



RESEARCH DATA ALLIANCE

# FAIR Data Maturity Model

Workshop #7

13th February 2020

# Agenda

5'	Welcome, objectives of the meeting
5'	Roundtable
5'	State of play
5'	Testing phase overview
25'	Testing insights – feedback
25'	Testing insights – general discussion
15'	Potential scoring mechanisms
5'	Action items and next steps

# 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**

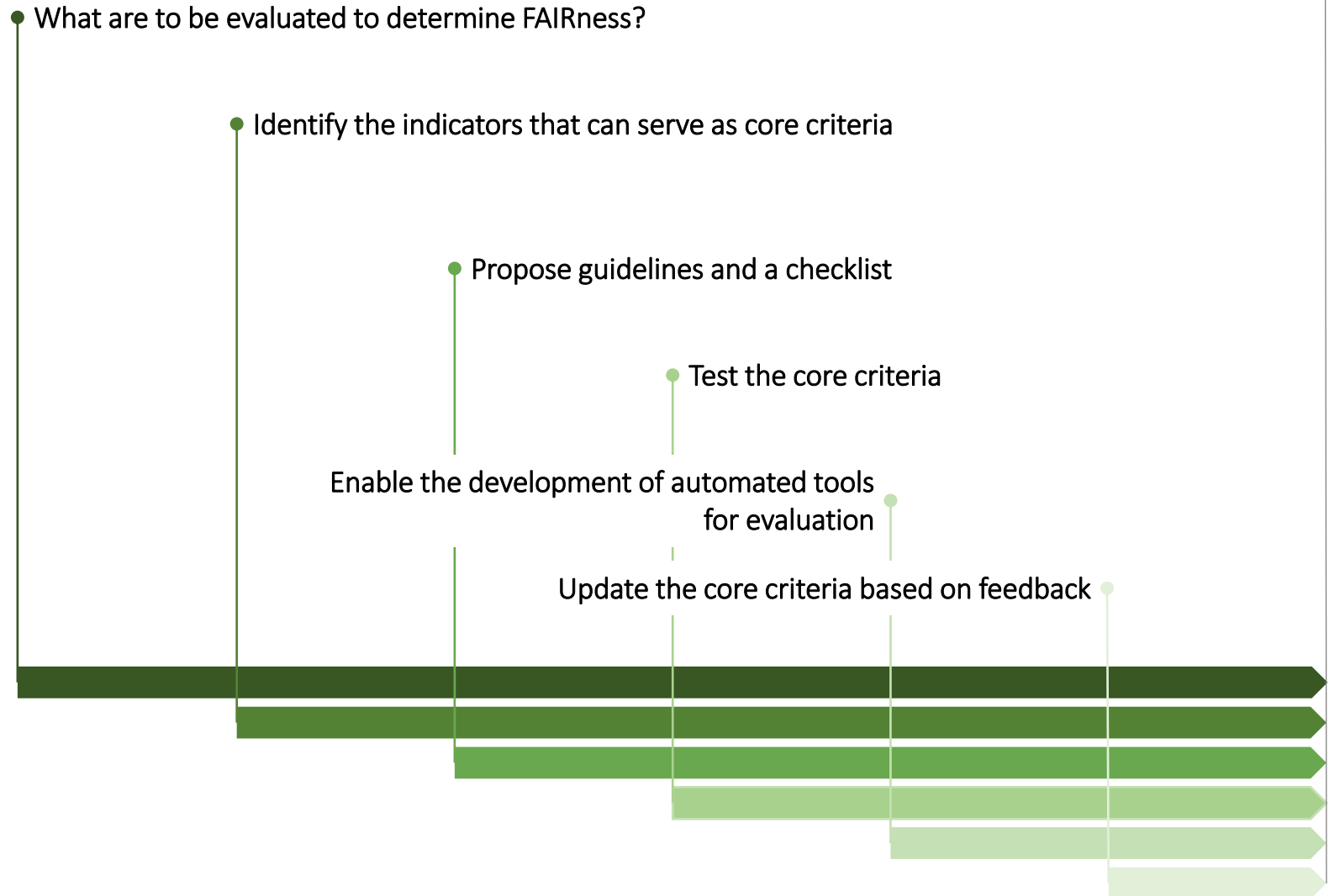


- Set of **core assessment criteria** for FAIRness
- FAIR **data maturity model & toolset**
- FAIR data **checklist**
- RDA recommendation

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

# Objectives

## FAIR data maturity model



# 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**

# Roundtable

In the chat window, please type...

- Your name
- Your affiliation
- Your role
  - Researcher
  - Librarian
  - Service provider
  - Policy maker
  - Funder
  
- Introducing the editorial team



State of play


# State of play

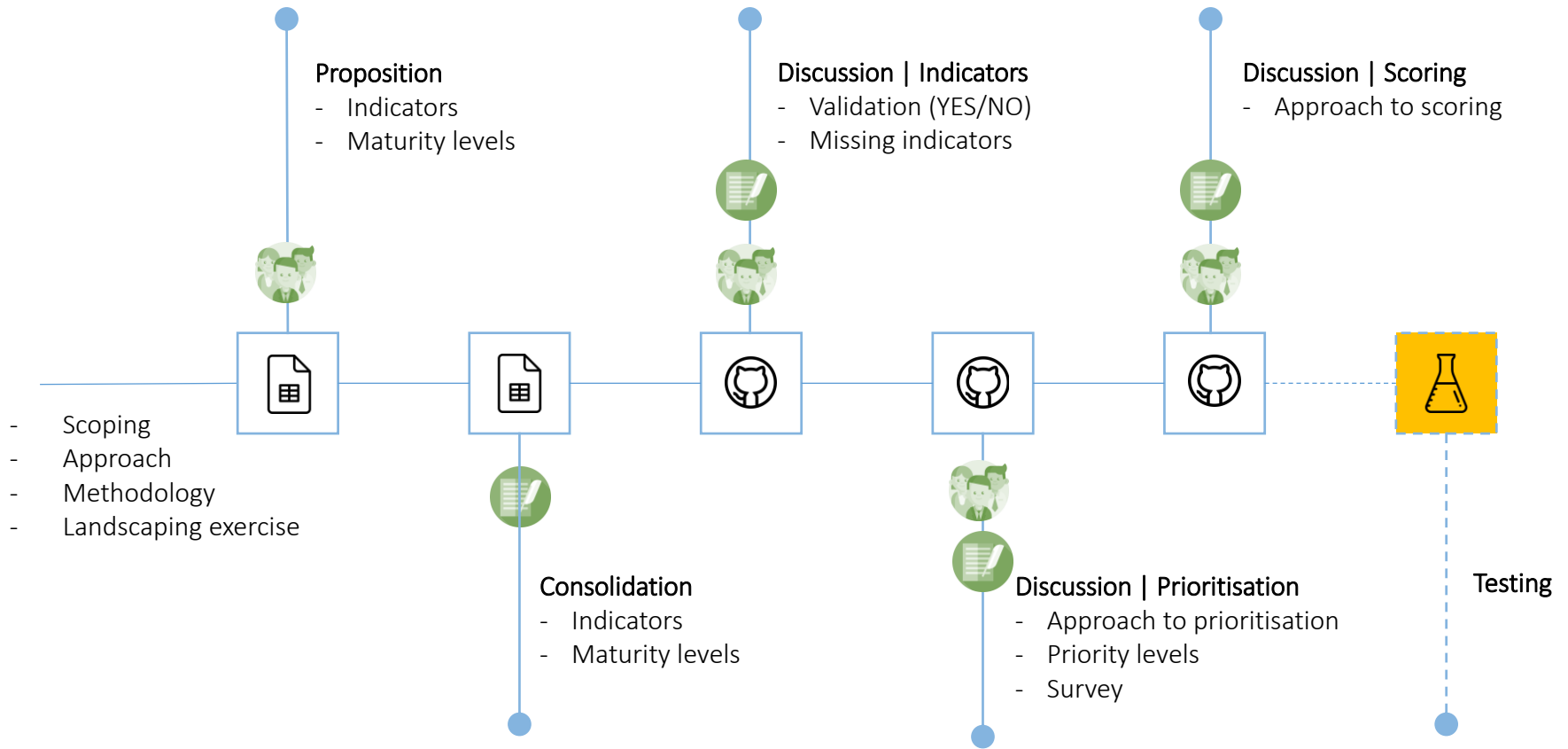
1. Definition	DONE
2. Development	DONE
i) First phase	DONE
ii) Second phase	DONE
3. Testing	ONGOING
4. Delivery	ON HOLD

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



# State of play

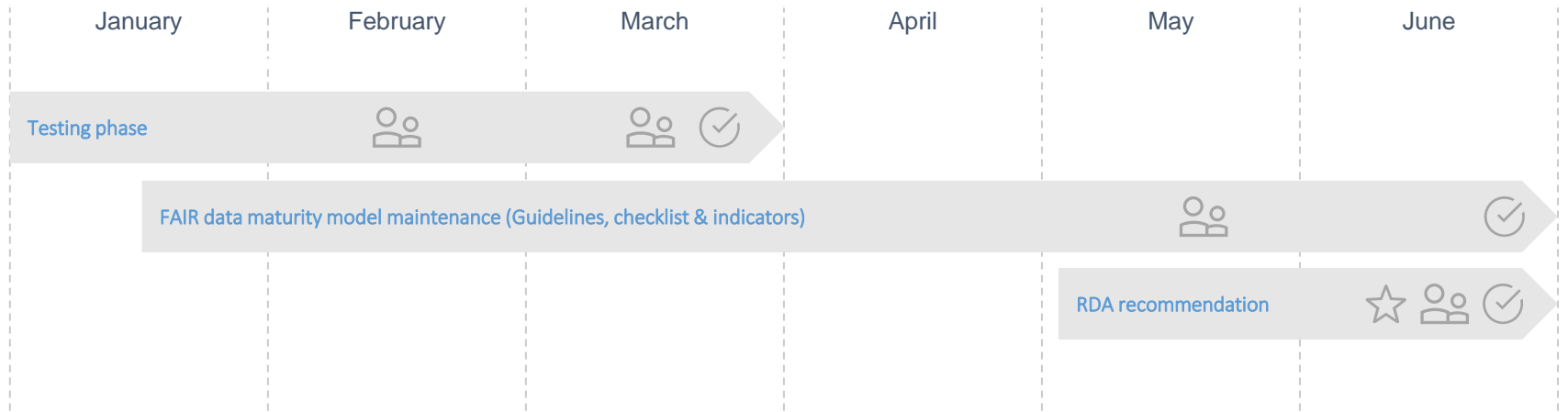
 Editorial team  
 Working group



# 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

2020



☆ End of RDA WG

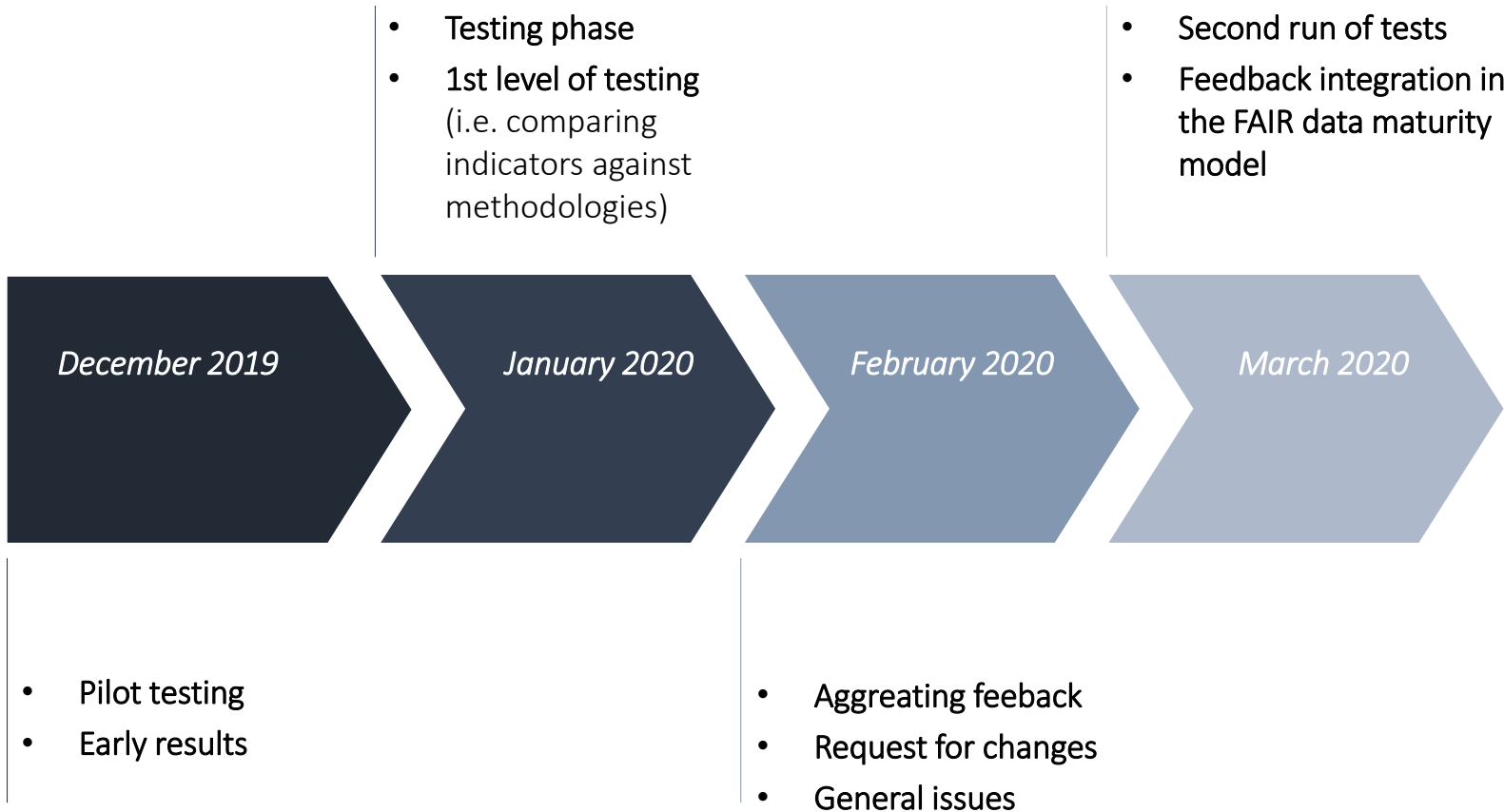
👤 Workshop

✅ Deliverable



Testing phase

# Testing phase | Overview



# Testing phase | Overview

- Thanks to all testers for their contribution
- 13 volunteers having different affiliations
- Various range of disciplines and entities
- Different approaches to the scoring

	Discipline / Domain	Affiliation / Tool	Tester	Entity
1	Earth Science	NCEI of NOAA	Ge Peng	Dataset
2	Engineering & Technical sciences	4TU.ResearchData	Egbert Gramsbergen, Paula Martinez-Lavanchy, Madeleine de Smaele, Marta Teperek	Dataset
3	Humanities, Spatial, Health, etc.	FAIRsFAIR	Anusuriya Devaraju	Methodology
4	Human-Environment Observatories (OHMs)	DRIIHM infrastructure	Romain David & Emilie Lerigoleur	Methodology
5	Biology	ODAM information system	Romain David & Daniel Jacob	Dataset
6	Agronomic & Biomedical	Agroportal	Romain David, Clément Jonquet & Emma Amdouni	Ontology
7	<u>ALL</u> disciplines / domains	ARDC FAIR self-assessment tool	Kerry Levett & Nichola Burton	Methodology
8	Humanities & Social Sciences	DRI	Kathryn Cassidy & Natalie Harrower	Dataset
9	Astronomy	CDS	Françoise Genova	Dataset



## Comments on indicators

## General issues

## Specific issues

## Information needs

- There are (too) many indicators. However, others note that this level of granularity is useful, because it helps to think about all the aspects
- Testing the indicators provided suggestions for improving existing evaluation approaches or existing standards
- Some indicators depend on data characteristics (public, embargoed, private), domain practice and type of repository (domain-specific vs. domain-agnostic), 'perimeter scale' (infrastructure vs. dataset)
- Distinguishing indicators for metadata separate from indicators for data does not work for resources with embedded metadata
- Overlap between indicators (e.g. across principles F1/A1 and F2/R1)
- Some indicators are conditional, e.g. the ones on authentication, authorisation, references and consent – if not applicable, they should not 'count'
- Several indicators require compliance with community standards, but who defines them?
- If data is an ontology, a different set of indicators or priorities may be needed



Comments on indicators

**General issues**

Specific issues

Information needs

- FAIR principles are aspirational and ambitious, aiming at full machine-understandability, but current practices are not well aligned at this point in time
- Identification: there are various comments, some favour identification of metadata over identification of data, others data over metadata, others see both as equally essential, but there is also a comment that having separate identifiers is not common practice
- There seems to be a role for landing pages and other human-readable documentation in providing information, in addition to structured metadata
- Requests for adding maturity levels > scoring
- Data comes in different granularities: whole dataset or part of dataset or individual data items (e.g. observations, concepts)
- Different perspectives on metadata and how it relates to data:
  - repository level / collection level / dataset / data item level metadata
  - separate metadata records or embedded metadata





Comments on indicators

General issues

**Specific issues**

Information needs

- There may be a need to ‘profile’ indicators for specific cases, i.e. selecting a subset of indicators, adapting priorities, following discipline-specific guidelines
- Noteworthy that testers are often stricter than the priorities that the WG has defined, e.g. making essential:
  - machine-understandable community standards
  - standard, open-source protocols
  - machine-understandable knowledge representation
  - standard vocabularies
  - standard reuse licences
- One tester proposes to do away with all priorities entirely




Comments on indicators

General issues

Specific issues

Information needs

- Need for better explanation of terms, in particular ones that are vague or subjective, e.g. 'sufficient'
- Need for better definition of terms used in FAIR principles, e.g. *knowledge representation*, *FAIR-compliant vocabularies* – to take into account [https://doi.org/10.1162/dint\\_r\\_00024](https://doi.org/10.1162/dint_r_00024)
- Need for information on best practices that may be applied to increase FAIRness, e.g. identification, protocols, licences
- These issues to be addressed in the Guidelines, using suggestions and examples provide by testers



The MIT Press Journals

Books Journals Digital Resources About Contact

Home | Data Intelligence | List of Issues | Volume 2 , No. 1-2 | FAIR Principles: Interpretations and Implementation Considerations

**DATA Intelligence**

**FAIR Principles: Interpretations and Implementation Considerations**

[Show all authors](#)  
Annika Jacobsen, Ricardo de Miranda Azevedo, Nick Judy, Dominique Batista,

Quarterly  
Founded: 2018  
E-ISSN: 2641-435X

More About Data Intelligence ▾

Journal Resources

Posted Online January 31, 2020  
[https://doi.org/10.1162/dint\\_r\\_00024](https://doi.org/10.1162/dint_r_00024)

© 2019 Chinese Academy of Sciences Published under a Creative Commons Attribution 4.0 International (CC BY 4.0) license

# Testing insights | General discussion

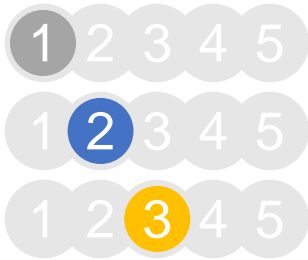
- 1 Even Open Data requires clarity on access conditions and reuse licence
- 2 Requirement for persistent identification of both metadata and data, but this is not always possible
- 3 Metadata standards may not include elements for all FAIR requirements (e.g. provenance, licences); role for documentation and landing pages?
- 4 Knowledge representation is about (1) controlled vocabularies used and (2) data models/formats



# Scoring mechanism

# Scoring mechanisms | Overview

## 5-level scale per indicator



- Five levels of compliance
- Per indicator – aggregated per FAIR area
- Non applicable or consideration/implementation as options
- Useful for giving credit for evolution and helping people to improve

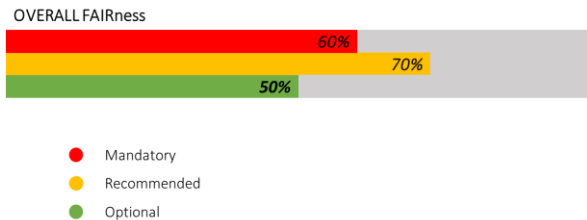
## FAIRNESS per area

	Essential	Important	Useful
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

- Measurement based on priorities
- Per indicator – aggregated per FAIR area
- Score determined based on the compliance to priorities
- Provides a ‘measure of FAIRness’

## Overall FAIRNESS



- Measurement based on priorities
- Per indicator – overall score
- Aggregated score
- Provides a quick view of how priorities are met -- but does not give detailed view



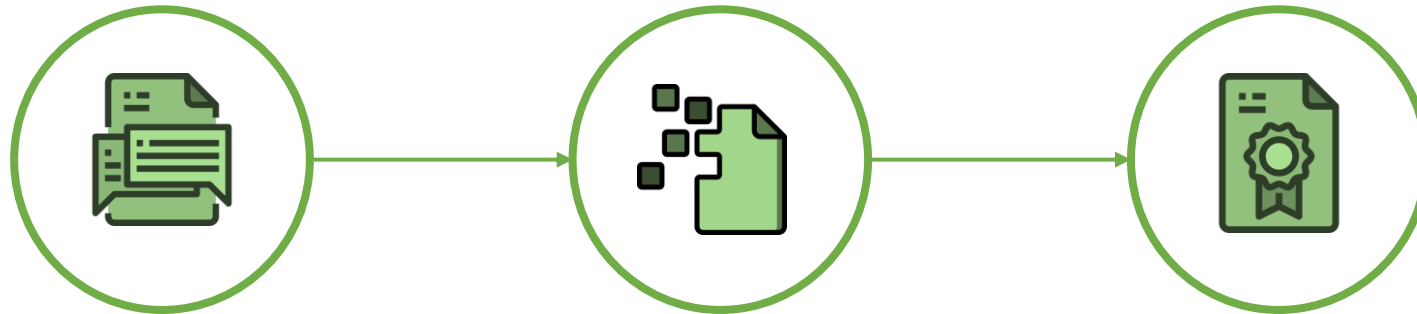


# Action items & Next steps

# Guidelines | first draft

GUIDELINES	INTRODUCTION	<ul style="list-style-type: none"><li>• Introduction</li><li>• Objectives</li><li>• Use of the document</li></ul>
	FRAMEWORK	<ul style="list-style-type: none"><li>• Indicators</li><li>• Maturity levels</li><li>• Prioritization</li><li>• Indicators description</li></ul>
	IMPLEMENTATION	<ul style="list-style-type: none"><li>• How to evaluate</li></ul>

# Guidelines | further development



Working Group to share **remarks** and **suggestions** about the guidelines

**Testing phase** will bring out comments and suggestions for change and for additional guidance

**Stable version of the guidelines** to be published for the next RDA plenary



[https://docs.google.com/document/d/1pDGGL3-BbBJu18KlfZUI3AizKLHXGXdli\\_mPtpEWmeg/](https://docs.google.com/document/d/1pDGGL3-BbBJu18KlfZUI3AizKLHXGXdli_mPtpEWmeg/)



# Action item and next steps

Working Group members are invited to:

- Share feedback, comments & suggestions – on the [Guidelines](#)
- Discuss proposals for changes in priorities on GitHub (issues will be created)
- Contribute to GitHub discussion on [scoring](#)

We're also looking for **volunteers for further testing**; please contact us!

## WORKSHOP #8

15<sup>th</sup> RDA PLENARY IN MELBOURNE  
19 March 2020

11.30 – 13h00 (GTM+11) | Breakout 4

# 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-vIPqjN9LCarOFmc3](https://drive.google.com/open?id=11hyAYCKz_NV0Ob9-vIPqjN9LCarOFmc3)

- RDA FAIR data maturity model WG – **Guidelines**

[https://docs.google.com/document/d/1pDGGL3-BbBJu18KlfZUI3AizKLHXGXdli\\_mPtpEWmeg/](https://docs.google.com/document/d/1pDGGL3-BbBJu18KlfZUI3AizKLHXGXdli_mPtpEWmeg/)

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

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



Thank you!