Workshop report

FAIR data maturity model Working Group

Second face-to-face meeting - 23rd October 2019

Project	RDA FAIR data maturity model working group	Date & Time	23 October 2019 11.30 — 13:00 UTC
Type	Physical meeting	Location	14th RDA Plenary
Meeting Chairs	Edit Herczog	Issue date	14 November 2019

Objectives

The objective of the face-to-face workshop was twofold. First, the editorial team reported on the status of the different work packages (e.g. indicators, prioritisation, scoring, etc.) and presented the next steps. Secondly, two topics were selected by the audience to be further explored.

Agenda

1. Welcome, objectives of the meeting
2. Roundtable
3. State of play
4. Development – First phase
5. Development – Second phase
6. Panel discussion | Consensus
7. Development – Next steps
8. Next steps and Q&A

Useful links

- RDA FAIR data maturity model WG
- RDA FAIR data maturity model WG – Case Statement
- RDA FAIR data maturity model WG – GitHub
- RDA FAIR data maturity model WG – Collaborative document
- RDA FAIR data maturity model WG – Indicators prioritisation
- RDA FAIR data maturity model WG – Indicators prioritisation survey results
- RDA FAIR data maturity model WG – Mailing list
- RDA FAIR data maturity model WG – Workshop #5 material

Participants

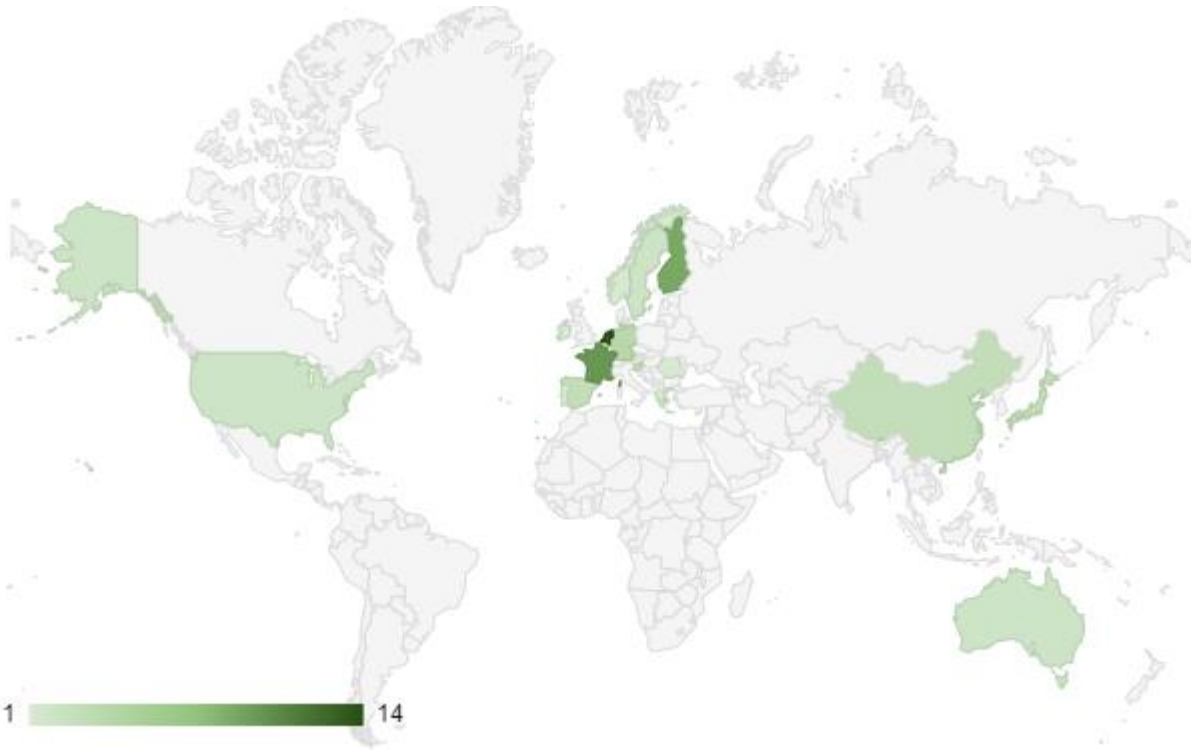
The workshop was well attended. Here below is a non-exhaustive list of the participants.

Name		***Affiliation***
Adam Dinsmore	UK	Wellcome Trust
Adam Leadbetter	IE	Marine Institute
Adeline Joffres	FR	CNRS France
Adrian Dusa	RO	RODA
Ana Slavec	SI	InnoRenew CoE
Anne-Caroline Delétoille	FR	Institut Pasteur
Barbara Magagna	AT	Umweltbundesamt
Barbara Sierman	NL	KB National Library of the Netherlands
Ben Schaap	NL	Godan
Birger Jerlehag	SE	University of Gothenburg
Carlos Casorrán	BE	EC DG RTD
Christophe Bahim	BE	PwC, Editor team
Christophe Bruch	DE	Helmholtz Association
Cynthia Love	AU	CSIRO
Damien Boulanger	FR	CNRS France
Daniel Mallman	DE	Forschungszentrum Jülich
Dimitri Szabo	FR	INRA
Donna Mc Rostie	AU	University of Melbourne
Edit Herczog	BE	Chair, Vision & values SPRL
Elli Papadopoulou	GR	ATHENA Research & Innovation Center
Erik Schultes	NL	GO FAIR
Eva Méndez	ES	Universidad Carlos III of Madrid
Fernando Aguilar	ES	CSIC
Frederic Andres	JP	NII
Gergely Sipos	NL	EGI

Heidi Laine	FI	CSC - IT Center for Science
Helen Glaves	UK	RDA TAB / UKKRI
Hervé L'Hours	UK	UK Data Archive
Hylke Koers	NL	SURFsara
Ilaria Carnevale	NL	Elsevier
Ingrid Dillo	NL	DANS / H2020 FAIRsFAIR
Iris Alfredsson	SE	Swedish National Data Service
Ivana Ilijašić Veršić	NO	CESSDA ERIC
Izabela Witkowska	NL	Utrecht University
Janez Štebe	SI	University of Ljubljana, Social science data archives
Jessica Parland-von Essen	FI	CSC - IT Center for Science
Jingchao Tan	CN	CAAS
Joanne Yeomans	NL	Leiden University
Kana Asano	JP	Japan Science and Technology Agency (JST)
Katja Fält	FI	Tampere University
Konstantinos Repanas	BE	European Commission DG RTD
Kristiina Himanen	FI	University of Helsinki
Leah Riungu-Kalliosaari	FI	CSC - IT Center for Science
Leyla Garcia	DE	ZBMED
Limor Peer	US	Yale University
Louise Darroch	UK	British Oceanographic Data Center
Makx Dekkers	ES	Independent Consultant, Editor team
Mari Elisa Kuusniemi	FI	University of Helsinki
Maria Teperek	NL	TU Delft
Mario J. Gaspar da Silva	PT	INESC-ID
Mark Greenslade	FR	IPSL
Matt Cannon	UK	Taylor & Francis Group
Michel Schouppe	BE	EC DG RTD
Milan Ojsteršek	SI	University of Maribor (FERI)
Miriam Braskova	NL	Erasmus University
Mohamed Yahia	FR	Inist-CNRS / Datacite
Mustapha Mokrane	NL	DANS

Natalie Harrower	IE	DRI
Nina Järviö	FI	The Federation of Finnish Learned Societies
Patricia Herterich	UK	Digital Curation Center
Pekka Orponen	FI	Aalto University
Philippe Boulben	FR	INIST-CNRS
Romain David	FR	INRA
Rousi Antti	FI	Aalto University
Sahar Farajnia	NL	Elsevier
Simon Lambert	UK	STFC
Stéphane Debard	FR	Institut Research for Development
Sven Rank	DE	Forschungszentrum Jülich
Wo Chang	US	NIST
Xuefu Zhang	CN	CAAS
Yann Le Franc	FR	e-science Data Factory
Yasushi Ogasaka	JP	Japan Science and Technology Agency (JST)
Yuantao Kou	CN	CAAS
Yuri Demchenko	NL	UVA

Here below is a map representing the provenance of the different participants

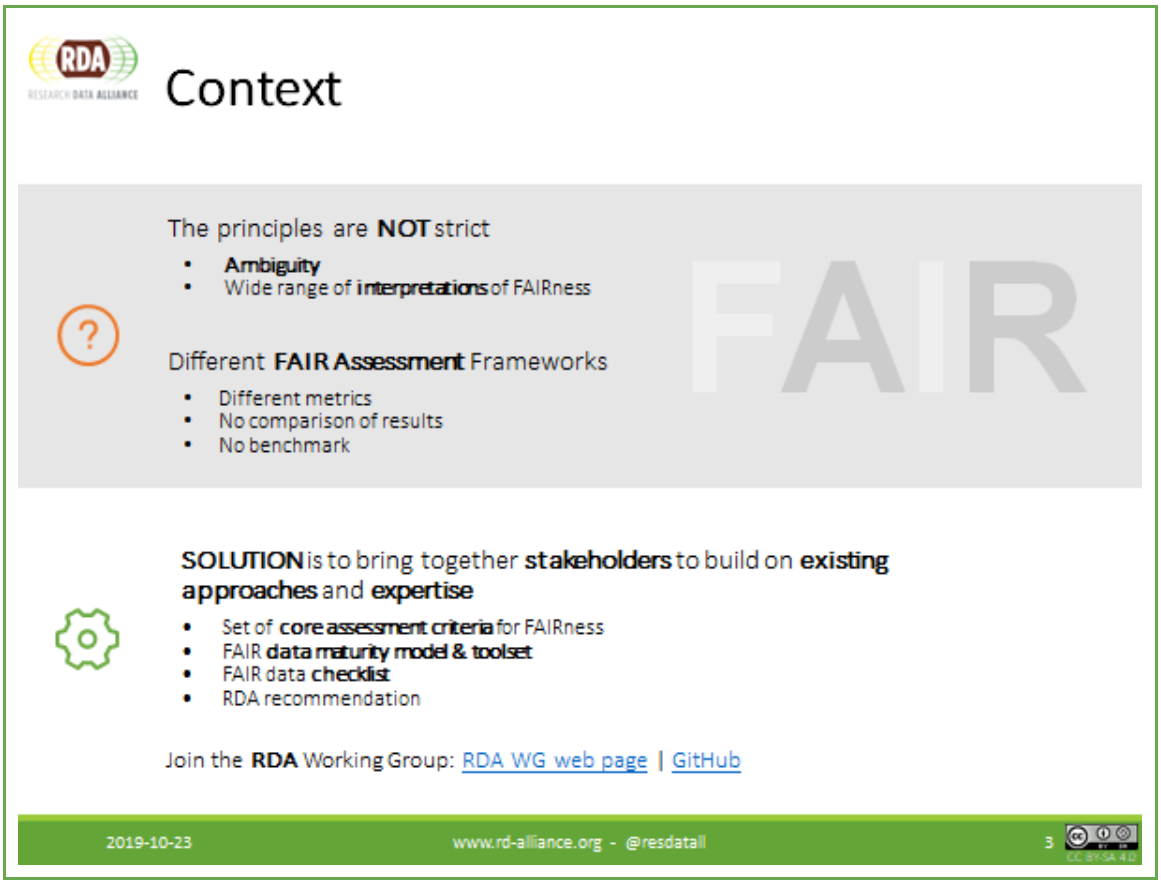

1
14

Content[1]

The workshop was designed in order to be as interactive as possible: to do so, the attendees were given the opportunity to ask questions through the slido app. In addition, a number of potential discussion items were put forward. The Working Group voted for their preferred discussion items, which were tackled during the panel discussion. Consequently, the meeting was fruitful and enabled lively discussions.

1. The Chairs opened the workshop, welcomed the participants and addressed the agenda. The approach to the Working Group was again presented:

    - Challenges rising from the different interpretations of FAIRness
    - Bringing together the relevant stakeholders to discuss and build on existing expertise and different approaches
    - Intended results: i) set of core assessment criteria for FAIRness ii) FAIR data maturity model & toolset iii) RDA recommendation and iv) FAIR data checklist.



[1] Please note that some of the slides are displayed for information purposes. The full presentation can be accessed via the RDA FAIR data maturity model WG web page.

Slide 3 | Welcome and objectives of the meeting

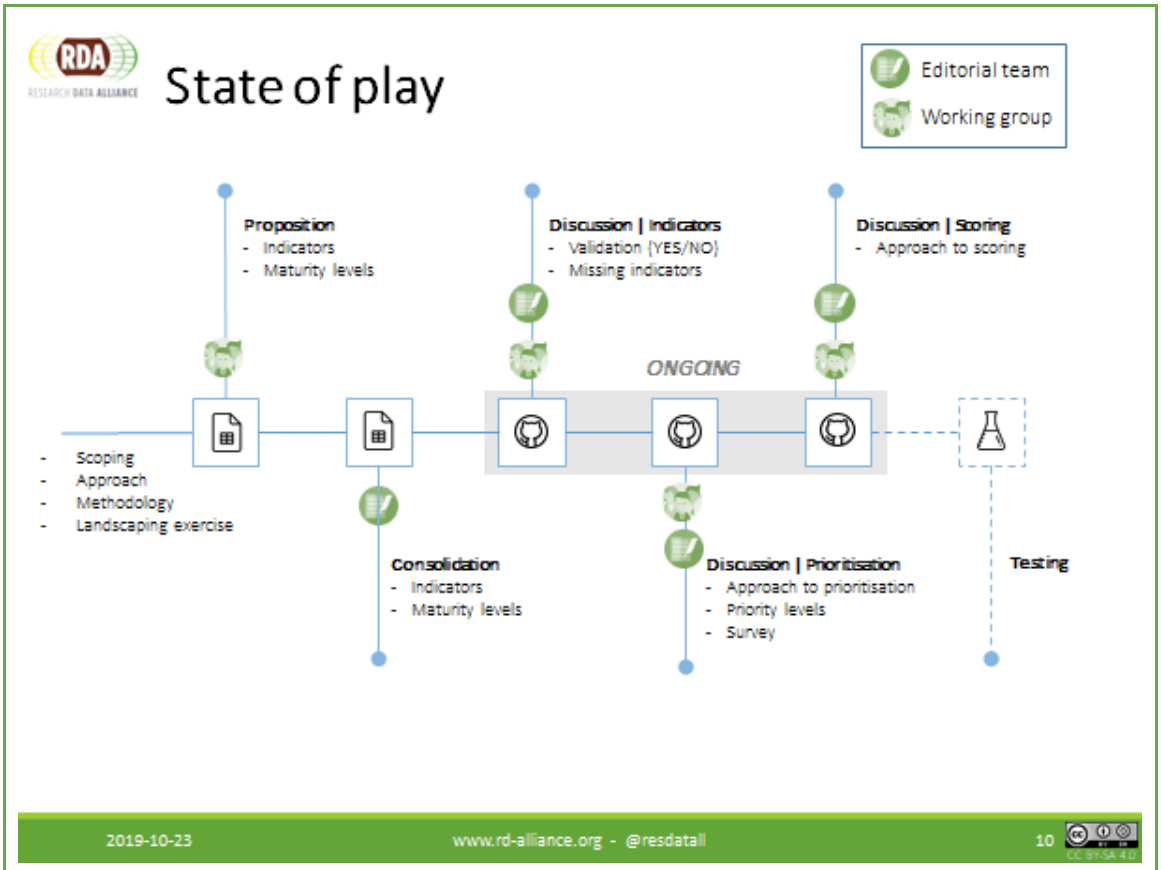
As usual, the Chairs insisted that despite all the challenges arising when designing indicators, the purpose of the WG was **NOT** to re-design the FAIR principles. As there are currently different interpretations of what the FAIR principles entail, the primordial goal is to build a common understanding.

In addition, the chairs reminded that all the presentations and reports are on the RDA FAIR data maturity model WG web page and the members are encouraged to participate via the dedicated GitHub repository.

2. The Chairs and the editor team introduced themselves, following, the participants were kindly invited to say where they come from and what are their roles in their organisation via the chat window.

   The Chairs took the opportunity to introduce Shelley Stall, Senior Director for the American Geophysical Union's Data Leadership Program. She will serve as a third co-chair for the RDA FAIR data maturity model Working Group.

3. The editorial team reported on the current state of development: what steps have been taken and what steps remain to be taken.



Slide 10 | State of play

As illustrated by the slide above, the Working Group was first invited to propose potential indicators to measure the FAIRness of a digital resource. The editorial team then consolidated all the contributions, which resulted in a finite set of 53 indicators and their respective maturity levels.

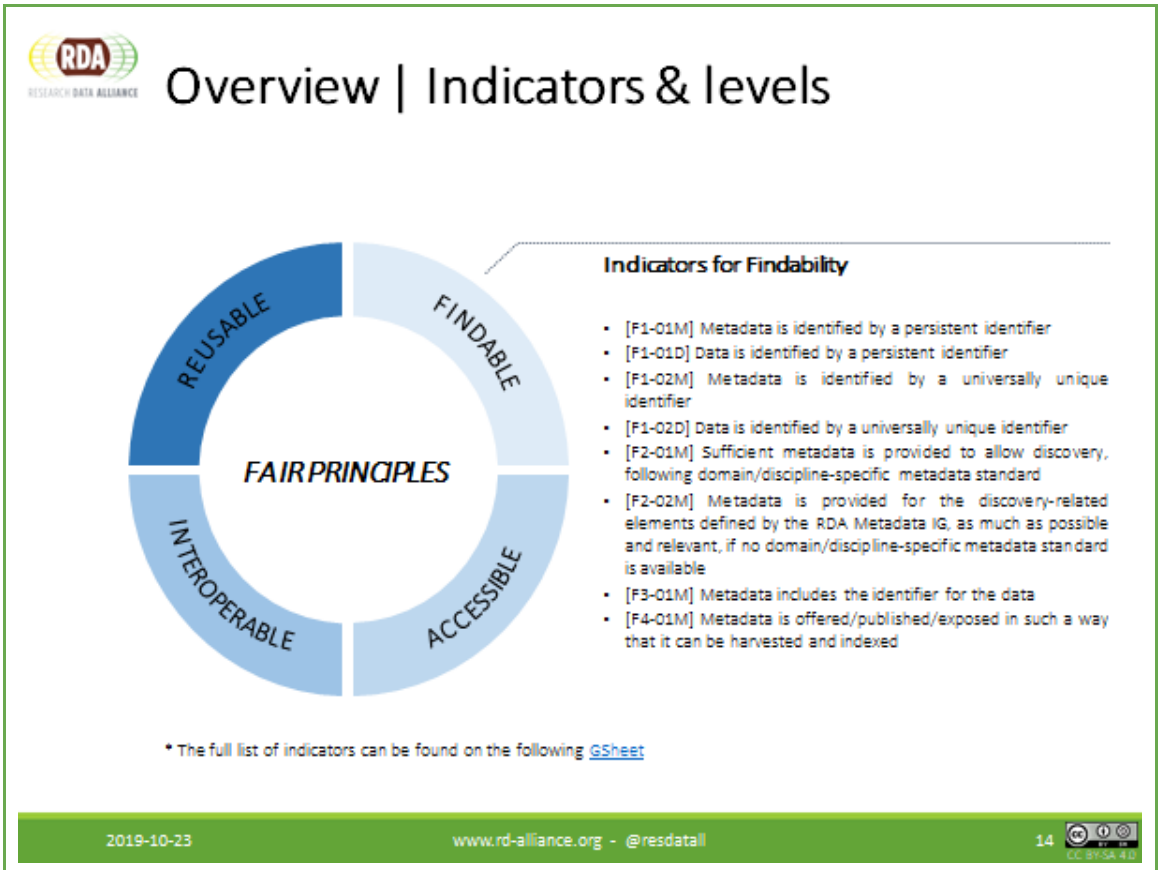
That consolidated set was shared for comments on the dedicated GitHub. Additionally, the editorial team made proposals for prioritisation and scoring. Discussions related to these three topics (i.e. indicators, prioritisation and scoring) were happening in parallel on the GitHub.

In order to facilitate the consensus process about prioritisation, the editorial team put together a survey. As a consequence, having reached a consensus on prioritisation, the discussion is being phased out.

As of today, the editorial team further investigates ways to score the FAIRness of a digital resource and started to look into an approach for testing. Additional information will be shared in the coming weeks about testing and it will be the focus of workshop #6, which is scheduled for the 4th of December.

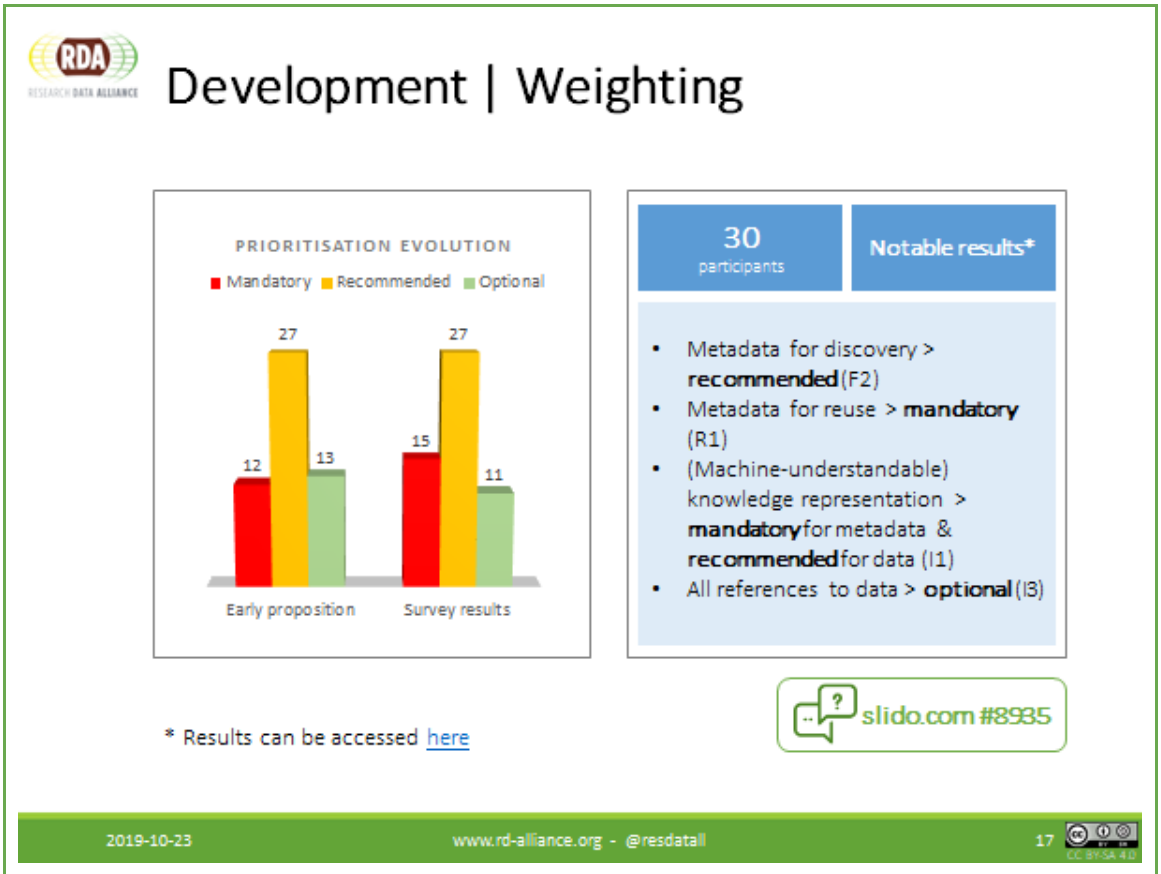
Slide 11 | State of play

Furthermore, it was reminded to the audience that the current state of the indicators, as of early October 2019, is now frozen, with the exception of the indicators for the principles that are concerned with 'richness' of metadata (F2 and R1). The current indicators will be used in a testing phase where owners of evaluation approaches are invited to compare their approaches (questionnaires, tools) against the indicators. As such, the current set of indicators can be seen as an 'alpha version'. In the first half of 2020, the indicators may be revised and improved, based on the results of the testing.



Slide 14 | Indicators & levels

The editorial team gave a glimpse of the indicators: metrics derived from the principles to measure the FAIRness of any digital resource. The editorial team also reminded the audience that, in scope of the charter, this Working Group needs to look specifically at *what* to measure and not *how*. The *how* part will come at a later stage.



Slide 17 | Weighting

The results of the survey related to prioritisation of the indicators were introduced. Overall, the Working Group members tend to be more strict towards evaluating FAIRness. In other words, 15 indicators were voted on mandatory. The editorial team put to the foreground some notable results:

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

Results of the survey can be consulted at the following address.

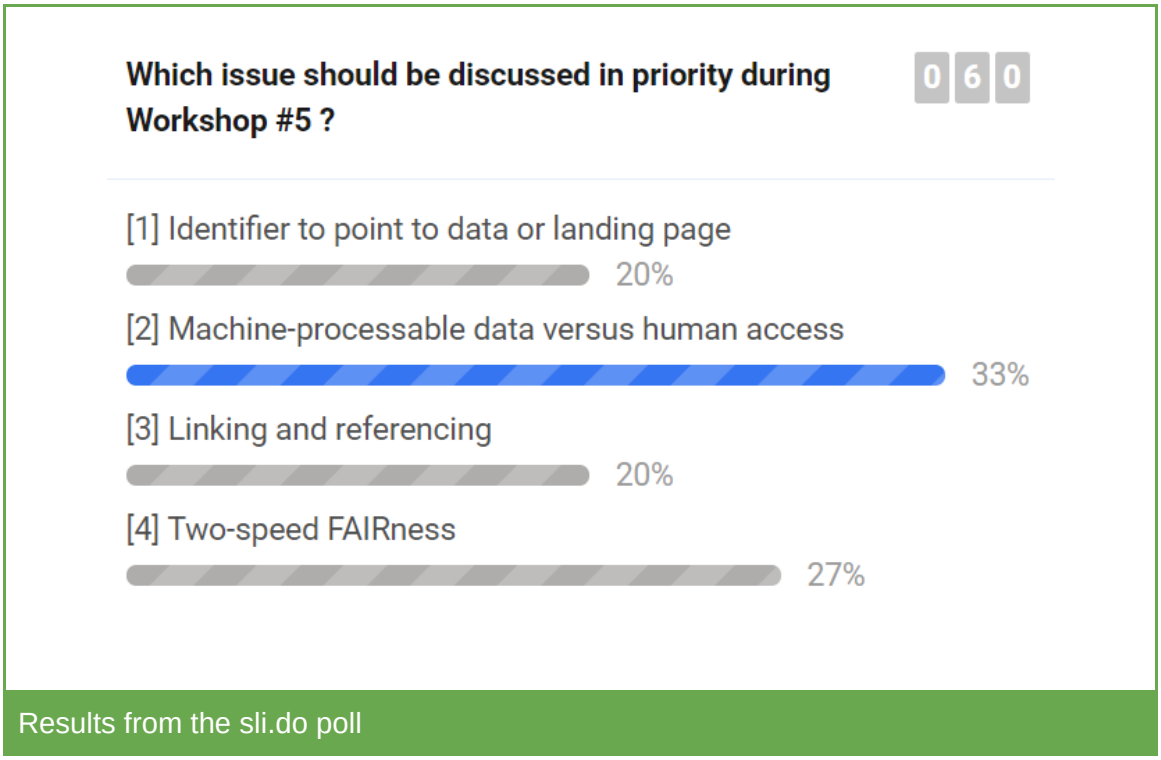
4. In order for the session to be interactive, the editorial team proposed a number of potential discussion items. On slide 20 in the presentation, we put forward four discussion items, two of which were tackled during the session.

- Identifier to point to digital object directly, or indirectly through a landing page
- Machine-processable data versus human access to the digital object

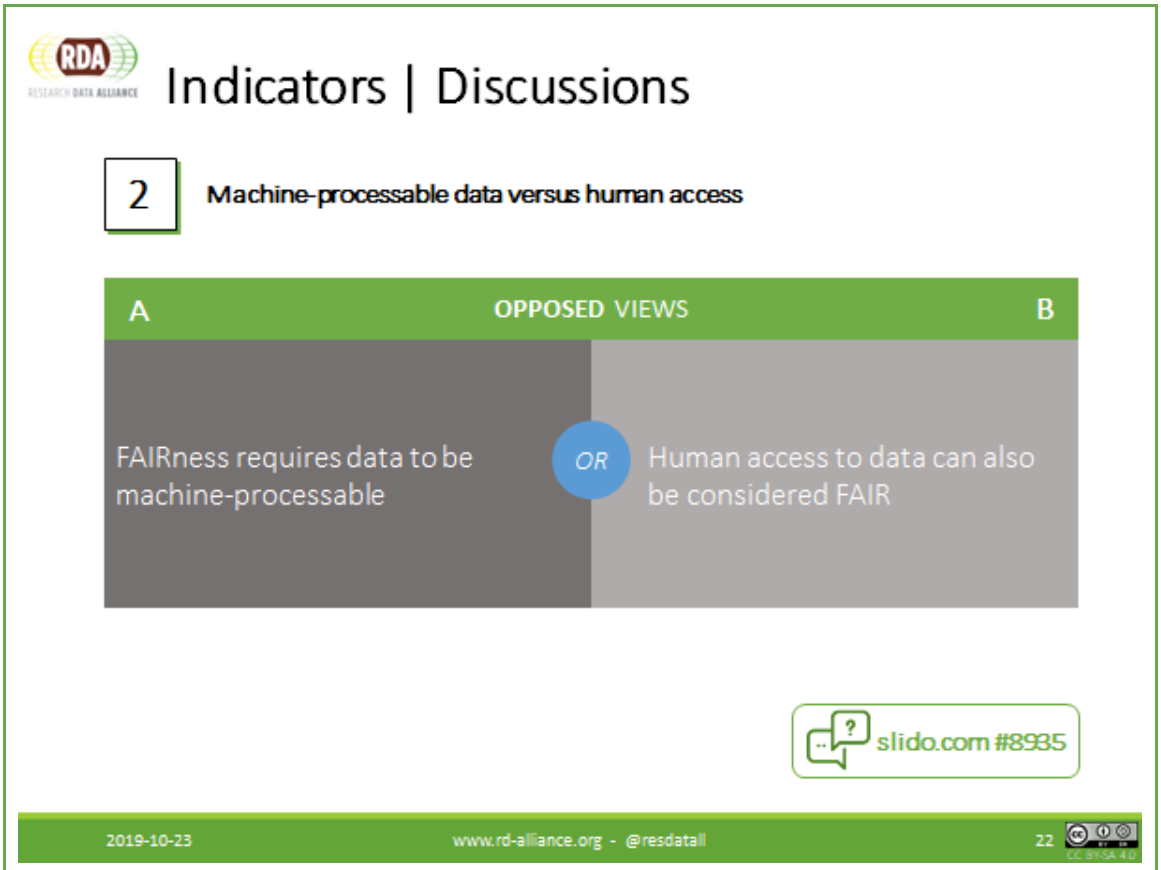
- Linking and referencing
- Two-speed FAIRness

On the grounds of the survey results, the following two issues were addressed:

- **Machine-processable data versus human access to digital object**
- **Two-speed FAIRness**



Results from the sli.do poll



Slide 22 | Indicators | Discussions

Here below are a number of comments put forward by the audience:

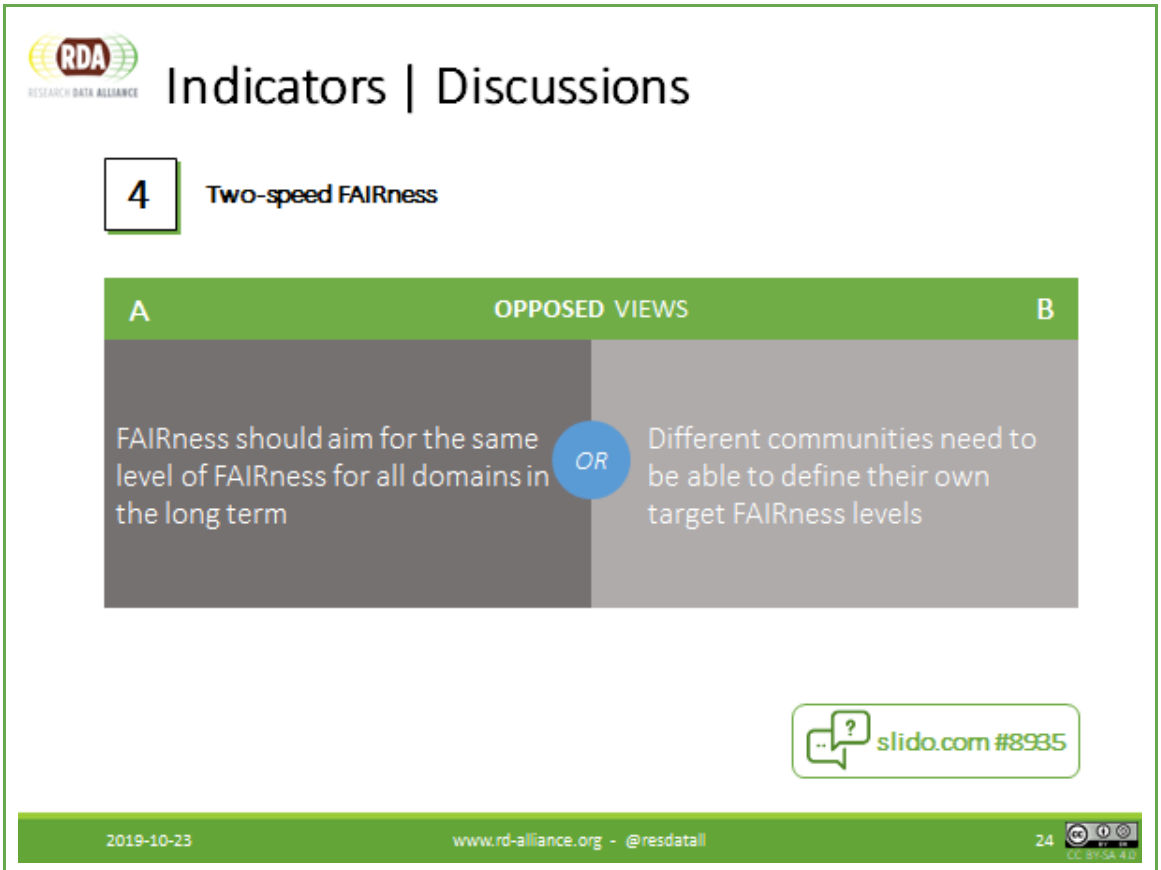
- It is important to note that some communities / people are already quite advanced in terms of FAIRness. For instance, there is a lot of data in the humanities that are not machine-processable and thus needs to be accessed by humans. Account must be taken that in some disciplines data are not machine-processable. If machine-processability becomes a criterion to satisfy, some communities will be left behind. FAIR is an inclusive movement; if a part of the community is left out, FAIR will lose relevance.
- In relation with the above point, machine-readable data is intended for data composed of numbers but when it comes to interpretation of more abstract information, machine-readability becomes a delicate matter. The same comment goes for PDF processing. Lot of historical data comes in PDFs - and these PDFs need to be findable and reusable. As such, having all the data in a machine-processable format is not easy.
- Machine-processable data is a broad concept and needs to be broken down into three distinct concepts: i) machine-readability, ii) machine-understandability and

iii) machine-interpretability. As for the moment, data can be read but not understood or interpreted.

- A & B should not be opposed views but rather complementary views. People should move from A to B. In the final set of indicators, there are currently indicators addressing the two opposed views. The testing phase will reveal if it makes sense or not.

More generic comments made by the audience:

- People are concerned that FAIRness will be a black and white decision, i.e. FAIR or not FAIR. Whereas FAIR is a journey and something that communities should strive for, i.e. FAIR is a means to an end, namely reuse of data. Instead, evaluators should measure the degree of FAIRness. That degree of FAIRness would help to pinpoint improvement areas. In addition, communities would need to decide what degree of FAIRness is accepted.
- The Working Group should define a 'FAIR' horizon for the communities.
- The Working Group should be careful when defining criteria. Communities need to be consulted for such an exercise. Communities should define what is FAIR enough. Everybody is working in silos at the moment.
- As a next step, we should drill down in the prioritisation of indicators and collect communities' opinions, especially on mandatory indicators.



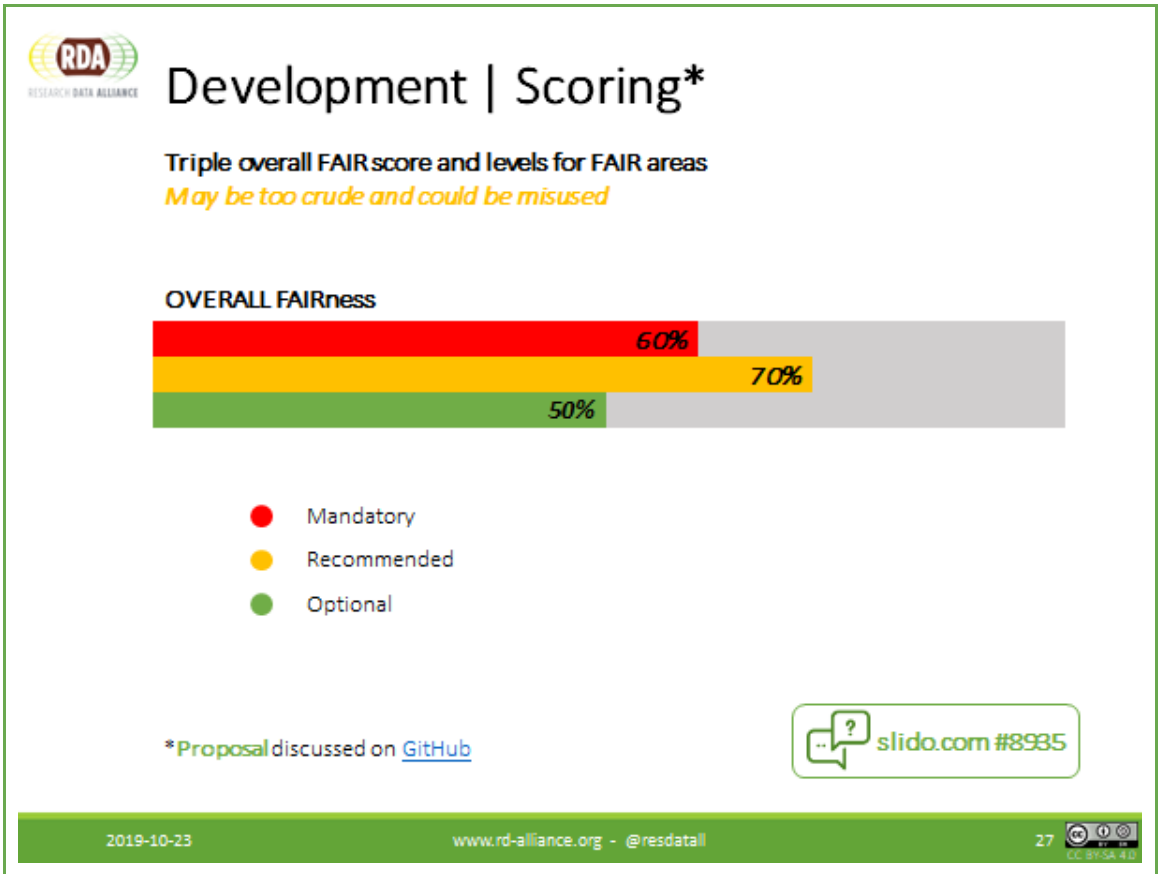
Slide 24 | Indicators | Discussions

Here below are a number of comments put forward by the audience:

- If communities are able to set a low level of FAIRness, it will benefit a cross-domain usage.
- Ideally, everybody should align. But doing so on a 'low level' will create issues, and most importantly the level of FAIRness will be poor.
- This would be unfair to ask a homogenous level of FAIRness because there are two different aspects to consider: i) intra-community and ii) inter-community.
- In the case one lowers the requirements for FAIRness, some communities that were performing well in terms of FAIRness will be demotivated. Indeed, it would mean that you could lower your level of FAIRness and still be alright. The bar should be raised to drive communities (i.e. encourage them to achieve more).

More generic comments made by the audience:

- Currently, researchers are obliged to make their datasets reusable, but at some point, researchers will be gone. To overcome problems with ongoing availability, there should be an infrastructure that supports FAIRness on the long term.
- In relation with the above point, ‘long term’ is too vague. The community needs to define a timeframe.
- It is necessary to have a score per FAIR area (F, A, I and R)



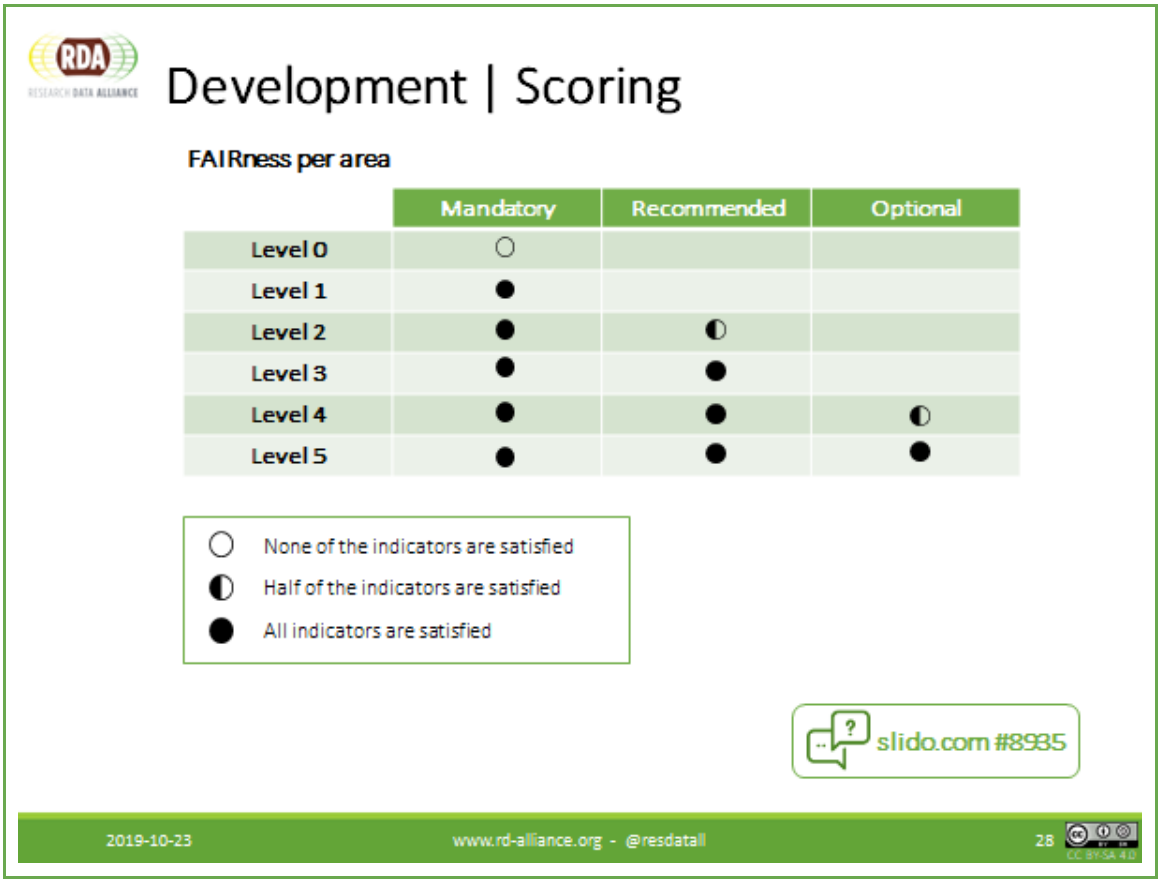
Slide 27 | Development scoring

Initially, a proposal was made about scoring the FAIRness of a digital resource using five levels. Through discussions to determine the best way possible to tackle the scoring of one digital resource's FAIRness, a proposal was made to have an overall FAIR score and levels for the FAIR areas. As illustrated in the picture above, one would a triple score for the overall FAIRness for a digital resource evaluated. Its means that, for instance, the resource satisfied 60% of the mandatory indicators, 70% of the recommended ones and only 50% of the optional ones.

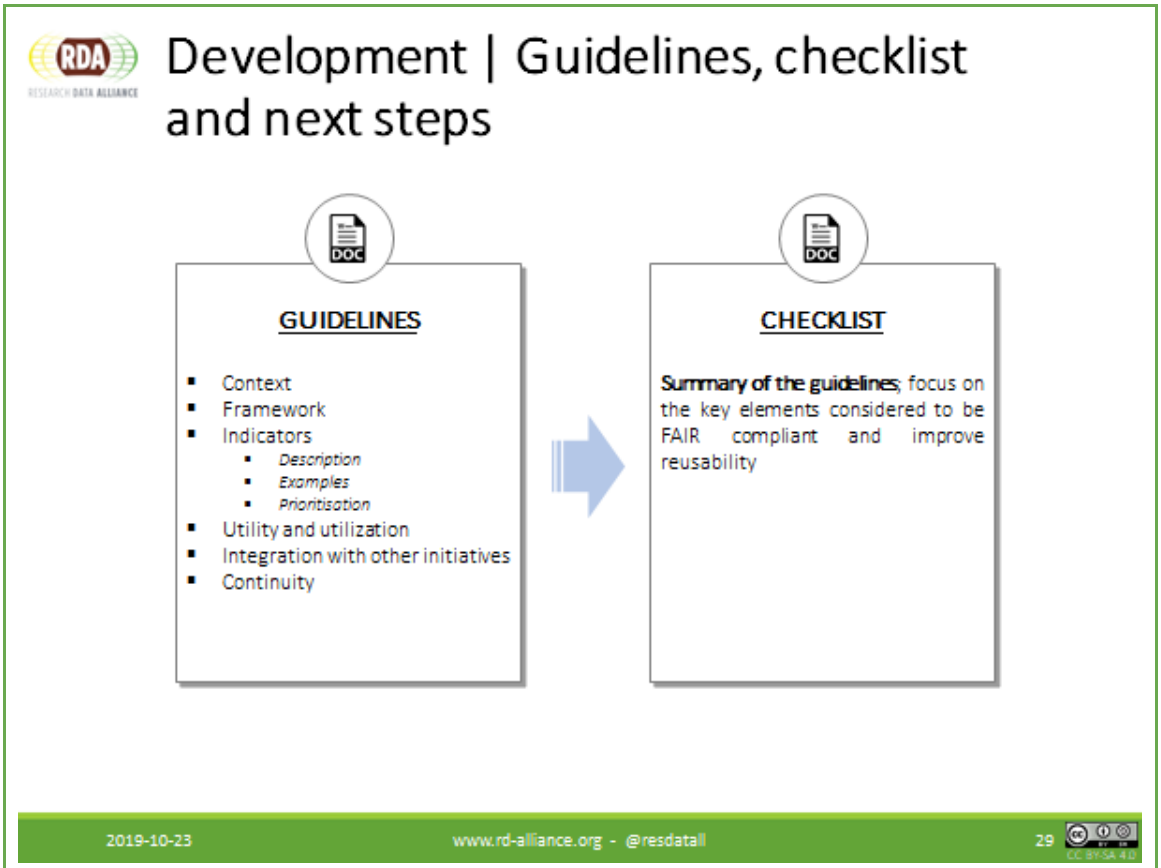
As for the levels, the approach is somewhat different. It consists of a tiered approach. As illustrated below, the FAIRness of a FAIR area is determined by levels.

- Level 0 – The resource did not comply will 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



Slide 28 | Development | Scoring



Slide 29 | Development | Guidelines, checklist and next steps

The editorial team then introduced the proposed approach for the development of the guidelines and checklist. The guidelines will have a purpose to help the evaluator with their approaches whereas the checklist will be a condensed summary of the key elements to be taken into account in evaluating the FAIRness of a digital resource.

Follow-up Action Plan

The working group was encouraged 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*