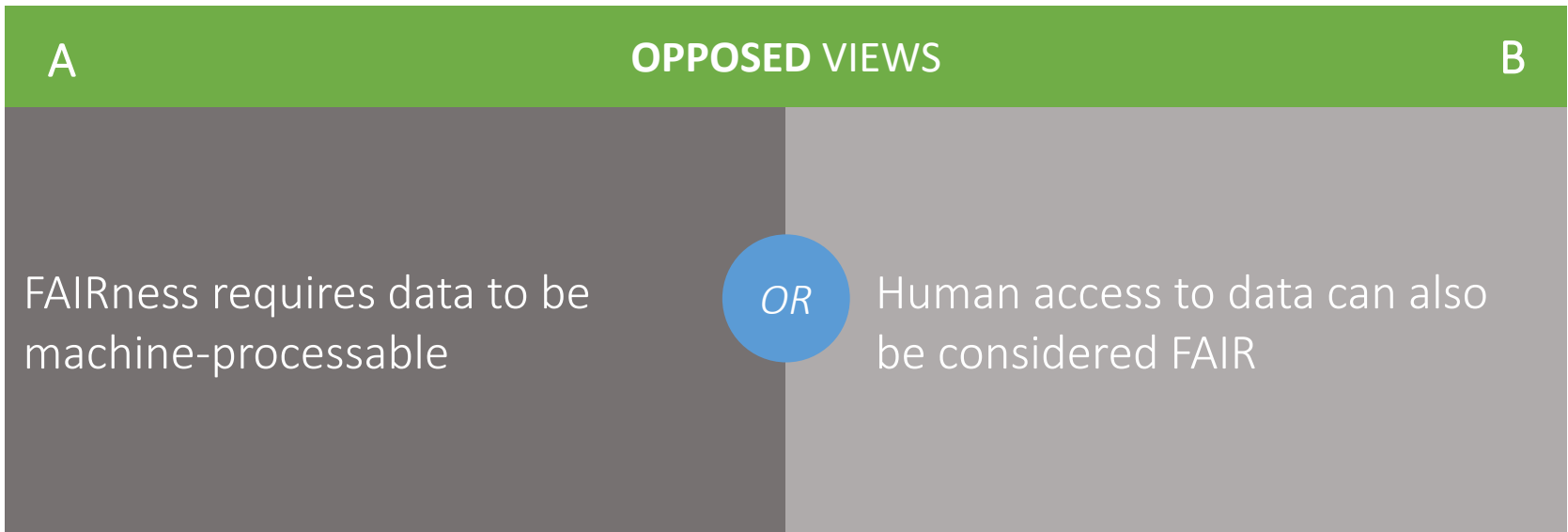# Indicators | Discussions

2 Machine-processable data versus human access

A	**OPPOSED** VIEWS	B
FAIRness requires data to be machine-processable	*OR*	Human access to data can also be considered FAIR

slido.com #8935

Indicators | Discussions

3 Linking and referencing

A	OPPOSED VIEWS	B
FAIRness requires rich linkages between metadata and other metadata, between metadata and other data between data and other data	OR	Metadata and data that have no links to other metadata and other data can also be considered FAIR

slido.com #8935

Indicators | Discussions

4	Two-speed FAIRness

A	OPPOSED VIEWS	B
FAIRness should aim for the same level of FAIRness for all domains in the long term	*OR*	Different communities need to be able to define their own target FAIRness levels

slido.com #8935

Development
Next steps

Development | Scoring

Core assessment criteria to evaluate and compare FAIRness

> FAIRness report for a resource under evaluation
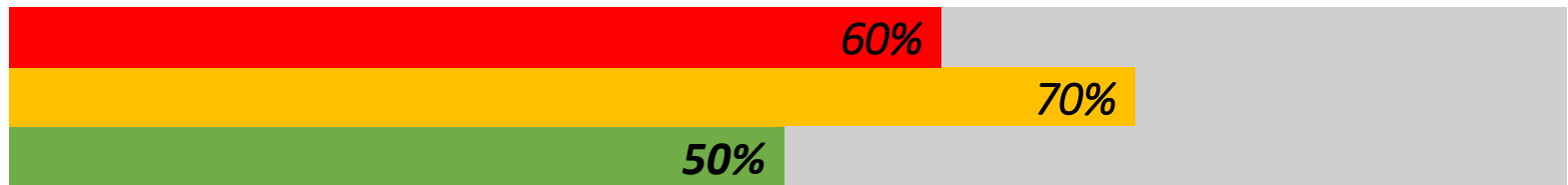
> > Indicators classified per importance

> 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*

Triple overall FAIR score and levels for FAIR areas
May be too crude and could be misused

OVERALL FAIRness

60%	
70%	
50%	

🔴 Mandatory

🟡 Recommended

🟢 Optional

*Proposal discussed on GitHub

slido.com #8935

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

slido.com #8935

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;

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

In scope for the WG

2nd 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

> 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_NVoOb9-vlPqjN9LCarOFmc3

> RDA FAIR data maturity model WG – Mailing list

fair_maturity@rda-groups.org

Thank you!