RDA and the FAIR Data Maturity Model Working Group

Edit Herczog RDA Global Council
Co-chair of RDA FAIR DMM WG

Agenda

1. RDA Global Foundation
2. GEDE Group
3. FAIR and the European Open Science
4. FAIR Data Maturity Model WG (FAIR DMM WG)
5. FAIR Maturity Model Criteria
6. FAIR DMM WG Workplan
7. FAIR DMM WG status of the work
8. Invitation to be a member and TESTING COMMUNITY

RDA GLOBAL FOUNDATION

THE RESEARCH DATA ALLIANCE

www.rd-alliance.org

building the social and technical bridges that enable open sharing of data

32 FLAGSHIP OUTPUTS

of which 4 ICT Technical Specifications

75 ADOPTION CASES

across multiple disciplines, organisations & countries

104 GROUPS WORKING ON GLOBAL DATA INTEROPERABILITY CHALLENGES

of which 35 WORKING GROUPS & 67 INTEREST GROUPS

8,642 INDIVIDUAL MEMBERS FROM 137 COUNTRIES

67,9% Academia & Research
14,2% Public Administration
12,7% Enterprise & Industry

50 ORGANISATIONAL MEMBERS & 8 AFFILIATE MEMBERS

RDA
RESEARCH DATA ALLIANCE

Vision

Researchers and innovators openly share data across technologies, disciplines, and countries to address the grand challenges of society.

Mission

RDA builds the social and technical bridges that **enable open sharing** of data.

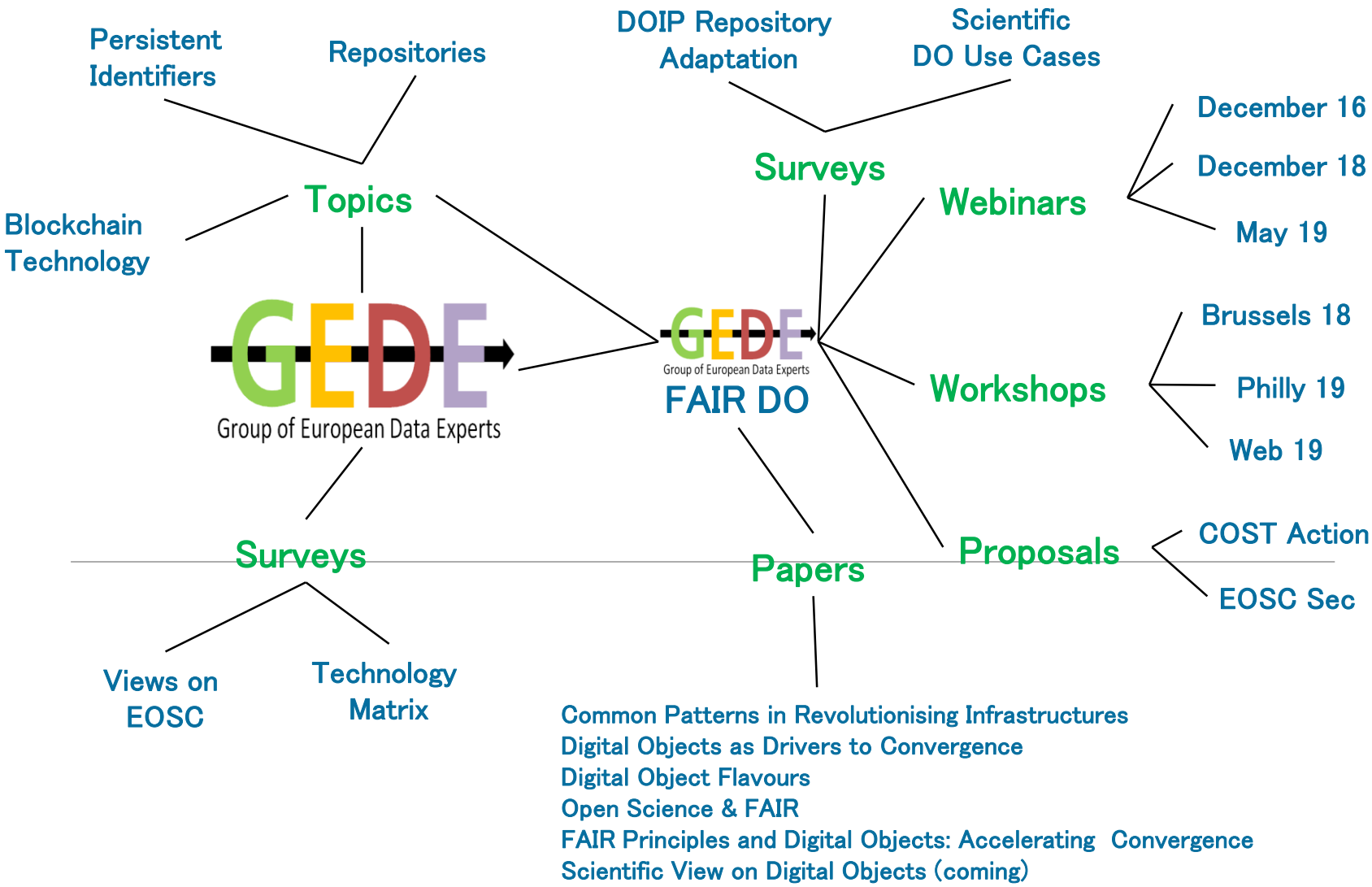
Why Join RDA as an Individual Member?

Individual Member Benefits

- ***Contribute*** *to acceleration of data infrastructure development*
- *Work and* ***share experiences*** *with collaborators throughout the world*
- ***Access*** *to extraordinary network of colleagues with various levels of experience, perspectives and practices*
- *Gain greater* ***expertise*** *in data science regardless of whether one is a student, early or seasoned career professional*
- ***Enhance*** *the quality and effectiveness of personal work and activities*
- ***Improve*** *one's competitive advantage professionally and positioning oneself for leadership within the broader research community*

THE GEDE GROUP

RDA SPECIAL GROUP (FROM ESFRI-S PLUS)



FAIR AND THE EUROPEAN OPEN SCIENCE

Open Science in the EU is a strategy by choice to respond the challenges

1. The Brutal reality inside the sector
   - Less than 20% of data measured by researchers is recorded
   - Less than 20% of the recorded data is resused at least once
   - 80% of researchers is used to gather the data they need
2. The Digitisation of the Research sector
   - Sensor enabled nature and society turns the Earth to a Laboratory, with real time data generated continuously
   - Laboratories with super equipment generate immense data
   - Computer ecosystem and AI needs high quality data
3. Research is capital, energy, and talent intensive, while resources are limited
   - Step change in effectiveness and efficiency is crucial
   - Sustainability is the name of the game
4. The EU prosperity should be Research and Science Data driven
   - economic, industrial and social policies, require better research

FAIR DMM WG for the EU HE

Based on an internet-type hour-glass model, the EOSC will need community-endorsed, internationally governed and enforceable set of protocols.

These protocols should be :

1. absolutely minimal, open and transparent so that all scientists, innovators, engineers and service providers understand them, see their value and can adhere to them, even if technology and data formats rapidly develop (as will be the case) tuned down to the very basics of what data and related services need, what they support at the most basic level and only where strictly necessary to make the EOSC work (comparable to TCP/IP, HTTP and HTML for the Internet)
2. Count for all Research Objects and they should enable the minimal requirements for Research Objects to be widely and effectively (re)-used
3. The FAIR principles14 will guide implementations to make research objects Findable, Accessible, Interoperable, ultimately to make them Re-usable and citable15
4. Within the scope of FAIR principles, the standards and protocols should again be restricted to the absolute minimum, to mitigate the risk that future developments will require adaptations of protocols

FAIR DATA MATURITY MODEL WG

Who we are

WG started the WG in January 2019

First plenary session at P13 in Philadelphia

Co chairs:
- Keith Russel from Australia
- Edit Herczog from Europe
- Shelley Stall from USA

TAB member:
- Jane Wyngaard from South Africa

Secretariat: Yolanda from USA

Editorial team: EC special support
- Makx Dekkers and the PWC team

129 members: 61 Female, 68 male

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

Case statement of the WG

Challenge

Ambiguity and wide range of interpretations of FAIRness

Lack of a common set of core assessment criteria and a minimum set of shared guidelines

Approach

Bring together stakeholders

Build on existing approaches and expertise

Intended results

RDA Recommendation of core assessment criteria

Generic and expandable self-assessment model

Self-assessment toolset

FAIR data checklist

Case statement of the WG

Target audiences

- Researchers, data stewards, other data professionals
- Data service owners, e.g. infrastructure, repositories
- Organisations that manage research data
- Policymakers

Connections

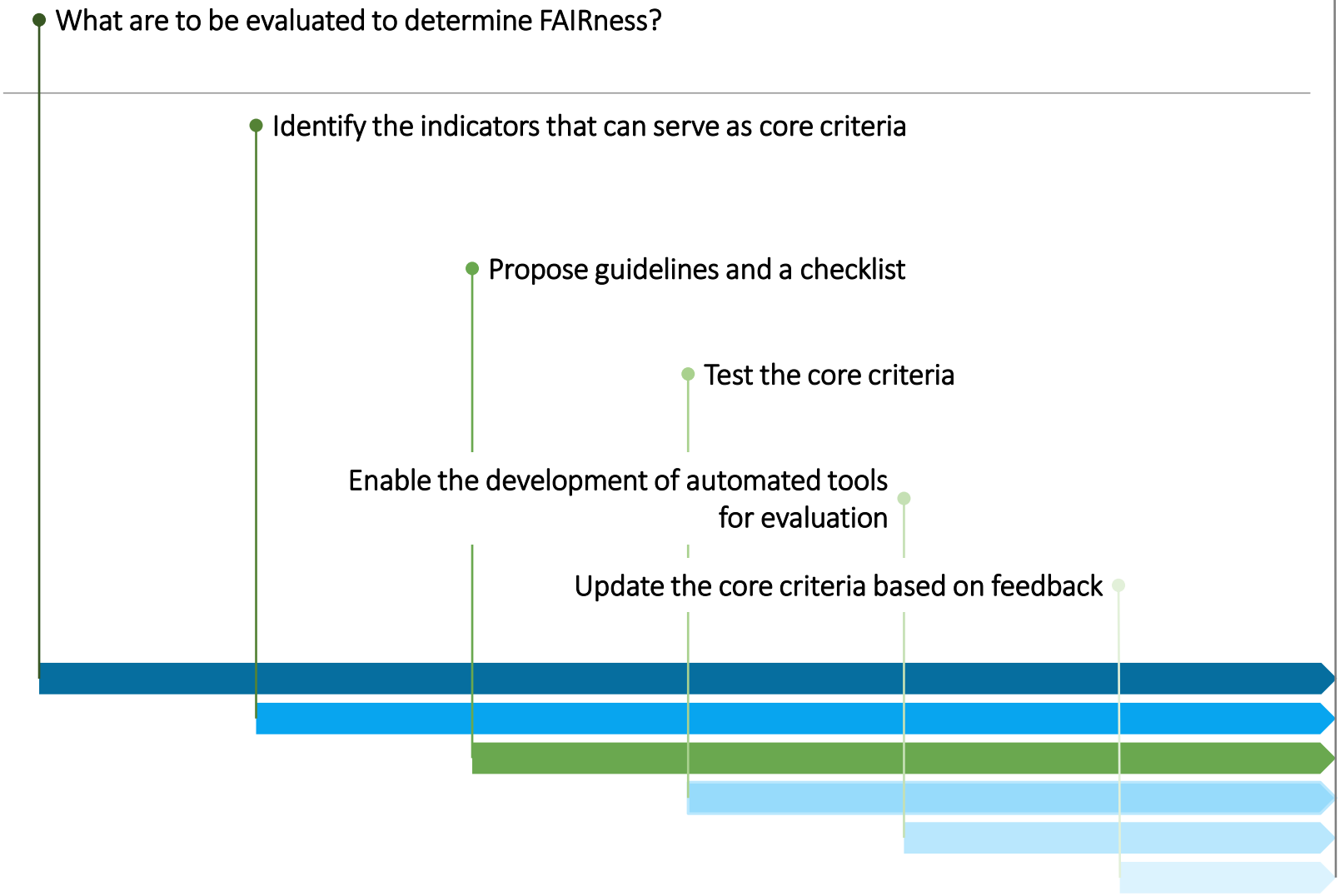
- RDA Disciplinary Framework Interest Group
- RDA Domain Repositories Interest Group
- Other RDA groups

Scope of the assessment

- Datasets
- Data-related aspects (e.g. algorithms, tools, workflows)

Objectives

FAIR data maturity model



Minimum CORE criteria

WHAT **NOT HOW**

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**

Criteria elements condensed from the FAIR principals

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

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

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

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

Proposed development methodology

Bottom-up approach comprising 4 phases

Definition

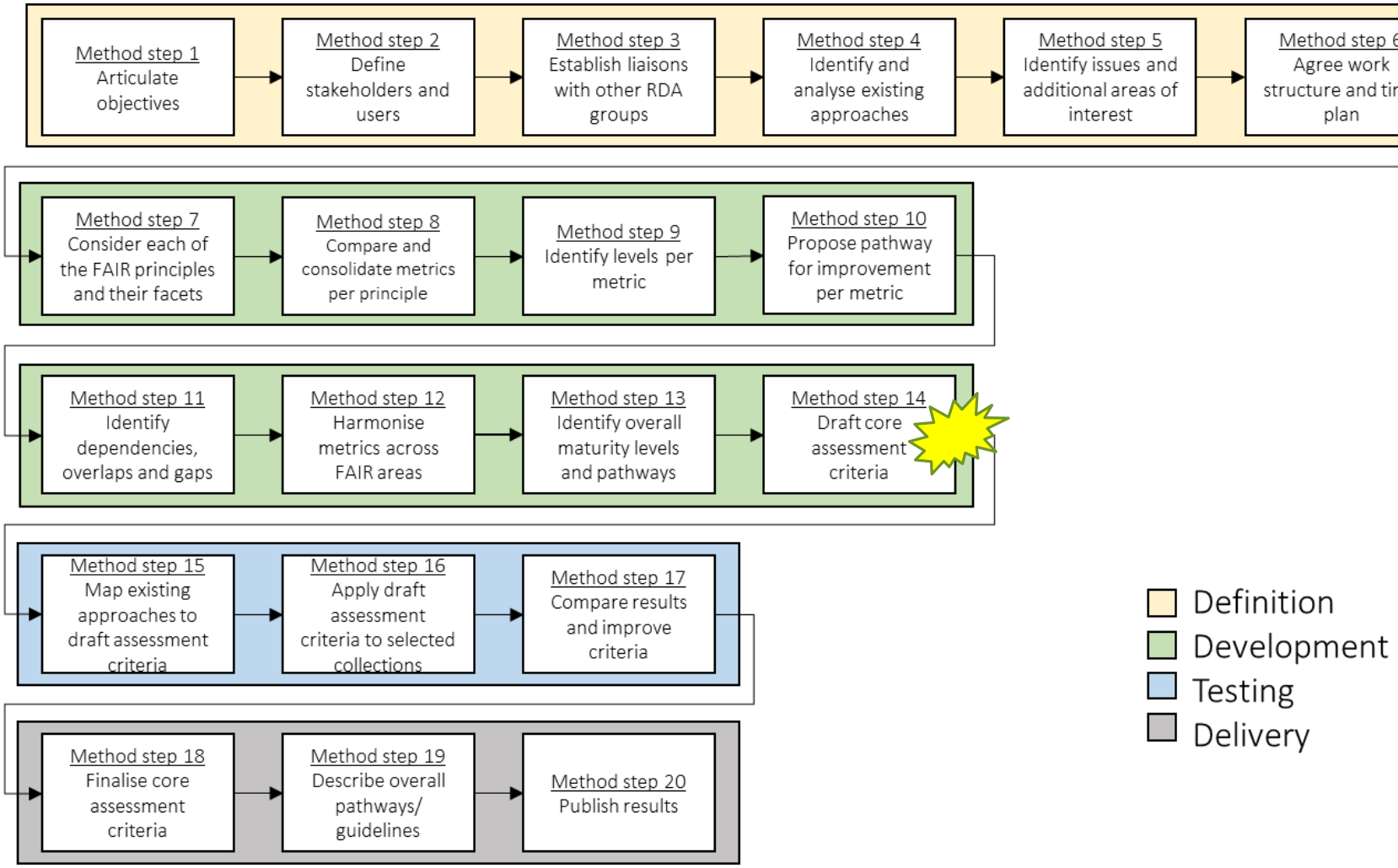
Development

- Assessment of the four FAIR principles in four ‘strands’
- Fifth ‘strand’: beyond the FAIR principles

Testing

Delivery

Overview of the methodology



Results of preliminary analysis - 2

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

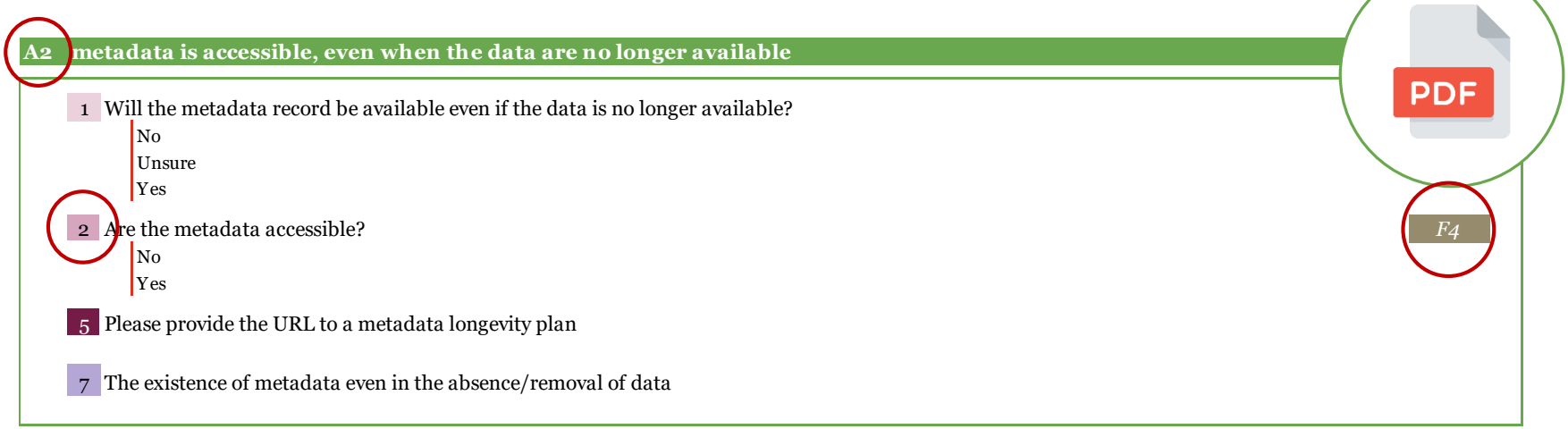
Results of preliminary analysis - 4

Five slide decks classifying questions

- FAIR – Findable [Link]
- FAIR – Accessible [Link]
- FAIR – Interoperable [Link]
- FAIR – Reusable [Link]
- Beyond the FAIR principles (X) [Link]

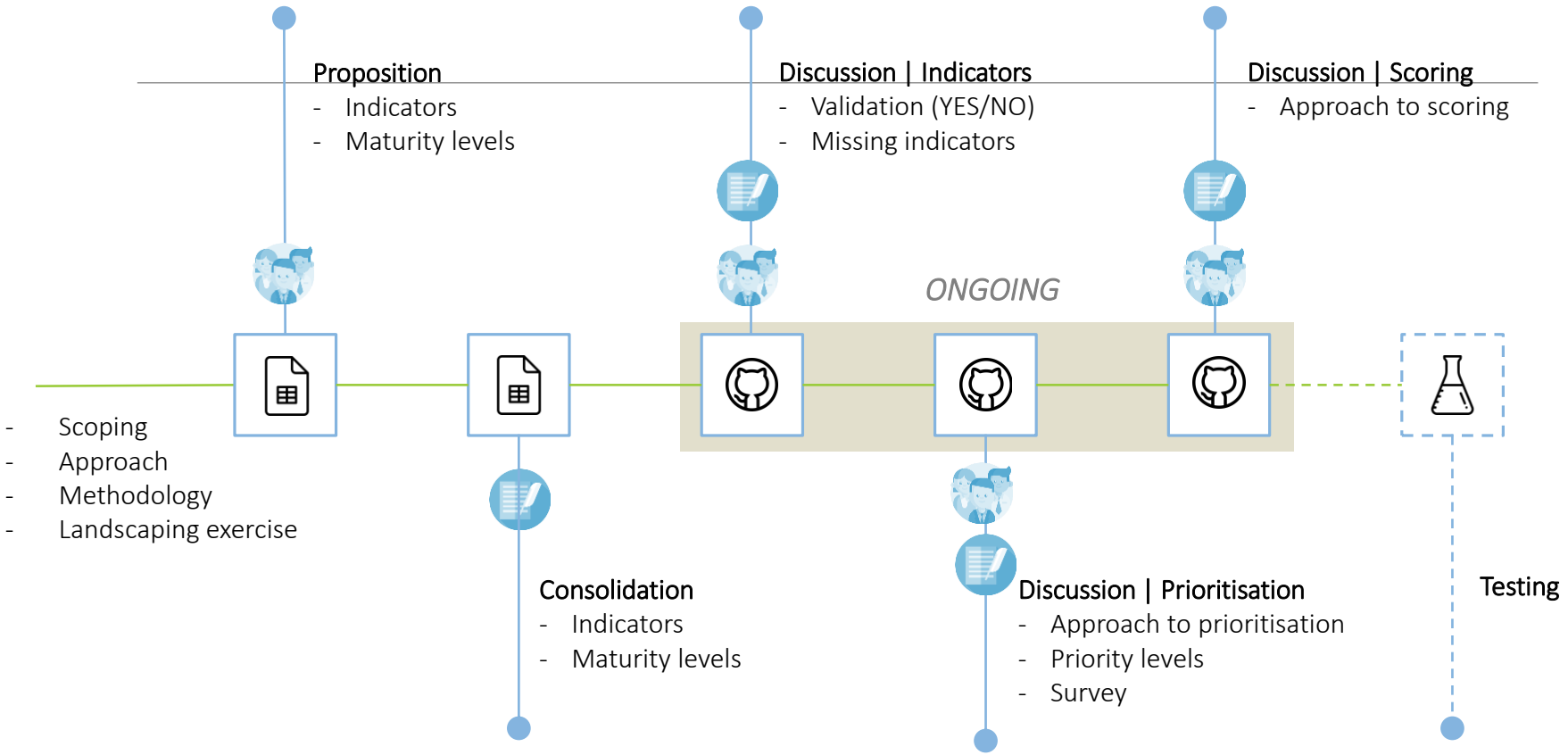
Example

Questions, options and potential overlaps



State of play

State of play

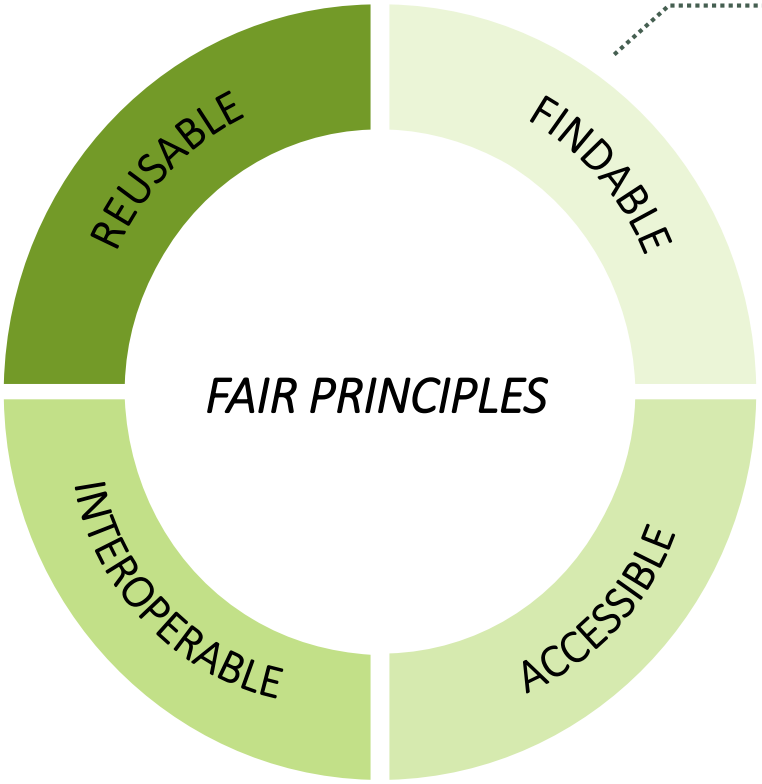


Overview | Indicators & levels

- 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



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

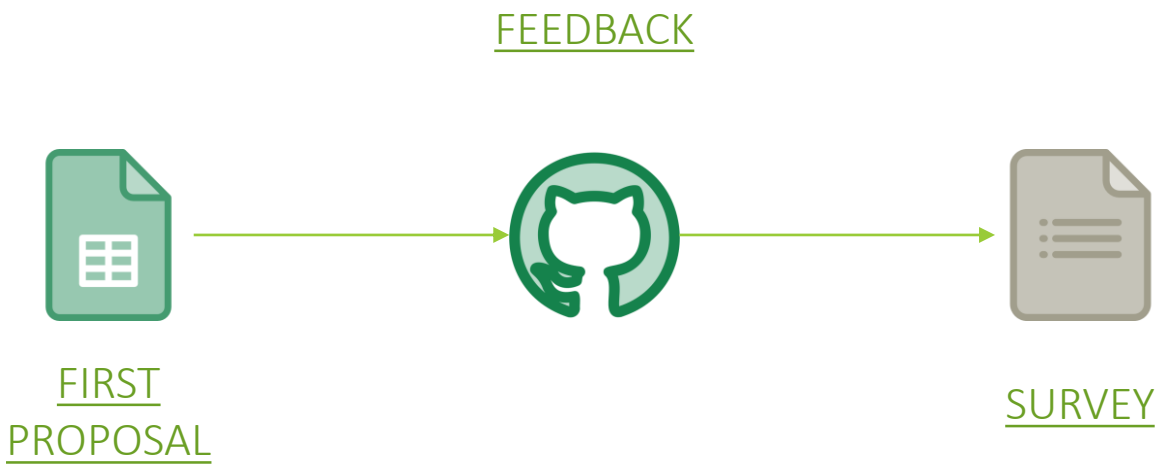
* The full list of indicators can be found on the following GSheet

Development
Second Phase

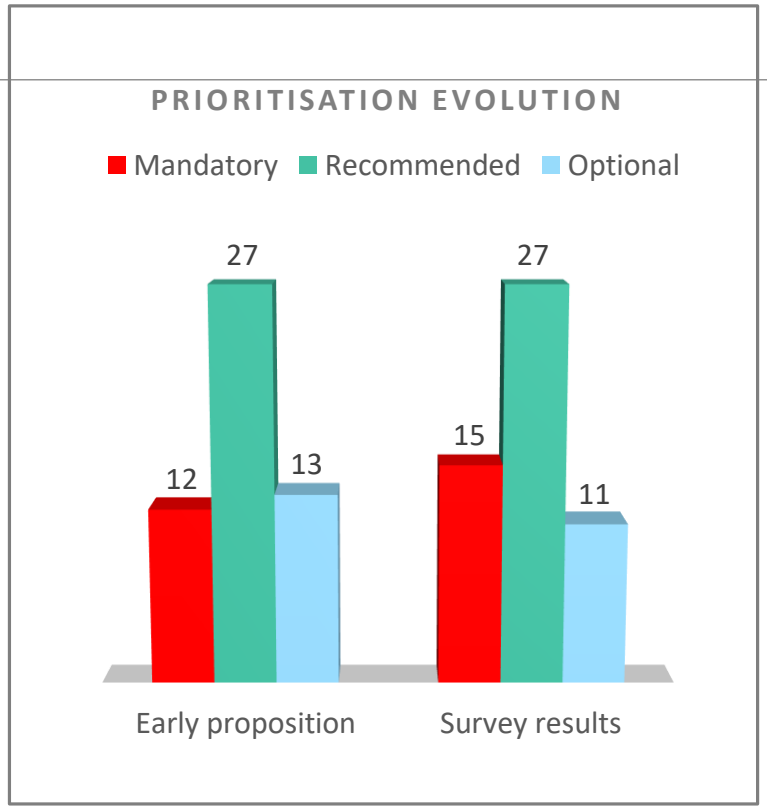
Development | Weighting

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)



Development | Weighting



30
participants

Notable results*

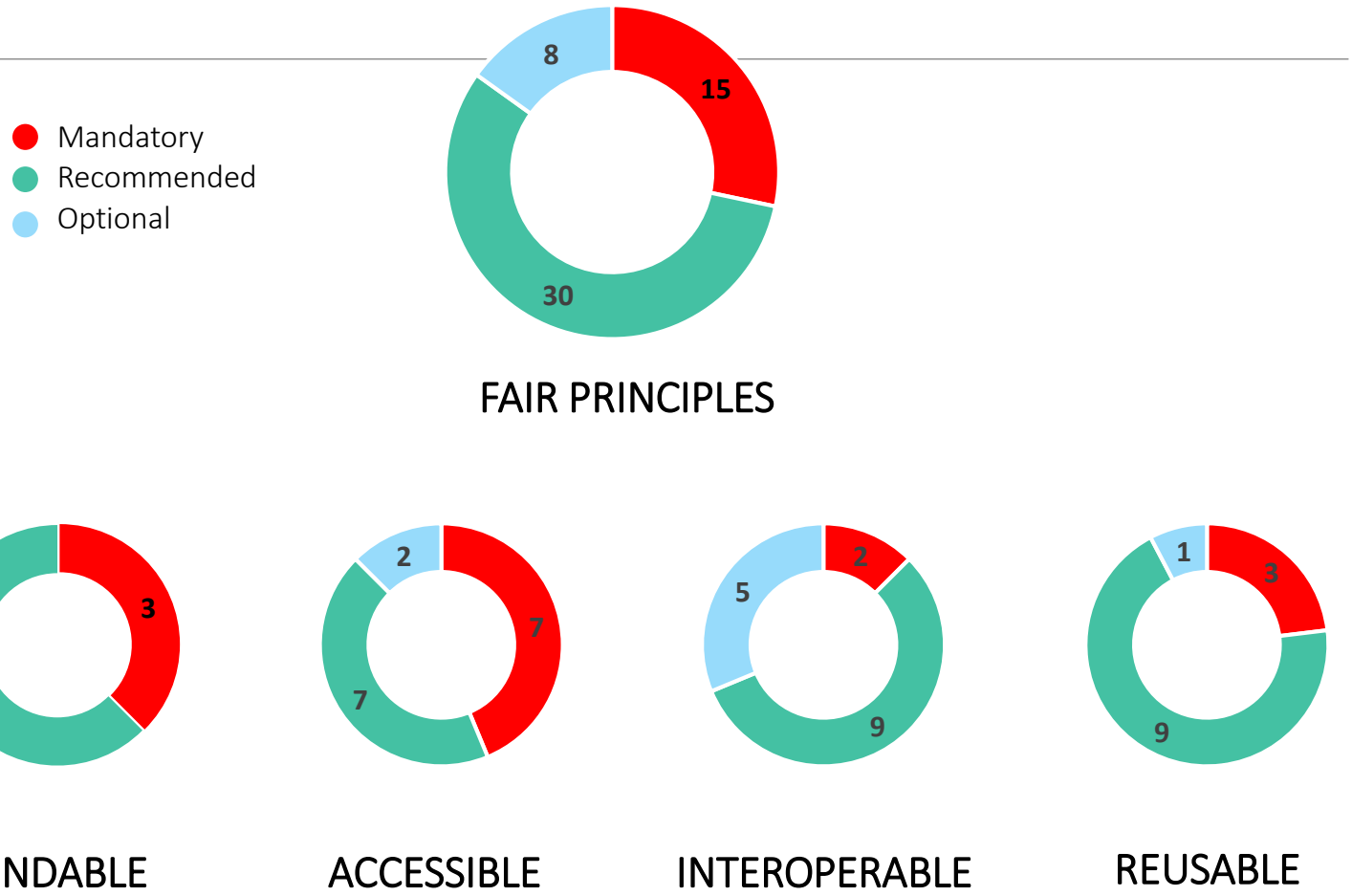
- Metadata for discovery > **recommended** (F2)
- Metadata for reuse > **mandatory** (R1)
- (Machine-understandable) knowledge representation > **mandatory** for metadata & **recommended** for data (I1)
- All references to data > **optional** (I3)

slido.com #8935

* Results can be accessed here

Development | Weighting Stats

Distribution of the weight of the indicators



Discussion

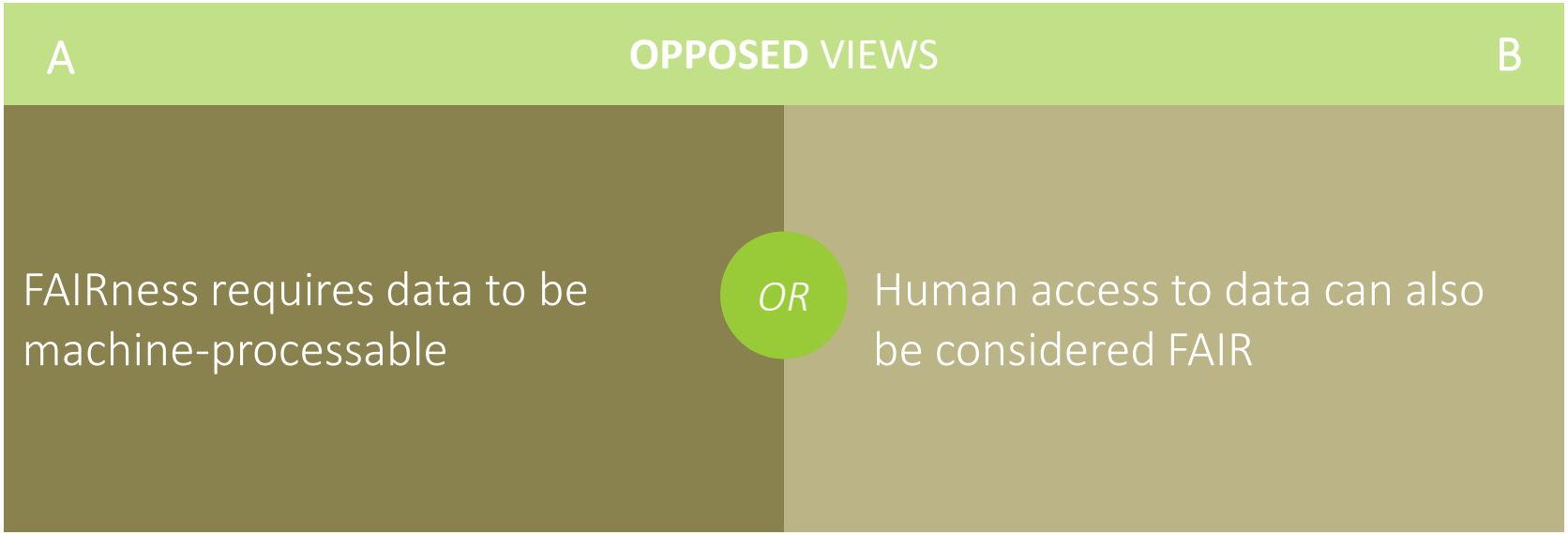
Indicators | Discussions

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

Indicators | Discussions

2 Machine-processable data versus human access



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

Development

Next steps

Development | Scoring

As presented during Workshop #3

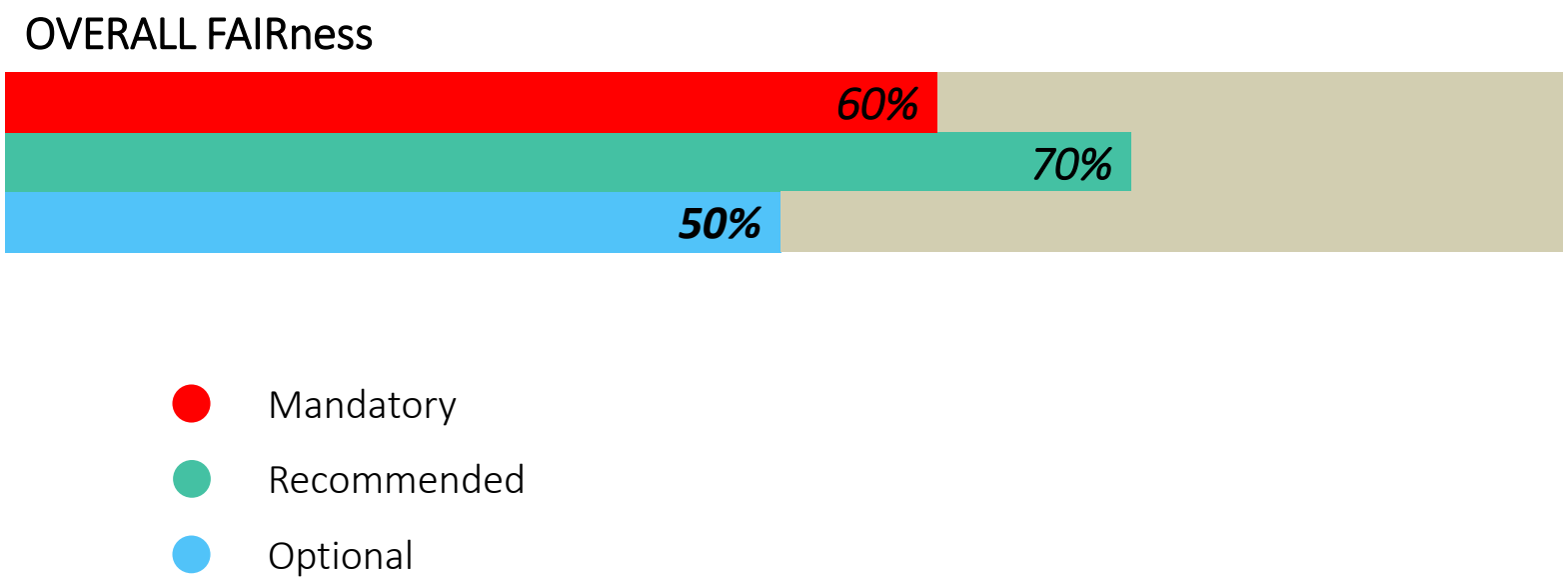
Core assessment criteria to evaluate and compare FAIRness

- 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



*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

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

RDA Plenary14 Helsinki FAIR Data Maturity Model WG meeting final report :
https://drive.google.com/file/d/1hzpFdVZsRwDzu3uaml_Pr9IdjObM4z-B/view

Thank you!