FAIR Data Maturity Model

presented by Edit Herczog Co-chair

e- IRG Workshop Geneva

20th of May 2019

Agenda

- Who we are
- Aim of the WG
- Methodology Timeline and Scope
  - Definition
  - Development
  - Testing
  - Delivery
- Actions and Next steps

Important:

- The Working Group started its work, but not issued yet results.
- This presentation is to explain the workplan and invite you to be part of the committed team

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
  - Vasilios Peristeras from Europe
- TAB member:
  - Jane Wyngaard from South Africa
- Secretariat: Lynn Yarmey 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)

Minimum CORE criteria

- **WHAT**
- **NOT HOW**

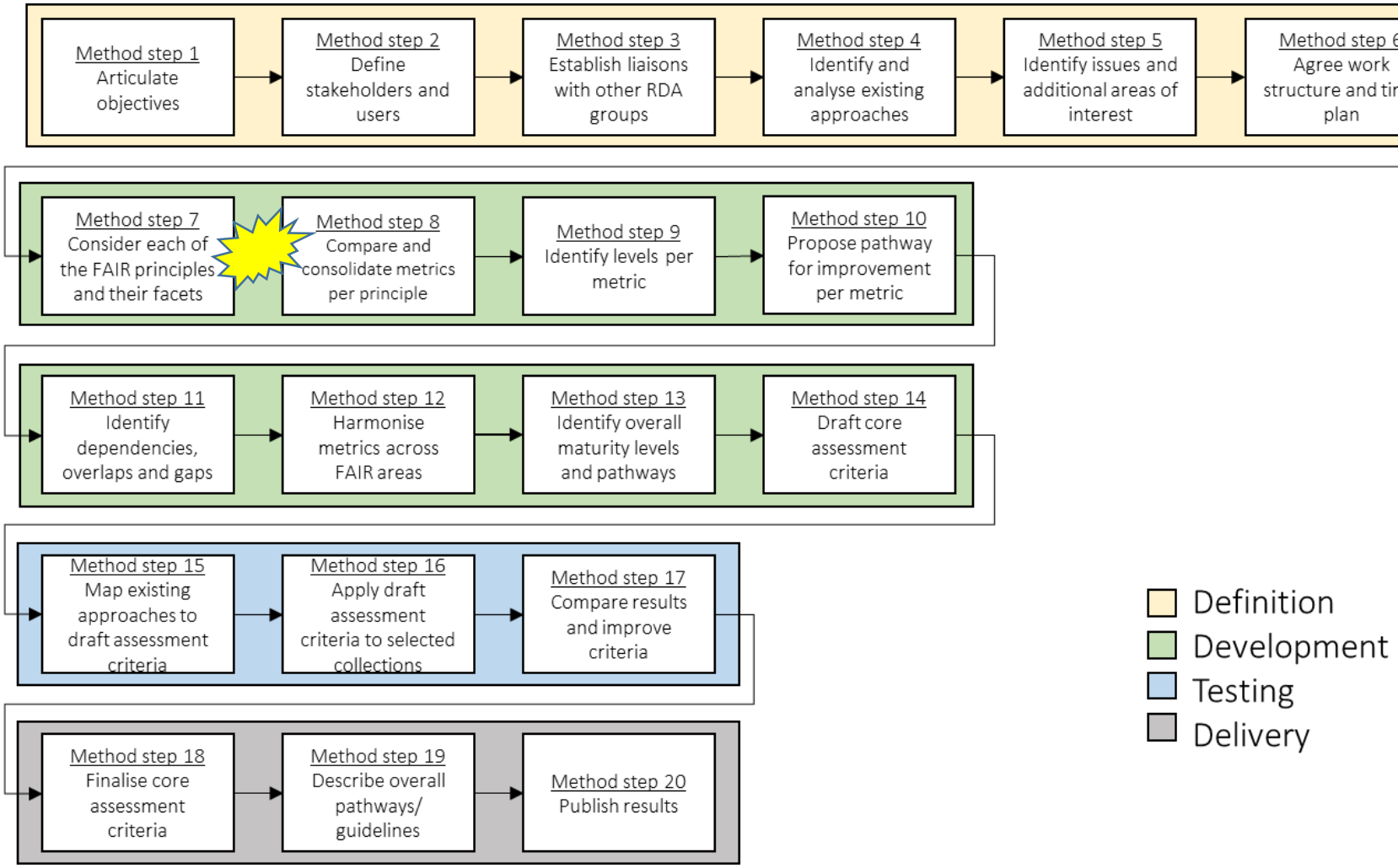
WG methodology, timeline & scope

Proposed development methodology

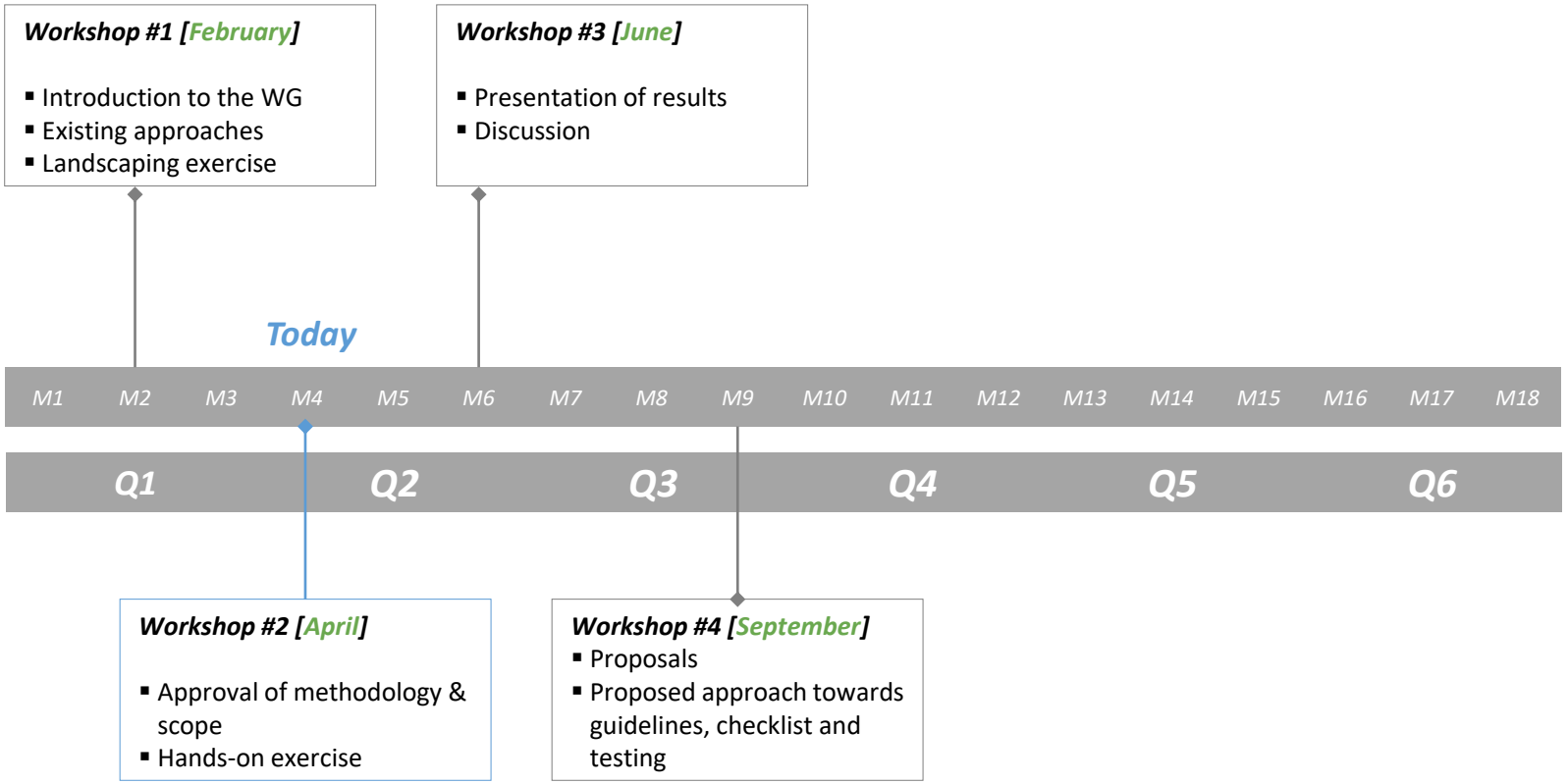
Bottom-up approach comprising 4 phases

- Definition
- Development
  - Assessment of the four FAIR principles in four ‘strands’
  - Fifth ‘strand’: beyond the FAIR principles
- Testing
- Delivery

Overview of the methodology



Proposed timeline



Survey results

- Respondents
  - Big Data Readiness
  - FAIR Metrics
  - FAIR evaluator
  - Data Stewardship Wizard
  - FAIR data assessment tool
  - FAIR enough? Checklist to evaluate FAIRness for researchers
  - Checklist for evaluation of Dataset Fitness for Use
  - Support your Data
  - Fairness assessment tools for crediting/rewarding research data sharing activities
- Some discussion items derived from the survey
  - Scope of the assessment
    - What does the tool assess? [e.g. DMP, dataset, way of conducting research, anything]
    - Cross-domain or domain-specific?
  - Audience [e.g. researcher, repository manager, data librarian, data steward]
  - Automation of the assessment [i.e. what proportion to automate and how]
  - Certification [e.g. quality label, scoring system]
  - Maintenance and governance [e.g. GitHub]
  - Guidance [e.g. checklist]

Summary of open issues

- Scope of the assessment
  - Data versus metadata, DMP, data sharing activities
  - General versus domain-specific
- Standards maturity
- Responsibilities
  - Criteria definition
  - Measurement execution
- FAIRness literacy
- Manual vs automated
- Scoring / Levels
- Certification

Results of preliminary analysis - 1

- Landscaping exercise as a ***starting point***
- Analysis of existing approaches
  - Publicly available documentation and the survey
  - Clustering questions and options
    - FAIR facets [e.g. F1, A2] per principle
    - Beyond the FAIR principles [e.g. data storage]
  - Identification of potential overlaps
- WG to compare questions and derive common aspects

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 - 3

Early observations

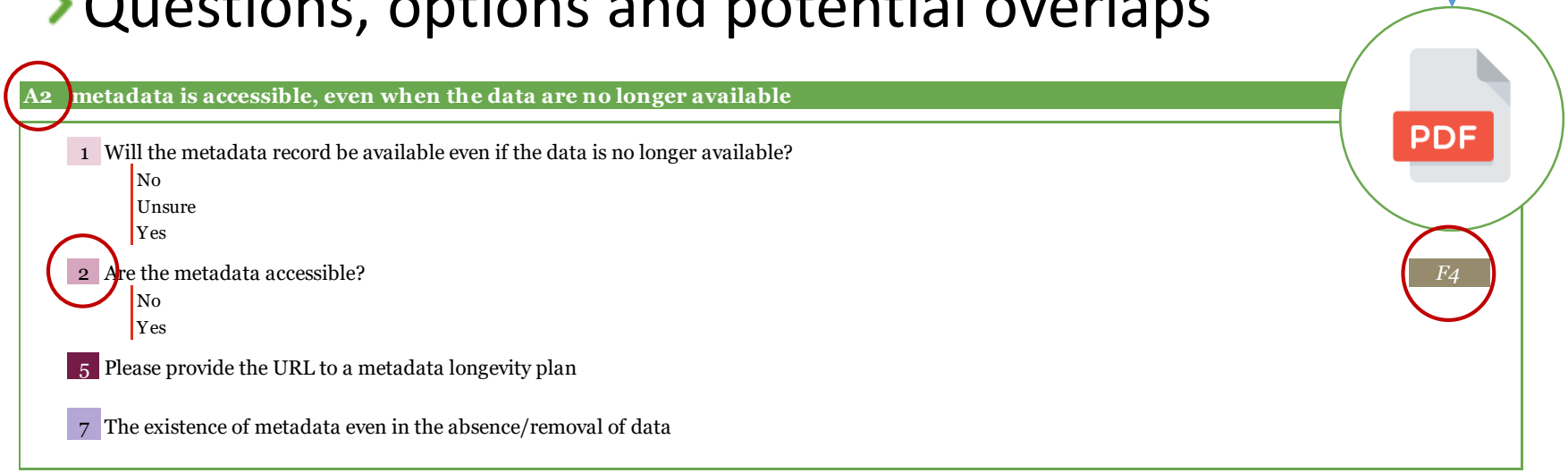
123 questions	**5** types of option	**4** scoring approaches

- On average, six questions per facet
  - Overlaps and different terminologies used
  - Some facets are underused [e.g. A1, A1.1, A1.2, A2]
  - Some facets are overused [e.g. F1, F2]
- Different options
  - YES/NO
  - TRUE/FALSE
  - URL
  - Multiple choice
  - Free text
- Different scoring mechanisms
  - Stars
  - Grade
  - Loading bar
  - None

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]
- Questions, options and potential overlaps

Example



Results of preliminary analysis - 5

- Beyond the FAIR principles
  - Characteristics of projects, workflows and tools
  - Open vs. closed/embargoed data
  - Curation, maintenance and governance
  - Certification (what and who/how)
  - Others ?
- Should the WG consider these additional aspects as one or more separate strands?

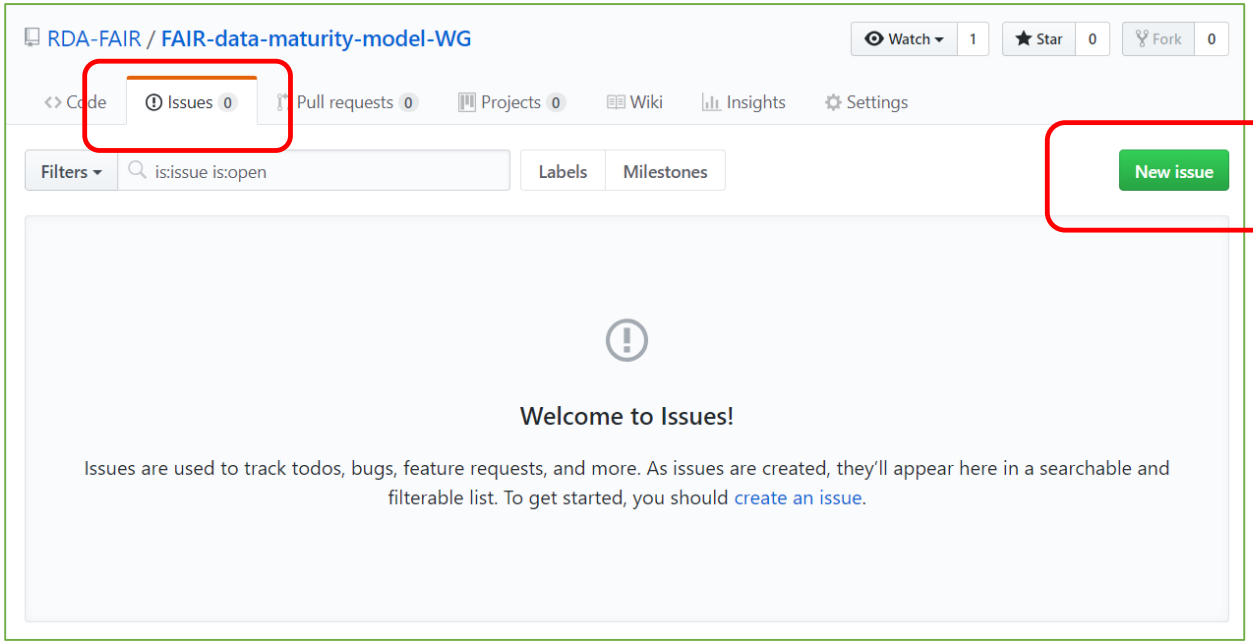
How to contribute - 1

- Contribution is sought and welcomed for

METHODOLOGY	ANALYSIS	AOB
E.G. - Missing items - Alternative approach - …	**E.G.** - Scope - Irrelevant items - Missing items - Additional aspects - …	…

How to contribute - 2

- Issue tracking on GitHub (Join GitHub)



- Create an issue:
  - Provide a clear title and a detailed description
  - Label and categorize the issue [e.g. Methodology Principle_F]

Proposed scope

	Proposed resolutions
ENTITY	**Dataset** and **data-related aspects** (e.g. algorithms, tools and workflows)
NATURE	**Generic** assessment (i.e. cross-disciplines)
FORMAT	**Manual** assessment
TIME	**Periodically** throughout the lifecycle of the data
RESPONDENT	People with **data literacy** (e.g. researchers, data librarians, data stewards)
AUDIENCE	Researchers, data stewards, data professionals, data service owners, organisations involved in research data and policy makers

Overview of discussions on GitHub

Findable: What does it mean? [GitHub]

- Human Findable
- Machine Findable
- Meaning of 'rich metadata'

'Flows' beyond the FAIR assessment [GitHub]

- Data flow
- Data flow legal issues
- People flow
- Financial flow
- Hardware infrastructure

Actions items & next steps

Discussion

- **Nature** of RDA recommendations & outputs
- How to keep you involved?

Action items

- Call for volunteers
- Development of the core assessment criteria on GitHub
  - Analysis of all the FAIR principles — Method step 7
    - FAIR – Findable [Link]
    - FAIR – Accessible [Link]
    - FAIR – Interoperable [Link]
    - FAIR – Reusable [Link]
  - Comparison and consolidation of the metrics per principle — Method step 8
  - Identification of levels per metric — Method step 9
  - Pathways of improvement per metric — Method step 10
- Online workshop #3
  - at 09:00 CEST on the 18 June 2019
  - at 17:00 CEST on the 18 June 2019

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 – **Mailing list**

fair_maturity@rda-groups.org

Be one of us!

- Second Workshop 18th of June, 9 -10.30 CET
- RDA 14th Plenary Helsinki, 23 – 25 October

- **Sign to RDA WG today**
- https://www.rd-alliance.org/groups/fair-data-maturity-model-wg

Thank you!