

CASE STATEMENT

for a new RDA Working Group
"FAIR Data Maturity Model: core criteria to assess the
implementation level of the FAIR data principles"

version 1.7

1. WG CHARTER

1.1 CONTEXT

Technological advancements have made science more data intensive and interconnected, with researchers producing and sharing increasing volumes of research data. To maximise the value of science, research data (sets) should have four foundational characteristics; they should be:

- 'Findable', i.e. discoverable with machine readable metadata, identifiable and locatable by means of a standard identification mechanism;
- 'Accessible', i.e. available and obtainable;
- 'Interoperable', i.e. both syntactically parseable and semantically understandable, allowing data exchange and reuse among scientific disciplines, researchers, institutions, organisations and countries; and
- 'Reusable', i.e. sufficiently described and shared with the least restrictive licences, allowing the widest reuse possible across scientific disciplines and borders, and the least cumbersome integration with other data sources.

Findability, Accessibility, Interoperability and Reusability – the FAIR principles – were first introduced in 2014 and intend to define a minimal set of community-agreed guiding principles and practices that allow both machines and humans to find, access, interoperate and re-use research data. The FAIR principles define characteristics that contemporary research data resources, vocabularies and infrastructures should exhibit to assist discovery and reuse by third-parties and they can be further refined into a range of facets that have the potential to: a) improve scientific research, b) contribute to growth and accelerate innovation in a global digital economy, c) increase the reproducibility of research and d) better inform citizens and society about the results and value of research (through thorough and comprehensible description of the data sets).

1.2 PROBLEM

The aspirational nature of the FAIR data principles and their rapid adoption at international level has led to an ambiguity and a wide range of interpretations of FAIRness since the principles do not strictly define how to achieve a state of FAIRness but rather they describe a continuum of features, attributes and behaviours that move a digital object closer to that goal. As a result, a number of incompatible methodologies to assess FAIRness have been developed already and relevant work is in under way by various groups.

Due to the lack of a common set of core assessment criteria for FAIRness, researchers and organisations cannot evaluate the readiness and implementation level of their datasets vis-à-vis the FAIR data principles in a coherent way. The majority of the available FAIR assessment frameworks: i) produce results which cannot be combined or compared and ii) do not allow a benchmark based on the comparison amongst peers. In addition, research performing

organisations and data infrastructures cannot develop or follow a minimum set of shared guidelines to climb up the ladder of FAIR because of the increased heterogeneity of the offered FAIR metrics tools.

1.3 OUTCOMES

The Working Group "FAIR data maturity model: core criteria to assess the implementation level of the FAIR data principles" will bring together stakeholders from different scientific and research disciplines, the industry and public sector, who are active and/or interested in the FAIR data principles and in particular in assessment criteria and methodologies for evaluating their real-life uptake and implementation level. The Working Group will develop as an RDA Recommendation a common set of core assessment criteria for FAIRness and a generic and expandable self-assessment model for measuring the maturity level of a dataset from the following perspectives:

- Data findability, i.e. how well it describes the data it produces or manages with rich metadata, assigns to data/metadata a globally unique persistent identifier and registers or indexes them in a searchable resource;
- Data accessibility, i.e. how well it allows the retrieval of its data/metadata by their identifier using a standardized communications protocol that is open, free and universally implementable;
- Data interoperability, i.e. how well it ensures that the precise format and meaning of exchanged and shared data/metadata is preserved and understood;
- Data reusability, i.e. how well it releases data/metadata with a clear and accessible data usage license, associated with detailed provenance and follows practices that promote the reuse and share of data, unless certain privacy or confidentiality restrictions apply.

In addition, the Working Group will design:

- A self-assessment toolset that enables researchers and organisations to evaluate and improve the readiness and implementation level of their datasets vis-à-vis the FAIR data principles.
- A lightweight version of the FAIR Data Maturity Model (aka FAIR data checklist), aiming to raise awareness on the main aspects related with the FAIR principles.

The outcomes of the Working Group will be possible to be applied not only to data in the conventional sense but also to data-related algorithms, tools, workflows, protocols and other data-related services produced or managed by the assessed entity.

2. VALUE PROPOSITION

Given that the outcomes of the Working Group will be in the form of generic and reusable building blocks, researchers and organisations will be in a position to easily apply and extend them in order to address FAIR-related assessment needs specific to their own thematic

disciplines and/or countries. That will increase the coherence and interoperability of existing or emerging FAIR assessment frameworks and it will ensure the combination and compatibility of their results in a meaningful way.

The outcomes of the Working Group "FAIR data maturity model: core criteria to assess the implementation level of the FAIR data principles" will benefit:

- Researchers, data stewards and other data professionals who are involved in the production and management of research data and have to follow good data management and data stewardship practises (which include the notions of data collection, annotation, archival and long-term care, either alone or in combination with newly generated data).
- Data services owners (data infrastructures, data repositories, owners of commercial and open-source tools), who are responsible for setting up and maintaining a data-related services and tools.
- Organisations that capture, generate, manage, share, protect and preserve research data.
- Policymakers who are responsible for defining data policies at international, European and national level.

The Working Group will provide to the aforementioned user categories an instrument with a three-fold nature:

- a. It will be descriptive, i.e. it will describe the as-is FAIR-related maturity level of a dataset,
- b. It will be prescriptive, i.e. it will provide guidance to researchers and organisations to improve the implementation of the FAIR data principles (aka 'FAIRness') through recommendations, and
- c. It will be comparative, i.e. it will allow a benchmark based comparison amongst peers.

In addition, the outcomes of the Working Group are expected to:

- Contribute to growth and accelerate innovation in a global digital economy: since data is becoming increasingly important for all aspects of the international economy, a common set of core assessment criteria and the FAIR Data Maturity Model will improve the readiness and capability of organisations to open up their data in a way that creates potential benefits for their investment plans (a specific example in Europe of the economic impact of opening up data is the Copernicus earth observation system).
- Provide savings in money: the outcomes of the Working Group will ensure money savings to researchers and organisations as it will deliver a reusable solution for measuring the FAIRness of their data. Also, it will contribute to the improvement of their readiness and implementation level of the FAIR principles, which will lead to money savings from the reuse of high-quality data, the combination of data sets across borders or disciplines and the avoidance of duplication.

-
- Provide savings in time for researchers and organisations aiming to implement the FAIR principles.
 - Increase transparency: better and faster implementation of the FAIR data principles will help to increase the reproducibility of research, which currently can be as low as 10-30% in key areas, such as cancer research. This can have a positive impact for the scientific principle of credibility, replication and further research given that the scientific community has repeatedly experienced instances of misconduct and erroneous analyses, which may endanger whole scientific fields

3. ENGAGEMENT WITH EXISTING WORK IN THE AREA

The Working Group "FAIR data maturity model: core criteria to assess the implementation level of the FAIR data principles" must build upon existing relevant efforts at international, European and sectorial level and will complement emerging activities (e.g. funded by the H2020 Work Programme 2018-20) that support the FAIR data uptake and compliance across borders/disciplines.

Research Data Alliance offers an ideal environment for an engagement of this kind because RDA can bring deep knowledge from the promotion of research data interoperability at disciplinary levels together with hands-on experience in leveraging such knowledge in order to improve interoperability amongst scientific disciplines too.

Two of RDA groups having this twofold nature are the Disciplinary Collaboration Framework Interest Group and the Domain Repositories Interest Group, which both enhance communication with other RDA IGs and WGs and represent the interests of specific disciplines in those groups. The Working Group "FAIR data maturity model: core criteria to assess the implementation level of the FAIR data principles" will work closely with the aforementioned Interest Groups aiming to: a) capture interoperability needs between disciplines and in view of specific scientific challenges, b) gather relevant input from different disciplines and c) develop and apply a structured methodology for prioritising, harmonising and efficiently articulating inter-disciplinary needs.

In parallel, the Working Group will investigate opportunities for collaboration with existing or emerging RDA groups that address any aspect that is relevant with the implementation of the FAIR data principles, such as the:

- Data Description Registry Interoperability (DDRI) WG
- DMP Common Standards WG
- Exposing Data Management Plans WG
- RDA/FORCE11 FAIR Sharing Working Group
- Metadata Standards Catalog WG
- WDS/RDA Assessment of Data Fitness for Use WG
- Research Data Repository Interoperability WG
- RDA/CODATA Legal Interoperability IG

-
- Data policy standardisation and implementation IG
 - Education and Training on handling of research data IG
 - Metadata IG
 - PID IG

In addition, the Working Group will engage with pertinent international and European actors and activities such as:

- FAIRmetrics.org: a group collaborating with a broad set of stakeholders to design a framework for evaluating "FAIRness" that enables both qualitative and quantitative assessment of the degree to which online resources comply with the FAIR Data principles.
- Horizontal or discipline-specific initiatives to measure the implementation of the FAIR data principles such as:
 - DANS FAIR data assessment tool: an online tool prototype which guides the user through a set of questions to assess a specific dataset.
 - ARDC FAIR Data self-assessment: a self-assessment tool designed predominantly for data librarians and IT staff to assess the 'FAIRness' of a dataset and determine how to enhance its FAIRness (where applicable).
 - CSIRO 5-star data rating tool: a tool that allows users to carry out a self-assessment based on 5 qualities of data – Findable, Accessible, Interoperable, Reusable and Trusted. For each quality, a number of specific questions have been curated to allow users to rate their data according to its current state.
- GO FAIR: a community-led initiative to contribute to and coordinate the coherent development of the Internet of FAIR Data & Services. GO FAIR is analysing the possibility to organise a FAIR certification mechanism of services, tools, organisations, and people (including data stewards) aiming to help research funders and other stakeholders to promote open science, for instance by enabling researchers to incorporate a certified service in their data stewardship plans.
- FORCE11 FAIR Data Management Plans Working Group: FORCE11, the international community/platform that hosted the open consultation for the definition of the 15 FAIR guiding principles in 2016, has established the "FAIR DMPs" Working Group aiming to provide a simple set of principles, along with examples of domain-specific implementations and recommendations for best practices, that emphasizes good data management, stewardship and machine-readability for making data FAIR.
- RDA/FORCE11 FAIR Sharing Working Group: a use cases-driven joint effort between RDA and Force11 to develop: a) a set of recommendations to guide users and producers of databases and content standards to select and describe them, or recommend them in data policies, and b) a curated registry, which enacts the recommendations and assists a variety of end users, providing well described, interlinked, and cross-searchable records on content standards, databases and data policies.
- CODATA: the Committee on Data of the International Council for Science (ICSU) that promotes global collaboration to improve the availability and usability of data for all areas of research.

- Science Europe: an association of European Research Funding Organisations (RFO) and Research Performing Organisations (RPO), active in the field of alignment of the research data management policies and templates.
- European Commission Expert Group on "Turning FAIR data into reality": established by the Commission, this expert group is working together with European and global initiatives towards a proposal for a FAIR Data Action Plan for consideration by the Commission, Member States and stakeholders in the research and data communities. The draft proposal presented by the Expert Group at the 2nd EOSC Summit on 11 June 2018, suggests the design of an agreed set of basic core FAIR metrics, which will be "standardised" and extendible in order to cover the needs and practises of different communities.
- EU-funded projects (e.g. EOSC pilot, EOSC hub, FREYA, Open AIRE Advanced etc.) supporting the first phase in the development of the European Open Science Cloud (EOSC):
- European Commission: DG RTD, DG CNECT, DG DIGIT and the Publications Office.

4. WORK PLAN

The Working Group "FAIR data maturity model: core criteria to assess the implementation level of the FAIR data principles" will build on top and combine the most salient characteristics of existing efforts for measuring the readiness and implementation level of a dataset vis-à-vis the FAIR data principles.

The outcomes of the Working Group (a common set of core assessment criteria for FAIRness in the form of a Recommendation and a generic and expandable self-assessment model for measuring the maturity level of a dataset FAIR) will be generic - and not specific to a certain discipline or country – and apply to any type of data in the conventional sense as well as to data-related algorithms, tools, workflows, protocols and other data-related services. They will be based on a core set of mutually exclusive and collectively exhaustive assessment criteria and be populated in a way that allows their extension in order to meet specific FAIR-related assessment needs, at national and/or discipline level (for example, for providing additional layers of detail for a number of discreet areas). Furthermore, design method will allow in the future the provision of estimations about the costs and benefits for organisations, both in economic and non-economic terms, for moving their datasets to a higher FAIR maturity level.

The outcomes will be developed following a progressive approach via a number of iterations. In each iteration, the current structure of the FAIR assessment criteria and the maturity model will be examined and validated in order to evolve to a revised version. The development process will be open, ensuring an active and continuous engagement of user communities and stakeholders in all development phases (including scoping, construction and testing). For that purpose, well-defined working and decision-making mechanisms will be defined and agreed from the beginning in order to facilitate the operation of the Working Group.

The main phases and deliverables will be the following:

Phase A: Initiation: during the first phase, the exact scope of the work will be defined including the objectives, the usage and the purpose of the assessment criteria and the model. Similar assessment criteria and models will be systematically analysed in order to identify components that could be reused either as they are or after applying some improvements, aiming to avoid the duplication of efforts.

Main outputs:

- Scope definition
- Literature review: an overview of existing approaches (generic or specific-purpose)

Timeline: M1 – M2

Phase B: Stakeholder identification: the initiation phase will be followed by the identification of the main actors who will be related with the outcomes of the Working Group from three perspectives: development process, execution and interest in the results.

Main outputs:

- Stakeholder matrix

Timeline: M3

Phase C: Design methodology: the Working Group will define and agree on a systematic, effective and efficient design methodology that will lead to results that are rigorous and both theoretically founded and empirically validated. The design methodology will follow an iterative approach, leveraging the most appropriate techniques for the population of the expected results. A special role should be foreseen for RDA Working Groups and Interest Group with pertinent objectives.

Main outputs:

- Design methodology

Timeline: M4 – M6

Phase D: Design: this phase will define all aspects with regard to the structure and the body of the FAIR assessment criteria and the model. The design phase will answer questions such as:

- How many different maturity stages will be foreseen?
- How many dimensions or layers will the model assess?
- Will be any documented maturation paths?
- How many questions will be included in the model?
- What will be the type of dependencies in the implementation of the foreseen model's capabilities or attributes (implicit / explicit)?
- Which techniques will be used for the population of the model (e.g. literature review, case study interviews, focus groups etc.)?
- Will be the measurement of the maturity quantitative and/or qualitative?

Main outputs:

- Core assessment criteria for FAIR
- FAIR data maturity model
- FAIR data checklist

Timeline: M7 – M16

Phase E: Testing: the assessment criteria and the model will be verified and validated following a well-defined evaluation methodology.

Main outputs:

- Testing results

Timeline: M13 – M16

- Delivery: when the main building blocks of the outcomes will be constructed, various characteristics regarding their distribution will be decided such as: what kinds of materials will be publicly available, in what format etc. The exact set of the outputs of this phase will be progressively decided by the members of the Working Group.

Timeline: M17 – M18

5. ADOPTION PLAN

The members of the Working Group will incorporate the outcomes of their work in their policies and practices that promote the implementation of the FAIR data principles at national, discipline and international level. The set of core assessment criteria for FAIRness will be used as the basis to examine the compatibility and alignment of their instruments (such as recommendations, frameworks, templates, toolsets etc.) and any corrective activities will be planned and implemented in a coherent way. In addition, all members will systematically promote the outcomes to their specific communities, aiming to raise awareness and support their real-life adoption.

Furthermore, the Working Group will create and publish a number of guidelines for the extension of the core assessment criteria and the FAIR maturity model by user communities and organisations with specific needs in the evaluation of the implementation of the FAIR data principles. That will allow and facilitate a wider adoption of the outcomes of the Working Group by existing and emerging initiatives.

6. INITIAL MEMBERSHIP

This is an initial membership, which gathers together representatives from organisations with experience in the area of FAIR data uptake, real-life implementation and compliance across borders/disciplines. The Working Group will progressively get in touch with disciplinary specific initiatives to get their input too.

Co-chairs:

- [Edit Herczog](#), Vision & Values SPRL
- [Vassilios Peristeras](#), International Hellenic University & KU Leuven
- Keith Russel, ARDC Australia

Members (currently confirmed list):

- Mercè Crosas (Co-chair in the FORCE11 WG on FAIR DMPs), Institute for Quantitative Social Science (IQSS), Harvard University
- Susanna-Assunta Sansone (co-chair in the RDA and Force11 FAIRsharing WG), Oxford e-Research Centre, University of Oxford
- Peter McQuilton (co-chair in the RDA and Force11 FAIRsharing WG), Oxford e-Research Centre, University of Oxford– confirmed
- Simon Hodson, CODATA
- Françoise Genova, CDS / ASTERICS
- Peter Wittenburg, IT Advisory board of Max Planck Society
- Natalie Harrower, Digital Repository of Ireland