



FAIR Data Maturity Model

Online meeting #1

21 & 22 February 2019

Agenda

- › Welcome, objectives of the meeting
- › Round table
- › Introduction to the Working Group
- › Survey results
- › Presentations from existing approaches
 - › DANS FAIR data assessment tool, FAIR checklist
 - › FAIR Metrics
 - › Data Stewardship Wizard
 - › RDA SHARC IG
 - › Dataset Fitness for Use
 - › ARDC FAIR self-assessment tool
- › Results of preliminary analysis
- › How to contribute
- › Logistics
- › Conclusion

- Short introduction of the chair and editor team
- All other participants, please type your name and affiliation in the chat window

Introduction to the Working Group - 1

› Problem:

- › 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

Introduction to the Working Group - 2

› 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

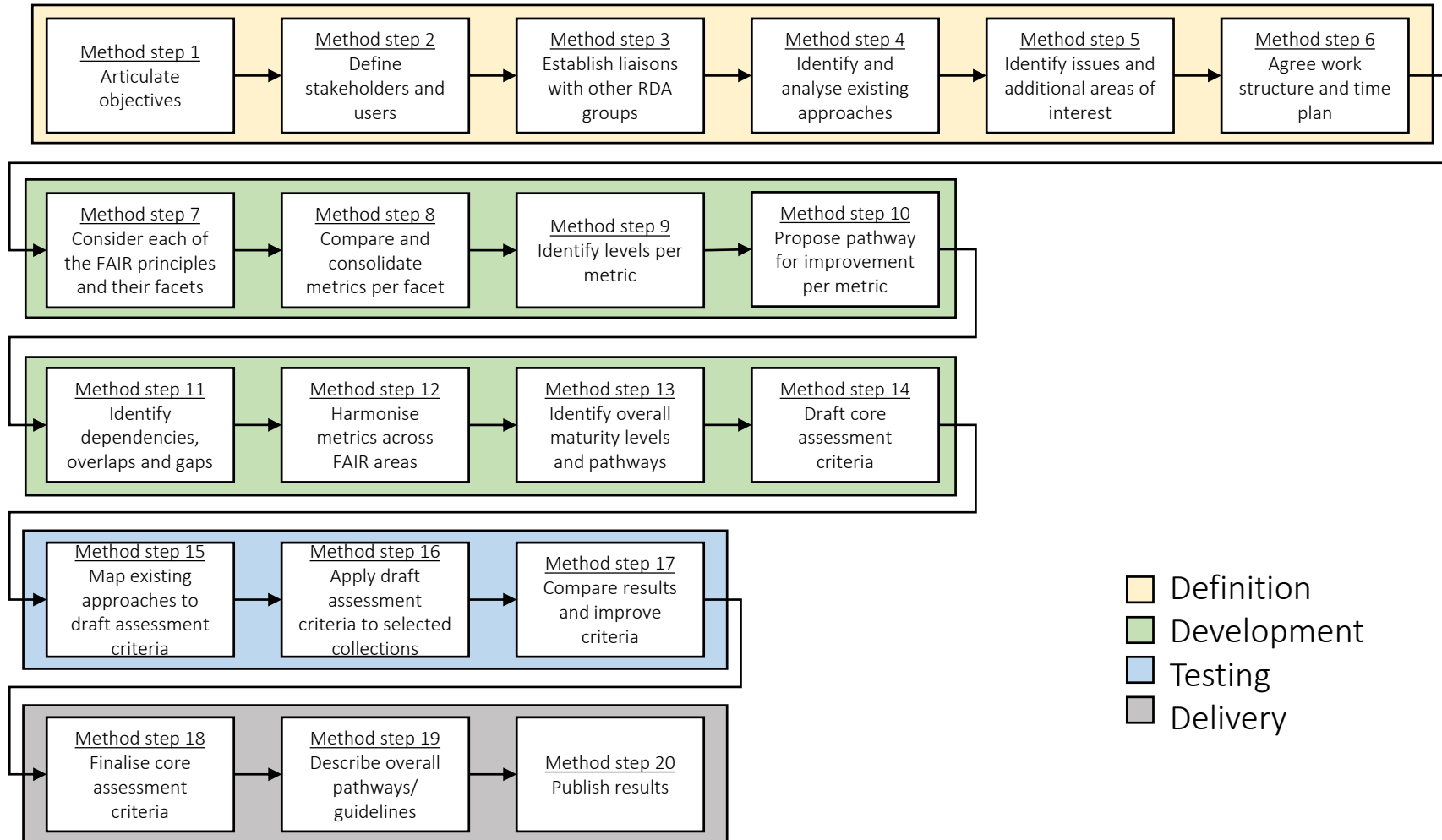
› Scope of the assessment:

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

Any questions about the ***approach*** outlined?

- ① Do you agree with the proposed approach and intended results?
- ② Do you have other suggestions concerning the scope of the work?
- ③ ...

Work methodology

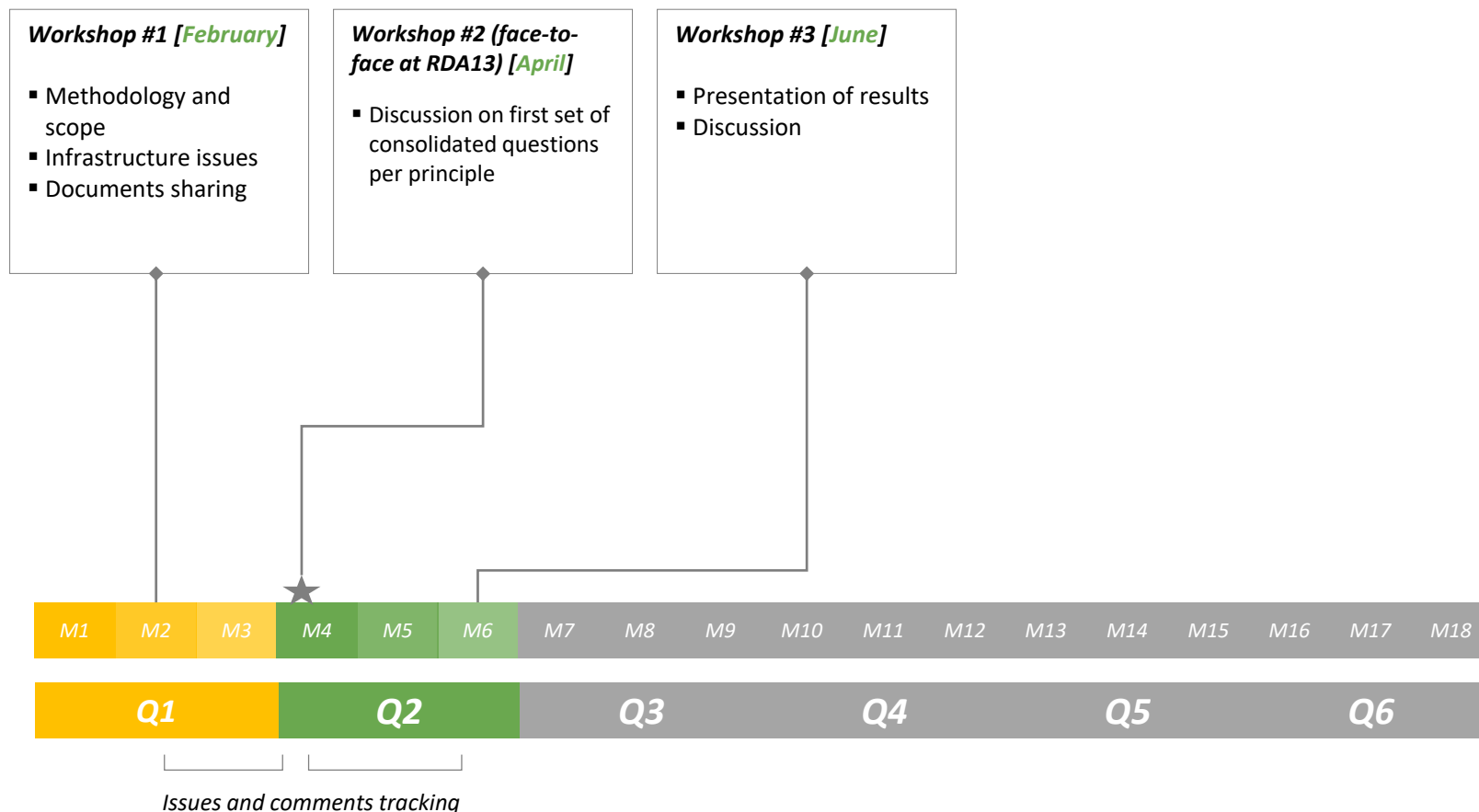


- Definition
- Development
- Testing
- Delivery

Introduction to the Working Group - 4

- Proposed approach to development
 - Consider the assessment of the four FAIR principles in four 'strands'
 - Possibly create a fifth strand for issues related to the environment around the FAIR principles, e.g.
 - Characteristics of projects, workflows and tools
 - Open vs. closed/embargoed data
 - Curation, maintenance and governance
 - Certification (what and who/how)

> Tentative timeline 2019



Any questions about the *methodology*

① Do you agree with the proposed methodology?

② ..

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]

Relevant initiatives

- › Presentation of existing approaches
 - › DANS FAIR data assessment tool, FAIR checklist
 - › **Eliane Fankhauser, DANS**
 - › FAIR Metrics
 - › **Luiz Olavo Bonino, GO-FAIR**
 - › Data Stewardship Wizard
 - › **Rob Hooft, DTL**
 - › RDA SHARC IG
 - › **Laurence Mabile & Romain David, University of Toulouse**
 - › Dataset Fitness for Use
 - › **Jonathan Petters, Virginia Tech**
 - › ARDC FAIR self-assessment tool
 - › **Keith Russell, ARDC**
- › Summary of lessons learnt and open issues
 - › **Makx Dekkers, editor team**



FAIRsFAIR

Fostering Fair Data Practices in Europe

Contributing to FAIR policy and practice in the EOSC: The FAIRsFAIR Project

Eliane Fankhauser

RDA FAIR Data Maturity Model WG,
First virtual meeting, 21/22 February 2019



FAIRsFAIR in a nutshell

- Budget: 10 million euro
- Time plan: 36 months
- Start: March 1 2019
- 22 partners from 8 MS
- 6 core partners



Overall aim

- Development and concrete realisation of an **overall knowledge infrastructure based on the FAIR data principles** on academic quality
 - data management
 - procedures
 - standards
 - metrics ...
- Delivering FAIR aspects of **essential Rules of Participation (RoP)** and regulatory compliance for participation **in the EOSC**
- Contribute to a **FAIR infrastructure of the EOSC**
- Implementation of recommendations from the EOSC HLEG and the Expert Group on FAIR Data.



WP2 (CSC)

- FAIR Practices: Semantics, Interoperability and Services

WP3 (DCC)

- FAIR Data Policy and Practice

WP 4 (DANS)

- FAIR Certification

WP6 (STFC)

- Competence Centre

WP7 (EUA)

- FAIR Data Science and Professionalisation

FAIRsFAIR work and the FAIR Data Maturity Model WG

- Technical implementation of FAIR principles: review of commonalities and gaps regarding semantic interoperability, use of metadata and PIDs [WP2, T2.1]
- Mapping emerging data assessment mechanisms with the FAIR principles to develop pragmatic concepts for FAIRness evaluations at dataset level [WP4, T4.5]
- Badging scheme for assessing the compliance of data resources with the FAIR principles [WP4, T4.5]
- Further development of FAIR data assessment tools including the [FAIRdat tool](#) [WP4, T4.5]

Main challenges

- Being coherent within the project (collaboartion accross WPs)
- Serving an EOSC Governance structure under development
- Creating synergies with all FAIR related projects, initiatives and activities in Europe and beyond





FAIRNESS ASSESSMENT CHALLENGES

RDA FAIR Maturity Model - February 20-21 2019

Luiz Bonino

FAIR PRINCIPLES

Findable:

- F1. (meta)data are assigned a globally unique and persistent identifier;
- F2. data are described with rich metadata;
- F3. metadata clearly and explicitly include the identifier of the data it describes;
- F4. (meta)data are registered or indexed in a searchable resource;

Interoperable:

- I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (meta)data use vocabularies that follow FAIR principles;
- I3. (meta)data include qualified references to other (meta)data;

Accessible:

- A1. (meta)data are retrievable by their identifier using a standardized communications protocol;
 - A1.1 the protocol is open, free, and universally implementable;
 - A1.2. the protocol allows for an authentication and authorization procedure, where necessary;
- A2. metadata are accessible, even when the data are no longer available;

Reusable:

- 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;

<https://www.nature.com/articles/sdata201618>

FAIR DATA PRINCIPLES - METADATA

Findable:

- F1. metadata are assigned a globally unique and persistent identifier;
- F2. data are described with rich metadata;
- F3. metadata clearly and explicitly include the identifier of the data it describes;
- F4. metadata are registered or indexed in a searchable resource;

Interoperable:

- I1. metadata use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. metadata use vocabularies that follow FAIR principles;
- I3. metadata include qualified references to other metadata;

Accessible:

- A1. metadata are retrievable by their identifier using a standardized communications protocol;
 - A1.1 the protocol is open, free, and universally implementable;
 - A1.2. the protocol allows for an authentication and authorization procedure, where necessary;
- A2. metadata are accessible, even when the data are no longer available;

Reusable:

- R1. metadata are richly described with a plurality of accurate and relevant attributes;
 - R1.1. metadata are released with a clear and accessible data usage license;
 - R1.2. metadata are associated with detailed provenance;
 - R1.3. metadata meet domain-relevant community standards;

<https://www.nature.com/articles/sdata201618>

FAIR DATA PRINCIPLES – DATA/DIGITAL RESOURCES

Findable:

- F1. data are assigned a globally unique and persistent identifier;**
- F2. data are described with rich metadata;**
- F3. metadata clearly and explicitly include the identifier of the data it describes;**
- F4. metadata are registered or indexed in a searchable resource;**

Interoperable:

- I1. metadata use a formal, accessible, shared, and broadly applicable language for knowledge representation.**
- I2. metadata use vocabularies that follow FAIR principles;**
- I3. metadata include qualified references to other (meta)data;**

Accessible:

- A1. metadata are retrievable by their identifier using a standardized communications protocol;**
 - A1.1 the protocol is open, free, and universally implementable;**
 - A1.2. the protocol allows for an authentication and authorization procedure, where necessary;**
- A2. metadata are accessible, even when the data are no longer available;**

Reusable:

- R1. metadata are richly described with a plurality of accurate and relevant attributes;**
 - R1.1. metadata are released with a clear and accessible data usage license;**
 - R1.2. metadata are associated with detailed provenance;**
 - R1.3. metadata meet domain-relevant community standards;**

<https://www.nature.com/articles/sdata201618>

FAIR DATA PRINCIPLES – SUPPORTING ELEMENTS

Findable:

- F1. (meta)data are assigned a **globally unique and persistent identifier**;
- F2. data are described with rich metadata;
- F3. metadata clearly and explicitly include the **identifier** of the data it describes;
- F4. (meta)data are registered or indexed in a **searchable resource**;

Interoperable:

- I1. (meta)data use a formal, accessible, shared, and broadly applicable **language for knowledge representation**;
- I2. (meta)data use **vocabularies** that follow FAIR principles;
- I3. (meta)data include qualified references to other (meta)data;

Accessible:

- A1. (meta)data are retrievable by their identifier using a **standardized communications protocol**;
 - A1.1. **the protocol** is open, free, and universally implementable;
 - A1.2. **the protocol** allows for an authentication and authorization procedure, where necessary;
- A2. metadata are accessible, even when the data are no longer available;

Reusable:

- 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**;

<https://www.nature.com/articles/sdata201618>

FAIRNESS ASSESSMENT CHALLENGES

WHY TO ASSESS?

- Because everybody is talking about FAIR and my resources should be seen as FAIR, whatever this means?
- To satisfy funders requirements?
- To serve as a guideline for achieving higher levels of interoperability and reuse with clarity on the concrete benefits (help improve)?

WHAT TO ASSESS?

- Metadata and data?
- Only metadata?
- Only data?
 - What do you mean by data?
 - In the FAIR principles, data refers to a variety of different resources, e.g., “traditional” data, services, software, APIs, vocabularies, ontologies, articles, etc.

HOW TO ASSESS?

■ Manual

- Takes advantage of human understandable artifacts, which are currently prevalent
- May lead to subjective assessments and, therefore, harder to compare resources
- Harder to scale
- Harder to evaluate FAIR for machines, which is the main goal of the FAIR principles

■ Automatic

- Requires more rigor on the assessed resources
- More likely to produce objective assessments
- Easier to scale
- Able to check if machines can, in fact, “work” with the (meta)data

HOW TO “READ” THE ASSESSMENTS?

■ Need for a scoring system

- One score for as 4 aspects of FAIR? Does not seem useful.
- One score per aspect (F, A, I and R)?
- One score per principle? What about the sub-principles?
- Is there a hierarchy among the principles? Is there an order of precedence? Or different weights?
- Is there an acceptable minimal FAIR level? Should it be across domains and applications or domain/community-dependent?
- Do we use a pass/fail approach or introduce intermediary compliance levels in each/some evaluation?

■ Need for a visual representation of the scores

- To facilitate quick perception of the FAIRness level, a visual representation of the FAIR scores is required, e.g., stars, bars, etc.

GENERAL CHALLENGES

- Clarify that nobody has been asked to be 100% FAIR. Many times a lower FAIRness level is perfectly adequate.
- How to deal with the conflicting forces that, from one side want to push the communities towards a better (and FAIRer) data landscape and, from the other side, want to preserve the *status quo* (existing “kingdoms”) but labeling themselves FAIR?
- Who will define the assessment criteria?
- Who will execute the assessments based on the defined criteria?
- Should we have a unique set of assessment criteria? Or a core set for general comparison and domain-specific sets on top of the core for the specific needs of a given domain/application?

■ Moving from metrics to maturity indicators

CURRENT STATUS OF THE FAIR METRICS

- The Maturity Indicator tests are also going to be "incremental". e.g. for the new I indicators there are "weak" and "strong" forms... with loose interpretation of "knowledge representation language" (e.g., CSV) vs strong interpretation (i.e. RDF)
- Full set of fully automatic evaluators almost complete
- Clear separation between the evaluation of metadata and data
- Used (together with the Data Stewardship Wizard) in the "FAIR Funders Pilot", involving Dutch ZonMW and Irish Health Research Board



Luiz Bonino

International Technology Coordinator – GO FAIR
Associate Professor BioSemantics – LUMC

E-mail: luiz.bonino@go-fair.org

Skype: [luizolavobonino](https://www.skype.com/people/luizolavobonino)

Web: www.go-fair.org

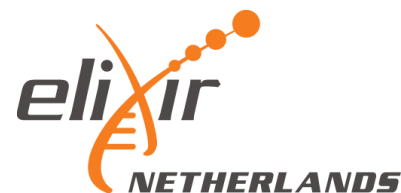
DTL



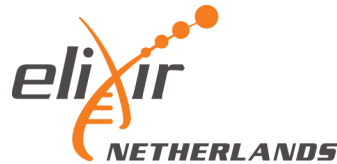
DUTCH TECHCENTRE FOR LIFE SCIENCES

DATA STEWARDSHIP WIZARD ROB HOOFT / ROBERT PERGL

21 and 22 February 2019



DTL



MOTIVATION FOR DATA STEWARDSHIP WIZARD

- Software tool for Smart Data Management Plans for FAIR Open Science
- *Help* researcher with Data Management
 - Smart questionnaire system
 - Expert system
 - Not: fulfil *requirements*
- **Target audience:**
 - Researcher (awareness of options, pointers)
 - Data Steward (checklist)
 - Data Expert (being found)
 - Funder (evaluate DMP)



DSW

DATA STEWARDSHIP WIZARD

<https://ds-wizard.org/>



APPROACH

- Hierarchical
 - Based on mind-map
 - Relevant questions in context
 - No attempt to limit it
- Links to supporting materials and experts
- Localizable: add your local experts and intranet resources

Current Phase
Before Submitting the Proposal

Design of experiment: ✓

- Data design and planning ✓
- Data Capture/Measurement ✓
- Data processing and curation ✓
- Data Integration ✓
- Data Interpretation ✓
- Information and Insight ✓

Summary Report

Design of experiment

Before you decide to embark on any new study, it is nowadays good practice to consider all options to keep the data generation part of your study as limited as possible. It is not because we can generate massive amounts of data that we always need to do so. Creating data with public money is bringing with it the responsibility to treat those data well and (if potentially useful) make them available for re-use by others.

Is there any pre-existing data?

Are there any data sets available in the world that are relevant to your planned research?

Desirable: *Before Submitting the DMP*
 Data Stewardship for Open Science: *atq*

No
 Yes
 Clear answer

Will you be using any pre-existing data (including other people's data)?

Will you be referring to any earlier measured data, reference data, or data that should be mined from existing literature? Your own data as well as data from others?

Desirable: *Before Submitting the DMP*
 Data Stewardship for Open Science: *ezi*

No



Data Stewardship for Open Science: Chapter 1.1

With kind permission of

Is there pre-existing data?

What's up?

For many decades if not centuries, virtually every experiment started with the collection or creation of 'observations' and in fact data. In social sciences and humanities the tendency to 'reuse' data that had been created earlier, in all kinds of surveys and increasingly of course from sources such as social media maybe already somewhat more established. However, in many of the hard experimental sciences, the generation of new data specifically generated to answer a hypothetical question is still so commonplace that careful thinking about the actual need to generate new data may just not be on the radar screen. Obviously, data creation will need to continue, but increasingly we have to ask the question whether such new data are absolutely necessary to answer the question we want to answer. With more and more data becoming available in reusable format, there may well be existing data collections 'Other People's' Data and associated Services (OPEDAS) that without or with some extra effort needed, can answer at least part of the question or least may be crucial for the interpretation of your own data.

Do

- Search for data sets (OPEDAS) that may be re-usable and can help you to reduce the number of new data sets you may have to generate (and

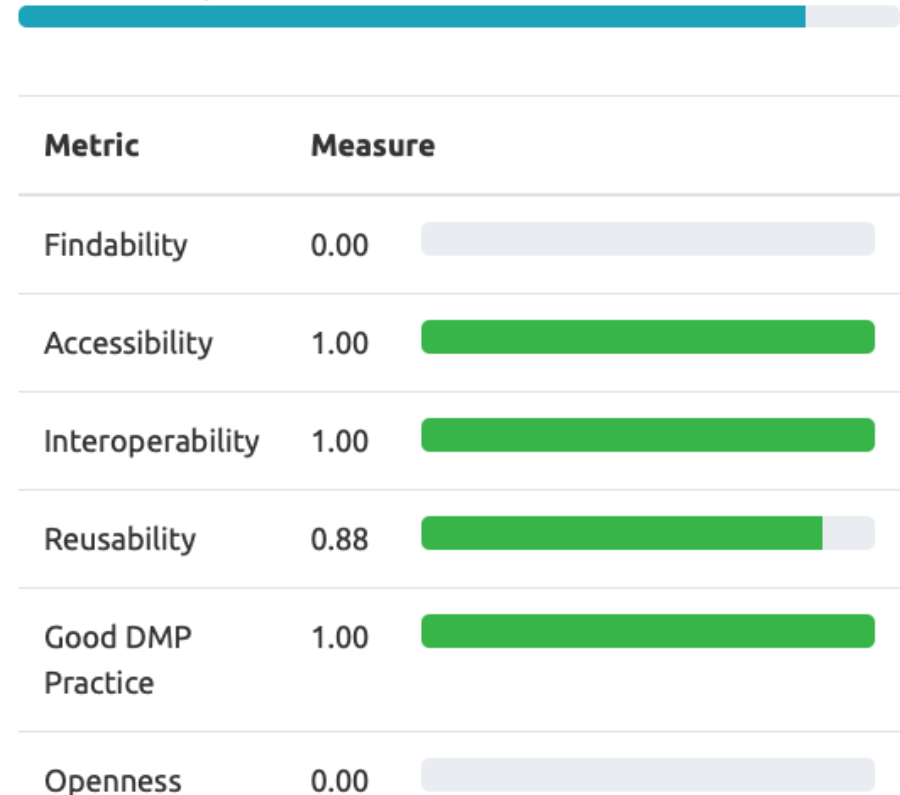


METRICS FOR FAIR

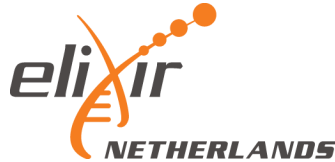
- No dedicated questions to probe the FAIR metrics
- *Instead:* Every question helps to measure
- Fully based on answers in the questionnaire

Data design and planning

Answered: 25/28



DTL



DMP

- No dedicated questions to fill a DMP template
- *Instead:* template engine (under development) uses the answers in the wizard to write a required DMP
- Fully based on answers in the questionnaire

Science Europe DMP

Organization: ELIXIR Global
Based on: Common ELIXIR Knowledge Model, 1.0.0
Project phase: Before Submitting the Proposal
Created at: 15.02.2019



Data Collection

What data will you collect or create?

The following instrument datasets will be acquired in the project:

- **Genomic data**
This dataset will be collected by an external party. For the ownership of the data we have made the following arrangements: "Ownership will stay with the external party for five years and then transfer to our library". The equipment is very well described and known.
- **Proteomics data**
This dataset will be collected by experts in the project, with our own equipment. The equipment is less well described or not completely standard, so we will need to take extra care documenting the process.

We also collect data from questionnaires, case report forms, and electronic patient records.
We will use the following reference datasets:

- **GRCh37lite**
We will use version "1.0.0-rc139" of this dataset. If a new version becomes available during the project, new analyses will be done with the new version.
- **GENCODE**
We will use version "2.0.0" of this dataset. If a new version becomes available during the project,

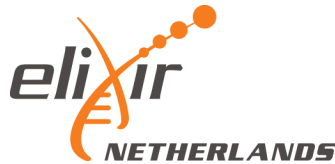
We will use the following already existing non-reference datasets:

- Last year's experimental data
- Statistics made by FIT CTU

Data Management Plan generated by Data Stewardship Wizard <<https://ds-wizard.org>>



DTL



EXPERIENCE / LESSONS LEARNED / OPEN ISSUES

- Worked on components (mind map) since 2013
- Very successful NL/CZ collaboration with clear task split
- Currently advertised as “source of inspiration for making a DMP” for researcher writing a proposal
- Many ideas on how the wizard can be further improved
- Interest in approach from many academic organisations
- Installation/collaboration at companies (e.g. DSM)
- Broad interest but adoption of new approach takes time
- Concrete plans with ZonMw funder (+HRB, Science Europe)
- Acquisition of funding for further development



RDA-SHARC fairness assessment tools for crediting/rewarding scientists data sharing activities

CONTEXT: the **rda-SHaring Reward & Credit ig**, *Corresponding authors : R. David, L. Mabile, A. Cambon-Thomsen*

What for? to foster data sharing by improving recognition of the work required

How? by providing a set of recommendations to guide researchers and other relevant stakeholders (research institutions administrators, funders, policy makers and publishers/editors) in moving through the necessary steps towards crediting and rewarding in the data/resources- sharing process (in progress); and to encourage the adoption of data sharing activities- related criteria in the research evaluation process at the institutional, national and European/international levels.

As part of it, **3 human readable assessment tools** are **under development** that will assess semi-quantitatively the fairness knowledge & practices of scientists:

1.1 extensive FAIRness external assessment grid

52 criteria so far

1.2 simplified FAIRness external assessment grid (can be used as a quick self-assessment grid)

18 essential criteria

<https://zenodo.org/record/2551500#.XGK4llxKg2w>

2.2 extensive checklist for fairness self-assessment (adapted from the 2 previous grids)

Fairness assessment grids

Objectives : credit & reward for FAIRness in researchers sharing behaviors
 -> necessity to improve FAIRness (understandable and step by step processes)

Main properties:

- As simple as possible (understandable by non IT people)
- Easy to complete (due to FAIR skills availability in evaluation processes)
- Based on informations given by researchers in careers doc / activity reports
- CC author license (can be re-used by anyone at the end of the implementation)

Assessment process: leading to recommendations to improve fairness

- Designed as a decision tree in each FAIR Principle
- 3 Level of criterion importance : **essential / recommended / desirable**
- 4 possible answers/criteria:
 - Never/NA If Mandatory Sometimes Always
- Evaluation based on scoring each answer for each F.A.I.R. principle
 ex: **Findable 2/8** Never/NA; **3/8** If Mandatory; **1/8** Sometimes; **2/8** Always
- Recommendations based on this scoring

1) FINDABLE (8 essential criteria)

Indexed identifier ?
 Identification Never/NA If Mandatory Sometimes Always
 Are each identifier identified by an internet accessible identifier?

Unique, global, persistent ID?
 Identification Never/NA If Mandatory Sometimes Always
 Are the data identifiers unique, global and persistent?

ID scheme?
 Identification Never/NA If Mandatory Sometimes Always
 Has any identifier scheme been used for data (e.g. DOI)?

Persistent metadata / data link ?
 Metadata availability Never/NA If Mandatory Sometimes Always
 Are the metadata linked to the dataset through a persistent identifier?

Metadata & authority linked ?
 Metadata availability Never/NA If Mandatory Sometimes Always
 Are the metadata of each related dataset a unique authority responsible for the dataset at a given time?

Datasets linked to authority ?
 Metadata availability Never/NA If Mandatory Sometimes Always
 Are all datasets linked to an authority (legal entity) through a unique and persistent identifier over time (e.g. institution, association or established body)?

Standards/dictionary for data description?
 Metadata description and searchability Never/NA If Mandatory Sometimes Always
 If relevant, has the researcher used and published standards for data description? If no, are the data standards and particularly missing data standards appropriately represented or appropriate activities undertaken? If so, do they work, has the researcher created and described new standards?

Data format/type description?
 Metadata description and searchability Never/NA If Mandatory Sometimes Always
 Are the types and formats of data provided (collected well described)?

Result for Findable: Never/NA If Mandatory Sometimes Always

2) ACCESSIBLE (3 essential criteria)

Data repositories?
 Repository Never/NA If Mandatory Sometimes Always
 Does the researcher use data repositories for the storage of data?

Efficient and rich services for various uses & users?
 Data security and services Never/NA If Mandatory Sometimes Always
 Does the researcher use efficient and rich services to access data (various formats, visualisation, practical tools and systems adapted to different types of use and users)?

Data access restriction justification?
 Access restriction Never/NA If Mandatory Sometimes Always
 In case of a non legal restricted access, is the restriction properly justified by the researcher?

Result for Accessible: Never/NA If Mandatory Sometimes Always

3) INTEROPERABLE (2 essential criteria)

Standard vocabularies, thesaurus, ontologies or data dictionary?
 Identification Never/NA If Mandatory Sometimes Always
 Are standard vocabularies, thesaurus or ontologies used for all data types present in datasets, to enhance interoperability between and within datasets? If not, is a well defined open data dictionary provided?

Interoperability criteria explained?
 Identification Never/NA If Mandatory Sometimes Always
 Are the interoperability criteria explained?

Result for Interoperability: Never/NA If Mandatory Sometimes Always

4) REUSABLE (5 essential criteria)

Relevant actions for data reuse potential?
 Data potential Never/NA If Mandatory Sometimes Always
 Have relevant actions been validated by the researcher to enhance the data reuse potential?

Provenance for raw and transformed data?
 Data availability Never/NA If Mandatory Sometimes Always
 Are the provenance and type of all data properly specified (stage of raw, primary, secondary, derivative)?

Information on methods and tools that permit the understanding, integrity of data?
 Reusability tools Never/NA If Mandatory Sometimes Always
 Does the researcher provide information on methods and tools that permit the understandability, integrity, value and usability of data intended to be kept on the long-term? (e.g. archiving, archival and long term reuse tools for protocols, software, required methods and contacts to create, host and understand data)

Data sharing arrangements meet data ethics and protection?
 Reusability right Never/NA If Mandatory Sometimes Always
 Do the data reuse control and data sharing arrangements meet the data producer and institutional ethics requirements?

Legal reuse restriction properly justified?
 Reusability right Never/NA If Mandatory Sometimes Always
 In case of a legal reuse restriction (such as personal data, trade and/or publicly security, national defence, social, commercial or financial activities, information security, security, security in research and commercial matters), is the restriction properly justified?

Result for Reusable: Never/NA If Mandatory Sometimes Always

TOTAL FAIR simple criteria evaluation results:
 .../18 'Never/NA' .../18 'If Mandatory' .../18 'Sometimes' .../18 'Always'
 *Indicators will be provided according to the criteria predominantly obtained

Fairness assessment grids

Lessons learnt from the first tests:

- Essential criteria not always understandable without training
- Implementation of some criteria can be time consuming / need technical advisor / operator

Possible open issues:

- Develop iterative assessment of the *researcher FAIRness Literacy*
- Help identify needs to build FAIRness guidelines for a better researcher sharing capacity
(based on rewards and credits / *How to do and step by step* tools)

Next steps:

- **Upcoming SHARC-survey launch to evaluate the external assessment extensive grid usability:** please participate!
- RDA P13 Sharc's session: please attend!
- Tools experimentation in specific networks (IMI FAIRplus; BiodiFAIRse; Citizen science networks...)

1) FINDABLE (8 essential criteria)	
Indexed identifier ? Identification	Never/NA <input type="checkbox"/> If Mandatory <input type="checkbox"/> Sometimes <input type="checkbox"/> Always <input type="checkbox"/>
<small>Are each subdataset identified by an indexed and persistent identifier?</small>	
Unique, global, persistent ID? Identification	Never/NA <input type="checkbox"/> If Mandatory <input type="checkbox"/> Sometimes <input type="checkbox"/> Always <input type="checkbox"/>
<small>Do the data identifiers uniquely global and persistent?</small>	
ID scheme? Identification	Never/NA <input type="checkbox"/> If Mandatory <input type="checkbox"/> Sometimes <input type="checkbox"/> Always <input type="checkbox"/>
<small>Has any identifier scheme been used for data (e.g. DOI)?</small>	
Persistent metadata / data link ? Metadata traceability	
<small>Are the metadata linked to the dataset through a persistent identifier?</small>	
Metadata & authority linked ? Metadata traceability	Never/NA <input type="checkbox"/> If Mandatory <input type="checkbox"/> Sometimes <input type="checkbox"/> Always <input type="checkbox"/>
<small>Are the metadata of each dataset linked to a central authority (responsible for the datasets at a given time)?</small>	
Datasets linked to authority ? Metadata traceability	Never/NA <input type="checkbox"/> If Mandatory <input type="checkbox"/> Sometimes <input type="checkbox"/> Always <input type="checkbox"/>
<small>Are all datasets linked to an authority (legal entity) through a unique and persistent identifier (e.g. metadata, collection or individual data)?</small>	
Standards/dictionary for data description? Metadata description and searchability	
<small>Does the researcher use well and updated description for data sharing? If not, are the data description and searchability following the standards recommended by community/department or appropriate authorities specified? If no standards exist, has the researcher created a well-structured data dictionary?</small>	
Data format/type description? Metadata description and searchability	Never/NA <input type="checkbox"/> If Mandatory <input type="checkbox"/> Sometimes <input type="checkbox"/> Always <input type="checkbox"/>
<small>Are the types and formats of data generated/collected well described?</small>	
Result for Findable: <input checked="" type="checkbox"/> Never/NA <input checked="" type="checkbox"/> If Mandatory <input type="checkbox"/> Sometimes <input type="checkbox"/> Always <input type="checkbox"/>	
2) ACCESSIBLE (3 essential criteria)	
Data repositories? Repository	Never/NA <input type="checkbox"/> If Mandatory <input type="checkbox"/> Sometimes <input type="checkbox"/> Always <input type="checkbox"/>
<small>Does the researcher use data repositories for the storage of data?</small>	
Efficient and rich services for various uses & users?	
Never/NA <input type="checkbox"/> If Mandatory <input type="checkbox"/> Sometimes <input type="checkbox"/> Always <input type="checkbox"/>	
<small>Does the researcher use efficient and rich services to access data (online formats, interoperability, practical tools and content related to different types of use and users)?</small>	
Data access restriction justification? Access restriction	Never/NA <input type="checkbox"/> If Mandatory <input type="checkbox"/> Sometimes <input type="checkbox"/> Always <input type="checkbox"/>
<small>Is there a well-justified access restriction, in the metadata properties, justified by the researcher?</small>	
Result for Accessible: <input checked="" type="checkbox"/> Never/NA <input checked="" type="checkbox"/> If Mandatory <input type="checkbox"/> Sometimes <input type="checkbox"/> Always <input type="checkbox"/>	
3) INTEROPERABLE (2 essential criteria)	
Standard vocabularies, thesaurus, ontologies or data dictionary? Identification	Never/NA <input type="checkbox"/> If Mandatory <input type="checkbox"/> Sometimes <input type="checkbox"/> Always <input type="checkbox"/>
<small>Are standard vocabularies, thesaurus or ontologies used for all data open projects? In addition, is each vocabulary or interoperable network well defined (stable)? If not, is a well-defined open data dictionary provided?</small>	
Interoperability criteria explained? Identification	
Never/NA <input type="checkbox"/> If Mandatory <input type="checkbox"/> Sometimes <input type="checkbox"/> Always <input type="checkbox"/>	
<small>Are the interoperability criteria explained?</small>	
Result for Interoperability: <input checked="" type="checkbox"/> Never/NA <input checked="" type="checkbox"/> If Mandatory <input type="checkbox"/> Sometimes <input type="checkbox"/> Always <input type="checkbox"/>	
4) REUSABLE (5 essential criteria)	
Relevant actions for data reuse potential? Data potential	Never/NA <input type="checkbox"/> If Mandatory <input type="checkbox"/> Sometimes <input type="checkbox"/> Always <input type="checkbox"/>
<small>What relevant actions have been undertaken by the researcher to enhance the data reuse potential?</small>	
Provenance for raw and transformed data? Data traceability	
Never/NA <input type="checkbox"/> If Mandatory <input type="checkbox"/> Sometimes <input type="checkbox"/> Always <input type="checkbox"/>	
<small>Are the provenance and type of all data properly specified (stage of raw, primary, transformed, secondary)?</small>	
Information on methods and tools that permit the understanding, integrity of data? Reusability tools	
Never/NA <input type="checkbox"/> If Mandatory <input type="checkbox"/> Sometimes <input type="checkbox"/> Always <input type="checkbox"/>	
<small>Does the researcher provide information on methods and tools that permit the understanding, integrity, value and usability of data shared to the best of the researcher's (e.g. metadata, archival access) know-how (e.g. for protocols, software, workflow methods and conditions to create, store and production code)?</small>	
Data sharing arrangements meet data ethics and protection? Reusability rights	
Never/NA <input type="checkbox"/> If Mandatory <input type="checkbox"/> Sometimes <input type="checkbox"/> Always <input type="checkbox"/>	
<small>Do the researcher control and data sharing arrangements meet the data protection and confidentiality ethics requirements?</small>	
Legal reuse restriction properly justified? Reusability rights	
Never/NA <input type="checkbox"/> If Mandatory <input type="checkbox"/> Sometimes <input type="checkbox"/> Always <input type="checkbox"/>	
<small>Is use of a legal reuse restriction (such as copyright, data, trade and public interests, national defence or security, confidentiality of external relations, intellectual property, security, trademark and commercial interests) in the metadata properly justified?</small>	
Result for Reusable: <input checked="" type="checkbox"/> Never/NA <input checked="" type="checkbox"/> If Mandatory <input type="checkbox"/> Sometimes <input type="checkbox"/> Always <input type="checkbox"/>	
TOTAL FAIR simple criteria evaluation results:	
.../18 'Never/NA' .../18 'If Mandatory' .../18 'Sometimes' .../18 'Always'	
<small>*ratios will be provided according to the criteria predominantly obtained</small>	

WDS/RDA Assessment of Data Fitness for Use WG

Goals:

- Specify criteria of dataset reusability expanding on FAIR principles
- Develop process by which a repository/data provider could assess their holdings for reusability

Outputs:

- Criteria for fitness for use, compared against CoreTrustSeal requirements and FAIR principles ([spreadsheet](#))
- Checklist for evaluation of dataset for fitness for use ([form](#)) ([pdf](#))
 - designed as a CoreTrustSeal certification add-on

WDS/RDA Assessment of Data Fitness for Use WG

Lessons learned/open issues

- CoreTrustSeal certification goes a ways towards providing for data reuse (covers F and A, less so I and R)
- Our practical assessment approach has caveats (also see notes)
 - Manual approach; hard to automate checks for metadata completeness and data correctness that require domain expertise to evaluate
 - Domain expertise of evaluator matters in assessment
 - Neglects data user perspective
- Many domains have not established metadata standards towards reusability



Australian Research Data Commons

FAIR Self Assessment Tool

PRESENTED BY

Keith Russell, Manager Engagements, 21 February 2019

Choices

- For research support staff
- Kept close to the principles
- Assessing a data set
- Hard to provide a score across disciplines
- Just a bar rather than a score
- With guidance included

FAIR Self-assessment Tool

Welcome to the ARDC FAIR Data self-assessment tool. Using this tool you will be able to assess the 'FAIRness' of a dataset and determine how to enhance its FAIRness (where applicable).

This self-assessment tool has been designed predominantly for data librarians and IT staff, but could be used by software engineers developing FAIR Data tools and services, and researchers provided they have assistance from research support staff.

You will be asked questions related to the principles underpinning Findable, Accessible, Interoperable and Reusable. Once you have answered all the questions in each section you will be given a 'green bar' indicator based on your answers in that section, and when all sections are completed, an overall 'FAIRness' indicator is provided.

Please be aware that additional explanatory information is provided within the tool. The (i) information button provides an overview of each of the FAIR high-level elements (Findable, Accessible, Interoperable and Reusable). Additionally, each question is hyperlinked, leading users to explanatory information and links to wider resources on related topics.

Findable ⓘ

Does the dataset have any identifiers assigned? Globally Unique, citable and persistent (e.g. DOI, PURL, ARK c)

Is the dataset identifier included in all metadata records/files describing the data? Yes

How is the data described with metadata? Brief title and description

What type of repository or registry is the metadata record in? Generalist public repository

Accessible ⓘ

Interoperable ⓘ

Reusable ⓘ

Total across F.A.I.R.

Uptake/Feedback

Used in workshops in Australia by institutions

Used in paper form

Used for developers in Agriculture and BioSciences projects

International interest

‘We want a score’

Links

The Self Assessment tool

<https://www.ands-nectar-rds.org.au/fair-tool>

Survey on 'How well does your repository enable FAIR?'

<https://www.slideshare.net/kgrussell/how-well-does-your-repository-support-fair-poll-results>

Training resources categorised by FAIR

<https://www.ands.org.au/working-with->

[data/fairdata/training](https://www.ands.org.au/working-with-data/fairdata/training)

Training for tool designers to enable FAIR

<https://www.ands.org.au/working-with-data/fairdata/fair-for-developers>

10 FAIR data things

<https://librarycarpentry.org/Top-10-FAIR/>

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

Any questions about the *lessons learnt and open issues* presented?

① Which open issues could be considered in this exercise?

② ...

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 [\[Link\]](#)
- > FAIRR – Reusable [\[Link\]](#)
- > Beyond the FAIR principles (X) [\[Link\]](#)

Example

> Questions, options and potential overlaps

A2 metadata is accessible, even when the data are no longer available

1 Will the metadata record be available even if the data is no longer available?

- No
- Unsure
- Yes

2 Are the metadata accessible?

- No
- Yes

5 Please provide the URL to a metadata longevity plan

7 The existence of metadata even in the absence/removal of data



PDF

F4

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?

Any opinions about the ***additional aspects*** to be considered?

① Which other aspects should the WG consider?

② ...

How to contribute - 1

› Contribution is sought and welcomed for

METHODOLOGY

E.G.

- › Missing items
- › Alternative approach
- › ...

ANALYSIS

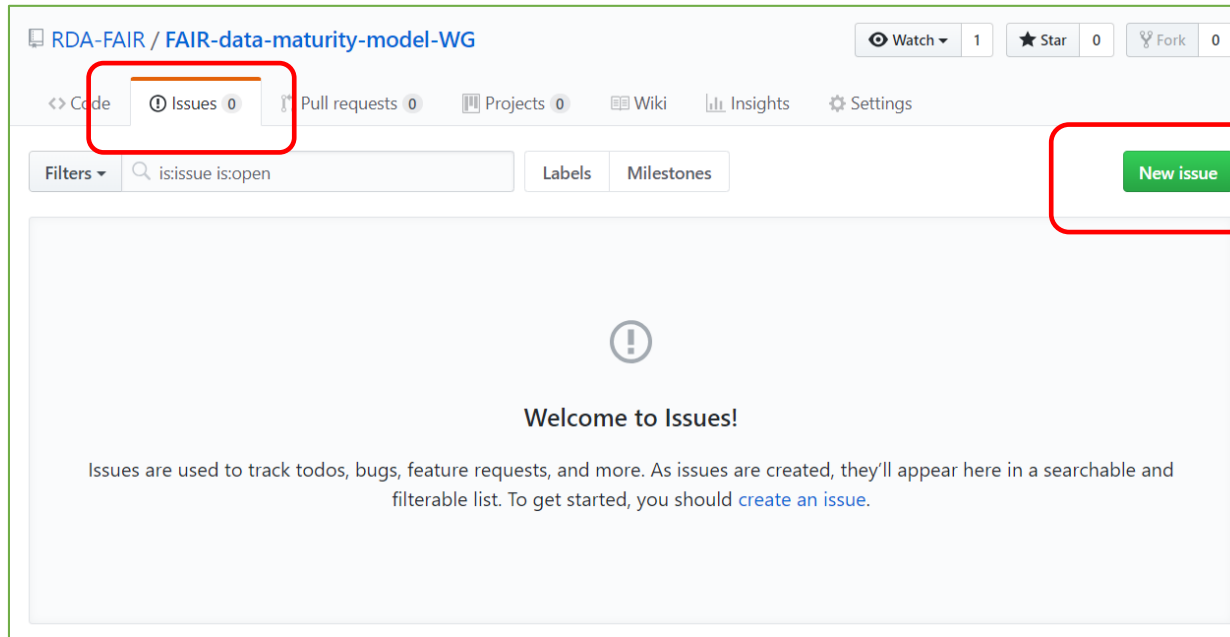
E.G.

- › Scope
- › Irrelevant items
- › Missing items
- › Additional aspects
- › ...

AOB

...

➤ 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**]

- › 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

Conclusion

› Action items

- › Feedback via [GitHub](#)
 - › [Work methodology](#)
 - › [Work process](#)
 - › [Tentative timeline](#)
 - › [Results of preliminary analysis](#)

› Next steps

- › *Issues and comments* review period
- › RDA 13th Plenary Session [Philadelphia, USA]
- › Online workshop #3