

**EOSC** FAIR  
Executive Board Working Group

## **Recommendations on FAIR Metrics for EOSC** ***Second Draft for consultation***

Françoise Genova, Observatoire Astronomique de Strasbourg, ORCID [0000-0002-6318-5028](https://orcid.org/0000-0002-6318-5028)

Jan Magnus Aronsen, University of Oslo, Norway, [0000-0003-2593-1744](https://orcid.org/0000-0003-2593-1744)

Oya Beyan, Fraunhofer FIT, Germany, [0000-0001-7611-3501](https://orcid.org/0000-0001-7611-3501)

Natalie Harrower, Digital Repository of Ireland, [0000-0002-7487-4881](https://orcid.org/0000-0002-7487-4881)

András Holl, Hungarian Academy of Sciences, [0000-0002-6873-3425](https://orcid.org/0000-0002-6873-3425)

Rob W.W. Hooft, Dutch Techcentre for Life Sciences, [0000-0001-6825-9439](https://orcid.org/0000-0001-6825-9439)

Pedro Principe, University of Minho, [0000-0002-8588-4196](https://orcid.org/0000-0002-8588-4196)

Kostas Repanas, European Commission, [0000-0002-7848-2834](https://orcid.org/0000-0002-7848-2834)

Ana Slavec, InnoRenew CoE, [0000-0002-0171-2144](https://orcid.org/0000-0002-0171-2144)

Sarah Jones, GÉANT, [0000-0002-5094-7126](https://orcid.org/0000-0002-5094-7126)

Edited by Françoise Genova & Sarah Jones, Co-Chairs of the EOSC FAIR Working Group

# Executive Summary

The report contains an analysis of activities relevant to the definition of FAIR Metrics in the EOSC context. It makes recommendations on the definition and implementation of metrics, proposes a set of metrics for FAIR data in EOSC to be extensively tested, offers an analysis of gaps and potential opportunities for extension, and defines priorities for future work.

The report analyses recent and on-going activities relevant to the definition of FAIR metrics for data and other research objects at the European and international levels, in particular in the FAIRsFAIR project and RDA. It also discusses the maintenance of *Turning FAIR into reality* Recommendations and Action Plan and of the FAIR guiding principles themselves.

It offers seven recommendations on the definition and implementation of metrics:

**Recommendation 1:** The definition of metrics should be a continuous process, taking in particular feedback from implementation into account. We recommend that the metrics are reviewed after two years initially, then every three years.

**Recommendation 2:** Inclusiveness should be a key attribute.

Recommendation 2.1: Diversity, especially across communities, should be taken into account.

Priorities on criteria should be adopted to the specific case.

Recommendation 2.2: FAIR should be considered as a journey. Gradual implementation is to be expected in general, and evaluation tools should enable one to measure progress.

Recommendation 2.3: Resources should be able to interface with EOSC with a minimal overhead, and the data and functionalities available and used should remain.

**Recommendation 3:** Do not reinvent the wheel: the starting point for FAIR data metrics at this stage should be the criteria defined by the RDA FAIR Data Maturity Model Working Group. Groups defining metrics for their own use case should start with these criteria, and then liaise with the RDA Maintenance Group to provide feedback and participate in discussions on possible updates of the criteria and on priorities.

**Recommendation 4:** Evaluation methods and tools should be thoroughly assessed in a variety of contexts with broad consultation, in particular in different domains to ensure they scale and meet diverse community practices.

**Recommendation 5:** FAIR metrics should be developed for digital objects other than data, which may require that the FAIR guidelines be translated to suit these objects, particularly software source code.

**Recommendation 6:** Guidance should be provided from and to communities for evaluation and implementation.

**Recommendation 7:** Cross-domain usage should not be a requirement by default, “just in case”; it should be based on use-cases, and metrics should be carefully tailored in that respect.

The definition of Metrics for FAIR Data is the most advanced to date. The RDA FAIR Data Maturity Model Working Group has published a model with 41 criteria allowing one to assess compliance of data with the FAIR principles, each one with a degree of priority. The model was approved as a RDA Recommendation by the RDA Council. It is “intended as a first version of the model”. The RDA FAIR Data Maturity Model

provides an essential toolbox for FAIR data metrics that has to be examined with a critical eye when defining criteria to evaluate the FAIRness of data in a given context. The degree of priority of the criteria should in particular be adapted to the specific requirements.

With respect to the other elements of the FAIR ecosystem, activities on the definition of FAIR for software are converging in the RDA/FORCE11/ReSA Working Group FAIR for Research Software, which has begun to work in July 2020. Its aim is to define the FAIR principles for research software and provide guidelines on how to apply them. This WG should bring another key component to the FAIR ecosystem.

This document proposes a set of target metrics for FAIR data in EOSC, on a timeline displaying several successive steps. **The criteria should be seen as a draft of a target for EOSC metrics, which will be met progressively by EOSC components, and should be tested thoroughly. It is essential to examine their applicability and to gather feedback in a wide range of contexts, in particular disciplinary contexts.** Taking the need for progressiveness into account by a stepped process, some criteria will appear at a given point of the process and stay, others will be in the list only temporarily, as a stage towards a more mature status. This concerns criteria on knowledge representation for discovery, licences and community standards. **The proposed timeline should also be extensively assessed.**

We propose to update the proposed metrics first after two years, then every three years. This proposal can be adjusted depending on the situation, for instance if the RDA criteria evolve with time. We propose that an RDA Working Group is created in the framework of the RDA Global Open Research Commons Interest Group, which gathers EOSC and other similar initiatives from other regions, to seek international agreement. The Working Group recommendations should be ratified at EOSC level, for instance by a Working Group of the EOSC Stakeholder Forum. Engaging communities and gathering their feedback will be essential.

The analysis of the status of the definition of metrics, gaps, and potential opportunities for extension, leads to define priorities for future work:

**Priority 1:** Support the assessment and improvement of the RDA FAIR Data Maturity Model.

Priority 1.1: Support disciplinary communities to clarify their requirements with respect to FAIR and identify cross-community use cases.

Priority 1.2: Test the FAIR data maturity model in a wide range of communities, in a neutral forum and seeking for international agreement, to fine-tune and customise the recommendations and guidance, assess the degree of priorities, identify adverse consequences and apply corrections.

**Priority 2:** Assess and test the proposed EOSC FAIR data metrics in a neutral forum, which could be a Working Group set up by the RDA Global Open Research Commons Interest Group to seek global agreement with the international EOSC counterparts.

**Priority 3:** Support the definition of evaluation tools; their thorough assessment and evaluation including inclusiveness; comparison of tools (manual, automated); identification of their biases and applicability in many different contexts, including thematic ones.

**Priority 4:** Support the definition of FAIR for software and of the assessment framework for key elements of the FAIR ecosystem, in particular PID services and semantics.

**Priority 5:** Define and implement governance of the principles, assessment frameworks and metrics, adapted to each specific case.

**Priority 6:** Provide guidance for and support to implementation: support data and service providers to progress in the FAIRness of their holdings.

Success will be strongly dependent on take-up by communities, especially the thematic ones. Three among the *Six Recommendations for Implementation of FAIR Practices* (FP), which detail support which should be provided, are fully endorsed as necessary for the implementation and take-up of FAIR Metrics, namely:

**FP Recommendation 1:** Fund awareness-raising, training, education and community-specific support.

**FP Recommendation 2:** Fund development, adoption and maintenance of community standards, tools and infrastructures.

**FP Recommendation 3:** Incentivise development of community governance.

# Table of Contents

<b>Executive Summary</b>	<b>2</b>
<b>Table of Contents</b>	<b>5</b>
<b>Introduction</b>	<b>6</b>
<b>Activities relevant to the definition of FAIR metrics in the EOSC context and more generally</b>	<b>7</b>
2.1. Metrics for FAIR data	8
2.1.1. The RDA FAIR Data Maturity Model Working group	8
2.1.2. FAIRsFAIR Data Object Assessment Metrics	11
2.2. Metrics for other research objects	12
2.2.1. FAIR Research Software	13
2.2.2. FAIR Semantics	15
2.3. Assessment of Turning FAIR into reality Recommendations and Action Plan	16
2.4. Maintenance of the FAIR guiding principles	16
<b>Recommendations on the Definition and Implementation of Metrics</b>	<b>18</b>
3.1. Metrics definition as a continuous process	19
3.2. Inclusiveness	21
3.2.1. Taking diversity into account	21
3.2.2. Consider FAIR as a journey	22
3.3. Not reinventing the wheel	23
3.4. Evaluation methods	24
3.5. FAIR metrics for other digital objects	25
3.6. Providing guidelines	26
3.7. Cross-domain aspects	26
<b>Summary of the Status of FAIR Metrics, Draft Metrics for EOSC, Gaps and Priorities</b>	<b>26</b>
4.1. Summary of the Status of FAIR Metrics	26
4.2. Recommended FAIR Metrics for EOSC	27
4.3. Gaps and potential opportunities for extension	32
4.4. Priorities for future work	33

# 1. Introduction

The EOSC FAIR Working Group (WG) is tasked to specify a set of metrics assessing the FAIRness of datasets and other digital objects which should be applied within the European Open Science Cloud (EOSC). The FAIR WG created a Metrics and Certification Task Force to coordinate this work, together with the companion activities on Certification.

FAIR is a recent conceptual toolbox. The principles (Figure 1) are loose and ambiguous enough to lead to different interpretations. It is thus necessary to clarify their meaning, and define criteria to assess FAIRness. Also FAIR can be achieved only in an ecosystem, and the principles should be applied to the different components of the ecosystem, since FAIR data maturity depends on the capabilities of these components. Therefore, FAIR assessment should not only include the FAIR data as an output but also a specific assessment of the relevant criteria for each component, as it is the combination of data and services which enable FAIR. For example, data are assigned a globally unique persistent identifier (PID) and made retrievable via standard communication protocols via services such as repositories.

<p><i>To be Findable:</i></p> <ul style="list-style-type: none"><li>F1. (meta)data are assigned a globally unique and persistent identifier</li><li>F2. data are described with rich metadata (defined by R1 below)</li><li>F3. metadata clearly and explicitly include the identifier of the data it describes</li><li>F4. (meta)data are registered or indexed in a searchable resource</li></ul> <p><i>To be Accessible:</i></p> <ul style="list-style-type: none"><li>A1. (meta)data are retrievable by their identifier using a standardized communications protocol</li><li>A1.1. the protocol is free, open and universally implementable</li><li>A1.2. the protocol allows for an authentication and authorization procedure, where necessary</li><li>A2. metadata are accessible, even when the data are no longer available</li></ul> <p><i>To be Interoperable:</i></p> <ul style="list-style-type: none"><li>I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation</li><li>I2. (meta)data uses vocabularies that follow FAIR principles</li><li>I3. (meta)data include qualified references to other (meta)data</li></ul> <p><i>To be reusable:</i></p> <ul style="list-style-type: none"><li>R1. (meta)data are richly described with a plurality of accurate and relevant attributes</li><li>R1.1. (meta)data are released with a clear and accessible data usage license</li><li>R1.2. (meta)data are associated with data provenance</li><li>R1.3. (meta)data meet domain relevant community standards</li></ul>
--

**Figure 1:** The FAIR guiding principles from Wilkinson et al. (2016).<sup>1</sup>

Incentives and Metrics for FAIR data and services was the focus of a chapter in the *Turning FAIR into reality* Expert Group Report<sup>2</sup> (TFiR, 2018), with one priority recommendation and a set of actions to develop metrics for FAIR Digital Objects (Recommendation 12):

*A set of metrics for FAIR Digital Objects should be developed and implemented, starting from the basic common core of descriptive metadata, PIDs and access. The design of these metrics needs to be guided by research community practices and they should be regularly reviewed and updated.*

<sup>1</sup> Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3**, 160018 (2016). <https://doi.org/10.1038/sdata.2016.18>

<sup>2</sup> <https://doi.org/10.2777/1524>

*Action 12.1: A core set of metrics for FAIR Digital Objects should be defined to apply globally across research domains. More specific metrics should be defined at the community level to reflect the needs and practices of different domains and what it means to be FAIR for that type of research. Stakeholders: Coordination fora; Research communities.*

*Action 12.2: Convergence should be sought between the efforts by many groups to define FAIR assessment. The European Commission should support a project to coordinate activities in defining FAIR metrics and ensure these are created in a standardised way to enable future monitoring. Stakeholders: Coordination fora; Research communities; Funders; Publishers.*

*Action 12.3: The process of developing, approving and implementing FAIR metrics should follow a consultative methodology with research communities, including scenario planning to minimise any unintended consequences and counter-productive gaming that may result. Metrics need to be regularly reviewed and updated to ensure they remain fit-for-purpose. Stakeholders: Coordination fora; Research communities; Data service providers; Publishers.*

The remit of the EOSC FAIR WG is not to duplicate the relevant work done in the European and international contexts, in particular by initiatives and projects which are supported by the European Commission, but to bring this work together into a coherent direction with a critical look. Section 2 of this report will provide a summary of the relevant activities in the EOSC context and more generally. Section 3 presents recommendations on the definition and implementation of metrics, while Section 4 describes the status of FAIR Metrics, proposes a draft for target FAIR Metrics for EOSC, discusses gaps and potential opportunities for extension, and defines priorities for future activities.

## 2. Activities relevant to the definition of FAIR metrics in the EOSC context and more generally

Most of the activities on the definition of FAIR metrics have been dealing with FAIR data. Significant progress has been made recently by the RDA FAIR Data Maturity Model Working Group<sup>3</sup>, which produced recommendations and guidelines<sup>4</sup> approved by the RDA Council as an RDA Recommendation on 6 July 2020. Activities are also on-going in the FAIRsFAIR<sup>5</sup> project, which is playing a specific role to enable a FAIR ecosystem in the EOSC. One of them is to produce data objects assessment metrics, a version of which (V0.3) were published in July 2020<sup>6</sup>. These activities and their results are described in Section 2.1.

Whereas the original paper on the FAIR guiding principles clearly states that “the principles should apply not only to ‘data’ in the conventional sense, but also to the other research outputs such as algorithms, tools, and workflows that led to the data” “to ensure transparency, reproducibility, and reusability”, the focus when assessing FAIR practices and defining metrics for FAIR has clearly been until now on research data. The report of the FAIR in Practice Task Force of the EOSC FAIR Working Group, *Six Recommendations for Implementation of FAIR practices*<sup>7</sup>, analyses the published practices and work to define better guidance to make other Research objects FAIR in their own right (not only in their role of

---

<sup>3</sup> <https://www.rd-alliance.org/groups/fair-data-maturity-model-wg>

<sup>4</sup> <https://doi.org/10.15497/RDA00050>

<sup>5</sup> <https://www.fairsfair.eu/>

<sup>6</sup> <https://doi.org/10.5281/zenodo.3934401>. An updated version (V0.4) published in October 2020 is available at <https://doi.org/10.5281/zenodo.4081213>.

<sup>7</sup> [https://ec.europa.eu/info/publications/six-recommendations-implementation-fair-practice\\_en](https://ec.europa.eu/info/publications/six-recommendations-implementation-fair-practice_en)

supporting FAIR data), and provides the references of the relevant publications. Section 2.2.1 summarises these findings with additional comments.

The EOSC Interoperability Framework developed jointly by the EOSC FAIR and Architecture Working Groups addresses four interoperability layers: technical interoperability, semantic interoperability, organisational interoperability and legal interoperability, and produces a set of recommendations. They state that every semantic artefact that is being maintained in EOSC should be FAIR, which aligns with the FAIR guiding principle that “(meta)data use vocabularies that follow FAIR principles” (I2). FAIR Semantics is briefly discussed in Section 2.2.2 with reference to the work done by FAIRsFAIR Work Package 2 in the domain, related on-going discussions in the RDA, and a recent position paper *Coming to terms with FAIR Ontologies* by Poveda-Villalon et al.<sup>8</sup>.

The FAIR Guiding Principles for scientific data management and stewardship were published in 2016 by Wilkinson et al. *Turning FAIR into reality* was published in 2018, fewer than 2 years after the publication of the Wilkinson et al. paper. Its Recommendations and Action plan continue to provide key guidance. However, many activities that further refine the understanding of FAIR and paths towards implementation have been performed since then. It is useful to examine these activities and to assess whether additional actions should be recommended. This has been the aim of the Second Workshop<sup>9</sup> of the FAIRsFAIR pan-project Synchronisation Force<sup>10</sup>, which was held virtually during the Spring of 2020. The conclusions of the Workshop relevant to metrics and the input gathered by the EOSC FAIR WG during a session of the EOSC Consultation Day (18 May 2020), *Governance, maintenance and sustainability of metrics and PIDs*<sup>11</sup> are discussed in Section 2.3.

Finally, the FAIR Guiding Principles have had an enormous impact: FAIR is a powerful concept, and it is in particular central to the EOSC viewed as the “Web of FAIR Data and Related Services for Science”. The intense activity related to FAIR provides feedback on the FAIR principles themselves, and it is useful at this stage to assess whether there is a need to maintain these principles, and if so, how this could be achieved. This question was discussed during the EOSC FAIR WG session of the EOSC Consultation Day, and during the FAIRsFAIR Synchronisation Force workshop. Section 2.4 reports on the outcomes.

## 2.1. Metrics for FAIR data

### 2.1.1. The RDA FAIR Data Maturity Model Working group

As recommended in Action 12.2 of *Turning FAIR into reality*, convergence was sought between the efforts of many groups working to define FAIR assessment, with the support of the European Commission. Science data sharing and usage is by essence an international endeavour, and the RDA was the appropriate coordination forum to tackle the subject. The RDA FAIR Data Maturity Model Working Group built on existing initiatives to look at common elements and involved the international community to prepare a recommendation specifying a set of indicators and priorities for assessing adherence to the FAIR principles, with guidelines intended to assist evaluators to implement the indicators in the evaluation

---

<sup>8</sup> Poveda-Villalon, M, Espinoza-Arias, P, Garijo, D., Corcho, O. Coming to Terms with FAIR Ontologies - A position paper, EKAW Conference, 2020.

<sup>9</sup> <https://www.fairsfair.eu/events/fairsfair-2020-and-second-synchronisation-workshop>

<sup>10</sup> <https://www.fairsfair.eu/advisory-board/synchronisation-force>

<sup>11</sup>

<https://www.eosc-hub.eu/eosc-hub-week-2020/agenda/governance-maintenance-and-sustainability-metrics-and-pids>



approach or tool they manage. The Working Group was chaired by Edit Herczog (Belgium), Keith Russell (Australia) and Shelley Stall (USA), and the editorial team led by Makx Dekkers. It gathered over 200 experts from more than 20 countries over 18 months between January 2019 and June 2020.

The 41 FAIR data maturity model indicators as defined by the RDA Working Group<sup>12</sup> are shown in Figure 2. Three levels of importance are defined and shown in the table:

- Essential: such an indicator addresses an aspect that is of the utmost importance to achieve FAIRness under most circumstances, or, conversely, FAIRness would be practically impossible to achieve if the indicator were not satisfied.
- Important: such an indicator addresses an aspect that might not be of the utmost importance under specific circumstances, but its satisfaction, if at all possible, would substantially increase FAIRness.
- Useful: such an indicator addresses an aspect that is nice-to-have but is not necessarily indispensable.

FAIR	ID	Indicator	Priority
F1	<a href="#">RDA-F1-01M</a>	Metadata is identified by a persistent identifier	●●● Essential
F1	<a href="#">RDA-F1-01D</a>	Data is identified by a persistent identifier	●●● Essential
F1	<a href="#">RDA-F1-02M</a>	Metadata is identified by a globally unique identifier	●●● Essential
F1	<a href="#">RDA-F1-02D</a>	Data is identified by a globally unique identifier	●●● Essential
F2	<a href="#">RDA-F2-01M</a>	Rich metadata is provided to allow discovery	●●● Essential
F3	<a href="#">RDA-F3-01M</a>	Metadata includes the identifier for the data	●●● Essential
F4	<a href="#">RDA-F4-01M</a>	Metadata is offered in such a way that it can be harvested and indexed	●●● Essential
A1	<a href="#">RDA-A1-01M</a>	Metadata contains information to enable the user to get access to the data	●● Important
A1	<a href="#">RDA-A1-02M</a>	Metadata can be accessed manually (i.e. with human intervention)	●●● Essential
A1	<a href="#">RDA-A1-02D</a>	Data can be accessed manually (i.e. with human intervention)	●●● Essential
A1	<a href="#">RDA-A1-03M</a>	Metadata identifier resolves to a metadata record	●●● Essential
A1	<a href="#">RDA-A1-03D</a>	Data identifier resolves to a digital object	●●● Essential
A1	<a href="#">RDA-A1-04M</a>	Metadata is accessed through standardised protocol	●●● Essential
A1	<a href="#">RDA-A1-04D</a>	Data is accessible through standardised protocol	●●● Essential
A1	<a href="#">RDA-A1-05D</a>	Data can be accessed automatically (i.e. by a computer program)	●● Important
A1.1	<a href="#">RDA-A1.1-01M</a>	Metadata is accessible through a free access protocol	●●● Essential
A1.1	<a href="#">RDA-A1.1-01D</a>	Data is accessible through a free access protocol	●● Important
A1.2	<a href="#">RDA-A1.2-01D</a>	Data is accessible through an access protocol that supports authentication and authorisation	● Useful
A2	<a href="#">RDA-A2-01M</a>	Metadata is guaranteed to remain available after data is no longer available	●●● Essential
I1	<a href="#">RDA-I1-01M</a>	Metadata uses knowledge representation expressed in standardised format	●● Important
I1	<a href="#">RDA-I1-01D</a>	Data uses knowledge representation expressed in standardised format	●● Important
I1	<a href="#">RDA-I1-02M</a>	Metadata uses machine-understandable knowledge representation	●● Important
I1	<a href="#">RDA-I1-02D</a>	Data uses machine-understandable knowledge representation	●● Important
I2	<a href="#">RDA-I2-01M</a>	Metadata uses FAIR-compliant vocabularies	●● Important
I2	<a href="#">RDA-I2-01D</a>	Data uses FAIR-compliant vocabularies	● Useful
I3	<a href="#">RDA-I3-01M</a>	Metadata includes references to other metadata	●● Important
I3	<a href="#">RDA-I3-01D</a>	Data includes references to other data	● Useful
I3	<a href="#">RDA-I3-02M</a>	Metadata includes references to other data	● Useful

<sup>12</sup> <https://doi.org/10.15497/RDA00050>

FAIR	ID	Indicator	Priority
I3	<a href="#">RDA-I3-02D</a>	Data includes qualified references to other data	● Useful
I3	<a href="#">RDA-I3-03M</a>	Metadata includes qualified references to other metadata	●● Important
I3	<a href="#">RDA-I3-04M</a>	Metadata include qualified references to other data	● Useful
R1	<a href="#">RDA-R1-01M</a>	Plurality of accurate and relevant attributes are provided to allow reuse	●●● Essential
R1.1	<a href="#">RDA-R1.1-01M</a>	Metadata includes information about the licence under which the data can be reused	●●● Essential
R1.1	<a href="#">RDA-R1.1-02M</a>	Metadata refers to a standard reuse licence	●● Important
R1.1	<a href="#">RDA-R1.1-03M</a>	Metadata refers to a machine-understandable reuse licence	●● Important
R1.2	<a href="#">RDA-R1.2-01M</a>	Metadata includes provenance information according to community-specific standards	●● Important
R1.2	<a href="#">RDA-R1.2-02M</a>	Metadata includes provenance information according to a cross-community language	● Useful
R1.3	<a href="#">RDA-R1.3-01M</a>	Metadata complies with a community standard	●●● Essential
R1.3	<a href="#">RDA-R1.3-01D</a>	Data complies with a community standard	●●● Essential
R1.3	<a href="#">RDA-R1.3-02M</a>	Metadata is expressed in compliance with a machine-understandable community standard	●●● Essential
R1.3	<a href="#">RDA-R1.3-02D</a>	Data is expressed in compliance with a machine-understandable community standard	●● Important

**Figure 2:** FAIR Data maturity model indicators v.1.0 as defined by the RDA Working Group

The document also discusses evaluation methods with two different perspectives: measuring progress and pass-or-fail.

Priority	Principle				
	Findable	Accessible	Interoperable	Reusable	Grand Total
Essential	7	8	0	5	20
Important	0	3	7	4	14
Useful	0	1	5	1	7
<b>Grand Total</b>	<b>7</b>	<b>12</b>	<b>12</b>	<b>10</b>	<b>41</b>

**Figure 3:** Distribution of priorities per FAIR area as defined by the RDA Working Group

The distribution of priorities per FAIR area is shown in Figure 3. One striking feature is that the distribution of priorities is very uneven across the four elements of FAIR: all the criteria for *Findable* are deemed 'essential', for *Accessible*, it is two-thirds of the criteria, half of the criteria for *Reusable*, and none of the criteria for *Interoperable*. This is a result of the WG participant input and open consultation process, and likely illustrates the fact that many communities are not yet ready to implement interoperability.

This result, however, is problematic for EOSC which aims to enable the federation of data and services across geographic and disciplinary boundaries. Interoperability is thus a critical requirement for EOSC; as a result, the EOSC FAIR Working Group is developing an EOSC Interoperability Framework<sup>13</sup> in collaboration with the Architecture Working Group. Indeed interoperability is a central requirement for some research communities as well, for instance astronomy, which may find that some of the priorities tagged as Essential are not actually essential to their community needs. As EOSC needs to work for all

<sup>13</sup> Draft for consultation: <https://www.eoscsecretariat.eu/sites/default/files/eosc-interoperability-framework-v1.0.pdf>

communities and prevent any barriers to uptake, we need to ensure that the criteria set are globally applicable, which requires compromise and taking as broad a view as possible.

The criteria defined at the international level by the RDA Working Group are a key contribution to the definition of core metrics for FAIR data. At this stage, they should be considered as a toolbox, in particular when determining the priority levels attached to the criteria. As stated in the RDA recommendation, the exact way to evaluate data based on the core criteria is up to the owner of the evaluation approaches, taking into account the requirements of their community. An example is described in Section 2.2.2.

The RDA Working Group did not revisit the FAIR principles themselves, but feedback from the participants illustrated some issues with the principles with respect to well established community practices, for instance the fact that in the FAIR principles data and metadata are separated, whereas many communities integrate data and metadata together and attach a unique PID to them, or store metadata at different levels, with some metadata available for instance at the collection level and others at the level of the data object.

Following the publication of the RDA Recommendation, the RDA FAIR Data Maturity Model Working Group has converted into a Maintenance Working Group with the objective to maintain and further develop the Maturity Model. The Maintenance Working Group will work on promoting and improving the FAIR Data Maturity Model, and more generally FAIR assessments. This encompasses (1) the establishment of formal or informal liaisons between FAIR assessment activities, (2) gathering feedback from the implementation of the FAIR data maturity model within thematic communities and (3) reaching an agreement on future work on FAIR assessments (e.g. integration of the FAIR Data Maturity Model in Data Management Plans).

### 2.1.2. FAIRsFAIR Data Object Assessment Metrics

At the European level, the FAIRsFAIR<sup>14</sup> project aims to supply practical solutions for the use of the FAIR data principles throughout the research data life cycle. Its Work Package 4 *FAIR Certification (of repositories)*<sup>15</sup> has a task (Task 4.5) which develops pilots for FAIR data assessment, with two primary use cases linked to the data management by a repository: a manual self-assessment tool offered by a trustworthy data repository to educate and raise awareness of researchers on making their data FAIR before depositing the data in the repository; and automated assessment of dataset FAIRness by a trustworthy repository<sup>16</sup>. The metrics are developed in stages, taking into account the outcome of the RDA FAIR Data Maturity Model Working Group, prior work from project partners, the WDS/RDA Assessment of Data Fitness for Use checklist<sup>17</sup>, and feedback from stakeholders interested in FAIR. The metrics will be used in tools adapted to different use cases.

July 2020 (V0.3) version of the metrics (*FAIRsFAIR Data Objects Assessment Metrics V0.3*) is shown in Figure 4. It includes 15 criteria<sup>18</sup> which address nearly all the FAIR guiding principles.

---

<sup>14</sup> <https://www.fairsfair.eu/>

<sup>15</sup> <https://www.fairsfair.eu/fair-certification>

<sup>16</sup> FAIRsFAIR D4.1 Draft recommendations on requirements for FAIR datasets in certified repositories  
<https://doi.org/10.5281/zenodo.3678716>

<sup>17</sup> <https://www.rd-alliance.org/group/wdsrda-assessment-data-fitness-use-wg/outcomes/wdsrda-assessment-data-fitness-use-wg-outputs>

<sup>18</sup> V0.4 has 17 metrics.

Identifier	Name
<a href="#">FsF-F1-01D</a>	<b>Data</b> is assigned a globally unique identifier.
<a href="#">FsF-F1-02D</a>	<b>Data</b> is assigned a persistent identifier.
<a href="#">FsF-F2-01M</a>	<b>Metadata</b> includes descriptive core elements (creator, title, data identifier, publisher, publication date, summary and keywords) to support data findability.
<a href="#">FsF-F3-01M</a>	<b>Metadata</b> includes the identifier of the data it describes.
<a href="#">FsF-F4-01M</a>	<b>Metadata</b> is offered in such a way that it can be retrieved by machines.
<a href="#">FsF-A1-01M</a>	<b>Metadata</b> contains access level and access conditions of the data.
<a href="#">FsF-A2-01M</a>	<b>Metadata</b> remains available, even if the data is no longer available.
<a href="#">FsF-I1-01M</a>	<b>Metadata</b> is represented using a formal knowledge representation language.
<a href="#">FsF-I1-02M</a>	<b>Metadata</b> uses semantic resources.
<a href="#">FsF-I3-01M</a>	<b>Metadata</b> includes links between the data and its related entities.
<a href="#">FsF-R1-01MD</a>	<b>Metadata</b> specifies the content of the data.
<a href="#">FsF-R1.1-01M</a>	<b>Metadata</b> includes license information under which data can be reused.
<a href="#">FsF-R1.2-01M</a>	<b>Metadata</b> includes provenance information about data creation or generation.
<a href="#">FsF-R1.3-01M</a>	<b>Metadata</b> follows a standard recommended by the target research community of the data.
<a href="#">FsF-R1.3-02D</a>	<b>Data</b> is available in a file format recommended by the target research community.

**Figure 4:** List of metrics for the assessment of FAIR data objects developed by FAIRsFAIR Task 4.4 (V0.3, July 2020)

FAIRsFAIR Data Object Assessment Metrics is still a work in progress, but it shows how the core criteria can be used to fit specific needs. The work illustrates a use-case driven approach to FAIR metrics, and the need for an iterative approach gathering feedback from the stakeholders for the specific usage of the criteria.

## 2.2. Metrics for other research objects

Most of the published practice, guidance and policy on other research objects concerns software, workflows and computational (executable) notebooks. As explained in the *Six Recommendations for Implementation of FAIR Practices, Turning FAIR into reality* also advocates that DMPs should be FAIR outputs in their own right (Recommendation 16: *Apply FAIR broadly*). Making DMPs ‘machine-actionable’ means making their content findable and accessible, exchanging that content with other systems in



standardised, interoperable ways, and potentially reusing that content. A standard for exchanging DMP content<sup>19</sup> has demonstrated the effective exchange of DMP data across several connected platforms.<sup>20</sup>

### 2.2.1. FAIR Research Software

*The text of this section is mostly copied from or inspired by the Six Recommendations for Implementation of FAIR Practices.*

Most of the published work<sup>21,22,23,24</sup> on FAIR suggests that whilst the FAIR foundational principles can apply to software, the guiding principles require translation for this purpose, though how much translation is still unclear. The paper *Towards FAIR principles for research software* (Lamprecht et al., 2019)<sup>25</sup> reviews previous work on applying the FAIR principles to software and suggests ways of adapting the principles to a software context. The principles proposed by the paper, and how they relate to the FAIR guiding principles for research data, are summarised in the paper's Table 1 (Figure 5).

---

<sup>19</sup> Walk, P., Miksa, T., & Neish, P. (2019). RDA DMP Common Standard for Machine-actionable Data Management Plans. Research Data Alliance. <https://doi.org/10.15497/RDA00039>

<sup>20</sup> <https://rda-dmp-common.github.io/hackathon-2020/>

<sup>21</sup> Chue Hong, N., & Katz, D. S. (2018). FAIR enough? Can we (already) benefit from applying the FAIR data principles to software? <https://doi.org/10.6084/M9.FIGSHARE.7449239.V2>

<sup>22</sup> Erdmann, C., Simons, N., Otsuji, R., Labou, S., Johnson, R., Castelao, G., Boas, B. V., Lamprecht, A.-L., Ortiz, C. M., Garcia, L., Kuzak, M., Martinez, P. A., Stokes, L., Honeyman, T., Wise, S., Quan, J., Peterson, S., Neeser, A., Karvovskaya, L., ... Dennis, T. (2019). Top 10 FAIR Data & Software Things. <https://doi.org/10.5281/ZENODO.2555498>

<sup>23</sup> Aerts, P. J. C. (2017). Sustainable Software Sustainability - Workshop report. Data Archiving and Networked Services (DANS). <https://doi.org/10.17026/DANS-XFE-RN2W>

<sup>24</sup> Doorn, P. (2017). Does it make sense to apply the FAIR Data Principles to Software? [https://indico.cern.ch/event/588219/contributions/2384979/attachments/1426152/2189855/FAIR\\_Software\\_Principles\\_CERN\\_March\\_2017.pdf](https://indico.cern.ch/event/588219/contributions/2384979/attachments/1426152/2189855/FAIR_Software_Principles_CERN_March_2017.pdf)

<sup>25</sup> Lamprecht, A.-L., Garcia, L., Kuzak, M., Martinez, C., Arcila, R., Martin Del Pico, E., Dominguez Del Angel, V., van de Sandt, S., Ison, J., Martinez, P. A., McQuilton, P., Valencia, A., Harrow, J., Psomopoulos, F., Gelpi, J. L., Chue Hong, N., Goble, C., & Capella-Gutierrez, S. (2020). Towards FAIR principles for research software. *Data Science*, 3(1), 37–59. <https://doi.org/10.3233/DS-190026>

Table 1

Summary of the proposed FAIR principles for research software and how they relate to the FAIR Guiding Principles for data. It is indicated whether a given FAIR data principle has been simply rephrased to adjust it to software, extended to cover a broader scope, reinterpreted to match the different context, discarded as it does not apply, or newly proposed as it only applies for research software

	FAIR for data	FAIR for software	Operation
F1	(Meta)data are assigned a globally unique and persistent identifier.	Software and its associated metadata have a global, unique and persistent identifier for each released version.	Rephrased
F2	Data are described with rich metadata.	Software is described with rich metadata.	Rephrased
F3	Metadata clearly and explicitly include the identifier of the data it describes.	Metadata clearly and explicitly include identifiers for all the versions of the software it describes.	Rephrased and extended
F4	(Meta)data are registered or indexed in a searchable resource.	Software and its associated metadata are included in a searchable software registry.	Rephrased
A1	(Meta)data are retrievable by their identifier using a standardized communications protocol.	Software and its associated metadata are accessible by their identifier using a standardized communications protocol.	Rephrased
A1.1	The protocol is open, free, and universally implementable.	The protocol is open, free, and universally implementable.	Remain the same
A1.2	The protocol allows for an authentication and authorization procedure, where necessary.	The protocol allows for an authentication and authorization procedure, where necessary.	Remain the same
A2	Metadata are accessible, even when the data are no longer available.	Software metadata are accessible, even when the software is no longer available.	Rephrased
I1	(Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.	Software and its associated metadata use a formal, accessible, shared and broadly applicable language to facilitate machine readability and data exchange.	Rephrased and extended
I2	(Meta)data use vocabularies that follow FAIR principles.	–	Reinterpreted, extended and split
I2S.1	–	Software and its associated metadata are formally described using controlled vocabularies that follow the FAIR principles.	Reinterpreted, extended and split
I2S.2	–	Software use and produce data in types and formats that are formally described using controlled vocabularies that follow the FAIR principles.	Reinterpreted, extended and split
I3	(Meta)data include qualified references to other (meta)data.	–	Discarded
I4S	–	Software dependencies are documented and mechanisms to access them exist.	Newly proposed
R1	(Meta)data are richly described with a plurality of accurate and relevant attributes.	Software and its associated metadata are richly described with a plurality of accurate and relevant attributes.	Rephrased
R1.1	(Meta)data are released with a clear and accessible data usage license.	Software and its associated metadata have independent, clear and accessible usage licenses compatible with the software dependencies.	Rephrased and extended
R1.2	(Meta)data are associated with detailed provenance.	Software metadata include detailed provenance, detail level should be community agreed.	Rephrased
R1.3	(Meta)data meet domain-relevant community standards.	Software metadata and documentation meet domain-relevant community standards.	Rephrased

**Figure 5:** Table 1 from *Towards FAIR principles for research software* (Lamprecht et al., 2019)

Most current publications on FAIR workflows suggest policies and processes to improve the FAIRness of workflows.<sup>26,27</sup> A common theme is that the same challenges faced when attempting to apply the FAIR guiding principles to software apply to workflows and executable notebooks; their characteristics mean that they are similar to software artefacts. Another challenge for workflows is that automated annotation and description strategies and tools are required because the burden of creating and maintaining metadata for workflows is much higher than for data. Considerable progress has been made on tooling and services to help make executable notebooks findable, accessible and reusable, by providing DOIs to identify them, reproducible environments to run them (Binder<sup>28</sup>, CodeOcean<sup>29</sup>) or to export them to other publishing formats. This has been supported by documentation and training that has aided adoption.

Metrics for FAIR software, as currently proposed, combine metrics based on FAIR data metrics with metrics based on software quality metrics. This will need to be clarified, in particular to identify which metrics will best help adoption of FAIR for software. In 2020, a joint RDA/FORCE11/ReSA working group has been setup on FAIR for Research Software (FAIR4RS)<sup>30</sup>. The Working Group has begun the work of reviewing and, if necessary, redefining FAIR guiding principles for software and related computational code-based research objects. We expect this to be the community forum for taking forward the FAIR principles for software, services and workflows. FAIRsFAIR Task 2.4 on FAIR Services added software to their initial remit. They will publish an assessment report for FAIR software in October 2020, which will be primarily a meta-analysis of the literature on FAIR software, which lays the foundation for further work under the umbrella of the RDA FAIR for Research Software Working Group.

### 2.2.2. FAIR Semantics

Task 2.2 of FAIRsFAIR on FAIR Semantics aims at improving the semantic interoperability of research resources by specifying FAIR metadata schemas, vocabularies, protocols, and ontologies. They produced a first set of recommendations<sup>31</sup> from their internal work and their initial discussions with the community of semantics experts during the workshop *Building the data landscape of the future: FAIR Semantics and FAIR Repositories*, co-located with the 14<sup>th</sup> RDA Plenary meeting in Helsinki (22 October 2019). The document contains a set of 17 preliminary recommendations aligned with individual FAIR principles and 10 best practice recommendations beyond the FAIR principles to improve the FAIRness of semantic artefacts. It was developed to trigger further discussion with the semantic expert community.

Communities are developing semantic artefacts, and the discussion should also include them since the recommendations may have a strong impact on existing practices which fulfil the community requirements. An appropriate path in this direction was the discussion of the document during the session *Moving towards FAIR Semantics*<sup>32</sup> of the Vocabulary Services Interest Group at the RDA 15<sup>th</sup> Plenary. FAIRsFAIR then organised two workshops on semantics artefact metadata on 29 April 2020 and 15 October 2020. The discussion will continue at RDA 16<sup>th</sup> Plenary in November 2020.

---

<sup>26</sup> Weigel, T., Schwarzmann, U., Klump, J., Bendoukha, S., & Quick, R. (2020). Making Data and Workflows Findable for Machines. *Data Intelligence*, 2(1–2), 40–46. [https://doi.org/10.1162/dint\\_a\\_00026](https://doi.org/10.1162/dint_a_00026)

<sup>27</sup> Goble, C., Cohen-Boulakia, S., Soiland-Reyes, S., Garijo, D., Gil, Y., Crusoe, M. R., ... Schober, D. (2020). FAIR Computational Workflows. *Data Intelligence*, 2(1–2), 108–121. [https://doi.org/10.1162/dint\\_a\\_00033](https://doi.org/10.1162/dint_a_00033)

<sup>28</sup> <https://mybinder.org/>

<sup>29</sup> <https://codeocean.com/>

<sup>30</sup> <https://www.rd-alliance.org/groups/fair-4-research-software-fair4rs-wg>

<sup>31</sup> FAIRsFAIR D2.2 FAIR Semantics: First recommendations <https://doi.org/10.5281/zenodo.3707985>

<sup>32</sup> <https://www.rd-alliance.org/plenaries/rda-15th-plenary-meeting-australia/moving-toward-fair-semantics>

Poveda-Villalon et al (2020)<sup>33</sup> compare the FAIRsFAIR approach with the 5-star schemes for publishing Linked Open Data and analyse their relationship to the FAIR principles. They advocate the need to involve the Semantic Web community in the conversation about FAIR Semantics.

### 2.3. Assessment of *Turning FAIR into reality* Recommendations and Action Plan

FAIRsFAIR has established a pan-project Synchronisation Force which liaises with its “European Group of FAIR Champions”<sup>34</sup>, the five ESFRI Clusters and the so-called ‘5b’ projects, the thematic and regional EOSC projects. In the framework of the Collaboration Agreement between FAIRsFAIR and EOSC Secretariat, the Synchronisation Force provides input for the EOSC Executive Board Working Groups, including the FAIR Working Group.

The second FAIRsFAIR Synchronisation Force Workshop was organised on-line as a series of eight sessions from April 29th to June 11th, 2020. Representatives from the EOSC Executive Board Working Groups, from the EOSC Clusters and ‘5b’ projects, and the members of the European Group of FAIR Champions, were invited to attend. The EOSC FAIR Working Group participated actively in the workshop. The workshop objectives were to measure the progress towards implementing the recommendations outlined in *Turning FAIR into reality*, and also to identify gaps in its Action Plan and propose additional actions. The workshop report<sup>35</sup> summarises the findings. TFiR Recommendation 12 *Develop metrics for FAIR digital outputs* was examined. It is proposed to add the following element: *developing a governance process for the maintenance and revision of metrics and associated assessment processes is important.*

The EOSC FAIR WG had organised a session about this issue during the EOSC Consultation Day on 18 May 2020, *Governance, maintenance and sustainability of metrics and PIDs*. The audience was polled about the need and method to maintain and sustain the recommendations, which provided useful input on the frequency of maintenance, the stakeholders to be consulted, the kind of information to be gathered and the best ways to gather it, and the bodies which should be in charge of the maintenance and their desirable characteristics. This input is taken into account in our recommendations and discussed in Section 3.1.

### 2.4. Maintenance of the FAIR guiding principles

The EOSC FAIR WG session of the EOSC Consultation Day also addressed the eventual need to govern the FAIR guiding principles, and who should maintain them (Figure 6).

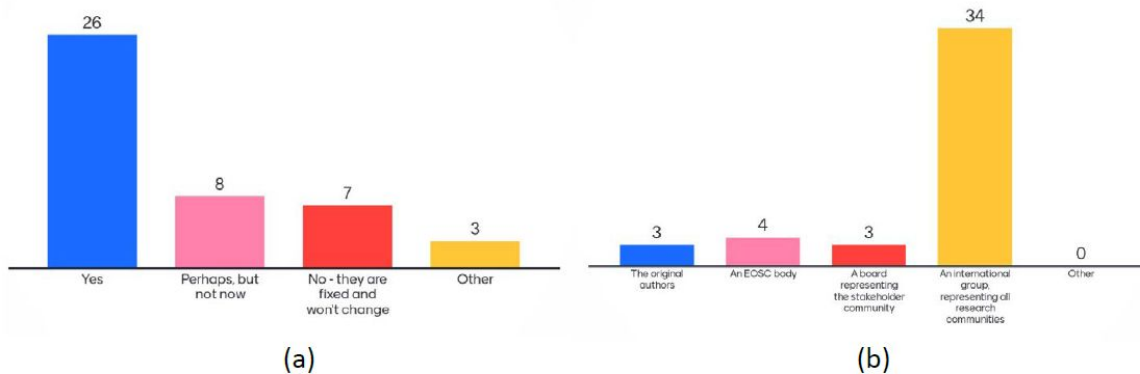
---

<sup>33</sup> [https://www.researchgate.net/publication/344042645\\_Coming\\_to\\_Terms\\_with\\_FAIR\\_Ontologies](https://www.researchgate.net/publication/344042645_Coming_to_Terms_with_FAIR_Ontologies)

<sup>34</sup> <https://www.fairsfair.eu/advisory-board/egfc>

<sup>35</sup> Second Report of the FAIRsFAIR Synchronisation Force (D5.5) <https://doi.org/10.5281/zenodo.3953979>





**Figure 6:** Results of the Mentimeter polls performed during the EOSC FAIR WG session of the EOSC Consultation Day on (a) Do the FAIR principles need to be governed; (b) Who should maintain them.

60% of the 44 participants agreed that the principles should be maintained, 18% that they should but not now, 16% stated that they are fixed and should not change. A significant majority of the participants thus agree that the FAIR principles should be maintained at some point.

When asked about how the maintenance should be performed, 77% of the participants considered that an international group representing all research communities should be in charge, with less than 10% of the participants in favour of each of the other proposed solutions, the original authors, an EOSC body, or a board representing the stakeholder community.

The question was also briefly raised during the FAIRsFAIR Synchronisation Force Workshop, based on concerns arising from some of the discussions held during the course of the work of the RDA FAIR Data Maturity Model Working Group. The issue is discussed in some details in the workshop report, which concludes: “Ultimately, it is not our recommendation that a governance process should be established to update and modify the FAIR guiding principles. They are what they are and, in any case, should serve as a guide, not as articles of faith from which nothing can deviate in any circumstances. Metrics for EOSC, for repositories and the FAIR ecosystem, should not necessarily feel bound to follow a strict interpretation of the existing FAIR principles. If there are issues which for pragmatic reasons or for accepted practice do not follow the apparent letter of the principles as published, then resultant guidelines and metrics - and the TFiR Action Plan or subsequent document - should be adjusted in a reasoned and transparent way. Effort would be better expended ensuring that metrics and certification are implemented with appropriate judgement, with transparency and feedback, and do not do an unnecessary disservice to good established practice.”

Putting things together one concludes that, at least for the moment, the FAIR principles should stay as they are, but a critical eye has to be kept on the way they are applied, with the possible need to identify or establish an international group representing in particular a large palette of research communities to maintain them if issues which cannot be solved pragmatically as proposed in the FAIRsFAIR report are identified. The point here is not to question Findability, Accessibility, Interoperability and Reuse as essential requirements and aims, but to assess the 15 guiding principles from Wilkinson et al. (2016).

### 3. Recommendations on the Definition and Implementation of Metrics

The definition of metrics is a difficult process, which, if not executed with sufficient care, risks causing unintended negative consequences for EOSC.

#### *The key role of research communities*

What is clear from the projects and consultations noted above is that research communities -- led by disciplines or groupings of disciplines -- need to define the specifics of what implementing FAIR means in their community, and in tandem, how to determine suitable metrics for that implementation.

There is no global research community with common file format recommendations, metadata standards, there is no formal knowledge representation commonly used across all fields of science, and only a few semantic resources are commonly usable for all areas of scholarly research. No such solutions exist, because the problem is field-specific and the research community is not uniform, but rather divided into different communities. This involves different customs and standards, and the communities are at different levels of preparedness or of progress, with some communities completely lacking such resources.

Community standards are central to FAIR. There must be agreed formats for data, common vocabularies, metadata standards and accepted procedures for how, when and where data will be shared. Research communities need to be supported to come together to define these practices and standards. Some have done so, but many lack the resources as this work is often undervalued and not rewarded. If they do not invest in the definition of standards where these are lacking, then some communities will be unable to fully engage in the Web of FAIR data. Levelling the playing field to enable broader cross-disciplinary research is a priority.

This is not to say that FAIR metrics are entirely specific to disciplines, but rather that disciplinary differences are real and must be used to shape best practice, and to create a FAIR environment that truly serves the discipline, does not hinder existing capacities, and is supported by the community, which brings key data and service providers and system users. For this reason, implementation of metrics requires thorough consultation. It will remain an ongoing process that must adapt over time.

#### *Recommendations for managing and mitigating risks linked to the implementation of metrics*

The risks and unintended consequences of the implementation of metrics are well understood by the community, and its concerns were strongly and unequivocally expressed during the consultation on the EOSC SRIA held during the Summer of 2020: Metrics and Certification are given a low priority in the survey of relevance of action areas, ranking only second-to-last with 39% of votes compared to 78% for the highest ranked priority, metadata and ontologies, followed by identifiers at 72%.

The FAIR Working Group adopts a cautious attitude with respect to FAIR metrics, understanding this requirement for caution to be fully aligned with community concerns. FAIRness should not be used to judge unfairly; the development of maturity over time and across communities has to be supported and binary judgements avoided. Metrics are not meant to be a punitive method for direct comparison between datasets of different areas, and between different communities, because communities will arrive at optimal FAIRness in different ways: a community should first set out what they aim to do with FAIRness,

why certain metrics are to be targeted and others not targeted, with a rough cost-benefit analysis of each criterion, define and manage implementation strategies, and then assess evolution over time with maturity.

The following recommendations enable the definition of implementation of metrics and minimise the risks.

**Recommendation 1:** The definition of metrics should be a continuous process, taking in particular feedback from implementation into account. We recommend that the metrics are reviewed after two years initially, then every three years.

**Recommendation 2:** Inclusiveness should be a key attribute.

**Recommendation 2.1:** Diversity, especially across communities, should be taken into account. Priorities on criteria should be adopted to the specific case.

**Recommendation 2.2:** FAIR should be considered as a journey. Gradual implementation is to be expected in general, and evaluation tools should enable one to measure progress.

**Recommendation 2.3:** Resources should be able to interface with EOSC with a minimal overhead, and the data and functionalities available and used should remain.

**Recommendation 3:** Do not reinvent the wheel: the starting point for FAIR data metrics at this stage should be the criteria defined by the RDA FAIR Data Maturity Model Working Group. Groups defining metrics for their own use case should start with these criteria, and then liaise with the RDA Maintenance Group to provide feedback and participate in discussions on possible updates of the criteria and on priorities.

**Recommendation 4:** Evaluation methods and tools should be thoroughly assessed in a variety of contexts with broad consultation, in particular in different domains to ensure they scale and meet diverse community practices.

**Recommendation 5:** FAIR metrics should be developed for digital objects other than data, which may require that the FAIR guidelines be translated to suit these objects, particularly software source code.

**Recommendation 6:** Guidance should be provided from and to communities for evaluation and implementation.

**Recommendation 7:** Cross-domain usage should not be a requirement by default, “just in case”; it should be based on use-cases, and metrics should be carefully tailored in that respect.

The raison d’être of the recommendations is discussed below.

### 3.1. Metrics definition as a continuous process

The metrics currently proposed are not a final canonical set of metrics - continual monitoring and updating is required. The RDA FAIR Data Maturity Model recommendation concludes on future maintenance, stating that “the FAIR Maturity Model described in this document is intended as a first final version of the model”, and that “the model should be further developed taking into account comments and contributions of a wide range of stakeholders”. The Working Group is now in Maintenance Mode, which is defined by RDA as having the purpose of managing the maintenance activities and supporting the adopters of the original recommendation. FAIRsFAIR started in March 2019 for 36 months, and will thus continue to

refine their findings after the end of this EOSC Working Group. As already explained, the definition of FAIR for software and semantics is just beginning.

Beyond the issues with the relevant project and activity planning with respect to the FAIR WG time, one has to remember the warning from Action 12.3 of *Turning FAIR into reality*, “Metrics need to be regularly reviewed and updated to ensure they remain fit-for-purpose.” Metrics will have to be maintained beyond initial feedback, and also beyond the FAIRsFAIR project, and how this will be done remains to be defined. For instance, CoreTrustSeal launched a thorough consultation about their criteria and guidance for certification of trustworthy digital repositories in 2019, after two years of operations. The community review led to a significant update of their guidance document, although the criteria remain the same<sup>36</sup>.

Based on community consultation during our session on governance of the EOSC Consultation Day, we propose that any metrics should be reviewed, first after two years, and then every three years, by a neutral governance body which we propose to be an RDA Group with EOSC and international participation.

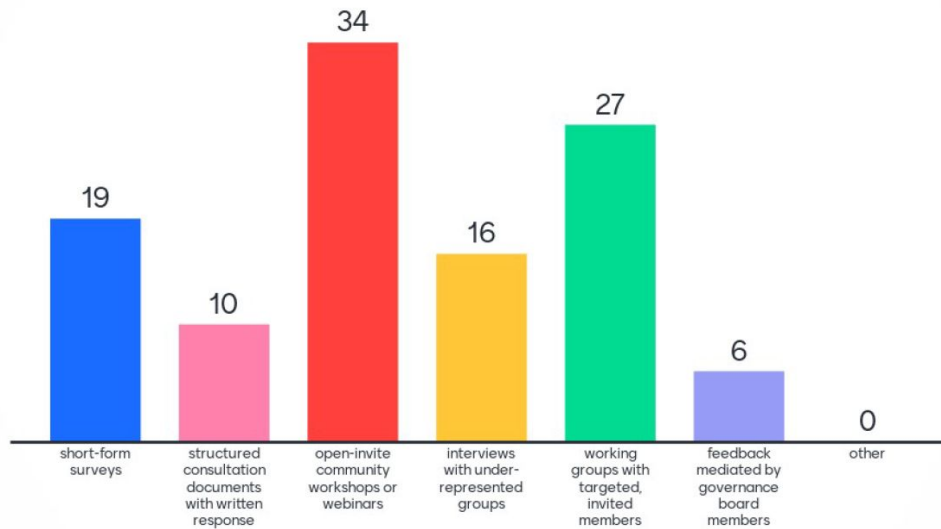
The session participants supported reviews every 3 years or more often. The criteria may evolve taking into account feedback from implementation. We would hope that community practice shifts towards higher compliance levels if incentives and support are put in place, at least within the 7 years of each Framework programme, allowing the minimal entry level to be increased. This desirable and expected progress is taken into account in the target metrics for EOSC proposed in Section 4.2.

The audience of the governance session was polled about the stakeholders to be consulted to review the recommendations (Figure 7). Each participant could select more than one answer. Services providers and scientific communities get a significant majority of votes of the 56 participants, respectively 86% and 66%, one third voted for policy makers and funders, one fifth for the end users. Research libraries, academic publishers, private sector, NGOs, students, PhD and beginners were proposed as possible additional stakeholders in the comments to the poll.

The participants agreed that the main kind of information to be gathered is feedback from implementation, challenges encountered, missing or unclear aspects, and recommendations for improvement. The ranking of the best ways of gathering the information is shown in Figure 7.

---

<sup>36</sup> <https://www.coretrustseal.org/why-certification/review-of-requirements/>



**Figure 7:** Results of the Mentimeter poll on the best way to gather information to maintain the recommendations (47 participants, several answers were possible)

EOSC Legal Entity and RDA were particularly cited, often together, when asked which body should be in charge of maintaining the recommendations. The importance of neutrality, independence, transparency, international point of view, diverse disciplinary background including underrepresented communities, and sustainability of the organisation in charge, were underlined as the characteristics the governance body needs.

The comment that maintenance should be performed in an international, neutral context, is well aligned with one of the conclusions of *Recommendations for Services in a FAIR Data Ecosystem* (Koers et al, 2020)<sup>37</sup>. They present the conclusions of a series of workshops gathering various stakeholders to define and prioritise recommendations for data and infrastructure providers to support findable, accessible, interoperable and reusable data within the scholarly ecosystem. They compare their findings to the Recommendations and Action Plan of *Turning FAIR into reality*. One of the recommendations they cannot readily map is “Foster global collaboration on FAIR implementation challenges and emerging solutions through organisations, such as the Research Data Alliance”, which is not a recommendation by itself in *Turning FAIR into reality*. They note that the recommendation clearly signals how the community values international collaboration and consensus-building.

## 3.2. Inclusiveness

### 3.2.1. Taking diversity into account

As also stated in Action 12.3 of *Turning FAIR into reality*, “the process of developing, approving and implementing FAIR metrics should follow a consultative methodology with research communities, including scenario planning to minimise any unintended consequences and counter-productive gaming that may result.” EOSC has to include communities at very different stages of making their data FAIR.

<sup>37</sup> Koers, H., Bangert, D., Hermans, E., van Horik, R., de Jong, M., Mokrane, M. Patterns, Vol. 1, Issue 5, 100058 (2020) <https://doi.org/10.1016/j.patter.2020.100058>

Some have well established, community-wide practices, others are just beginning, many are in between and may have already well established data services. An example test of the RDA WG criteria in the life sciences showed issues with the granularity of metadata attached to the collection or to the specific dataset, differences in how “persistent identification” should be understood, and with licences for different types of data publishing such as open access. Very similar findings were reached in astronomy, a discipline in which researchers find, access, reuse and interoperate data in their daily research work.

The EOSC will be a federation of existing resources. It will of course give access to new data, but it will primarily be a federation of existing data repositories and services, to be interfaced with existing data sharing frameworks. Taking community FAIR practices properly into account when defining general rules will be essential for uptake. Resources should be able to interface with EOSC with minimal overhead, and the data and functionalities available and used should remain.

One of the issues is that when disciplinary infrastructures and services are supported to provide data, they are supported to fulfill the requirements of their own community/ies. There are additional investment costs linked to additional requirements, which means that the adaptation of legacy data sharing systems to some of the FAIR criteria, when they come in addition to the community ones, will take time, even for providers which are already very efficient in terms of their community FAIR practices. It is thus very important that data providers with diverse disciplinary backgrounds test the proposed metrics and provide their feedback. The additional costs represent a significant barrier for those countries where science financing is weaker, but they can also present issues for research infrastructures (e.g. ESFRI) which have to mobilise additional resources at scale. Gradual implementation is to be expected in general, and it is essential to continuously evaluate the cost/benefit ratio.

Another issue for inclusiveness is that the weight of the different criteria can be different in the FAIR practices of different communities, for instance, depending on whether reproducibility is a key requirement (e.g. the life science community at the origin of the FAIR principles, which are a vehicle for reproducibility), or where reusability / interoperability is the starting point, such as in astronomy. The definition of “Essential” criteria will thus vary depending on the community practices and requirements. In addition, the uneven distribution of Essential criteria with none for the Interoperability principle, while understandable with respect to the current state of FAIR implementation in most communities, is also an issue. Machine actionability is of course a key aim of EOSC, even if it will not be implemented for all the resources from the start. The EOSC Interoperability Framework developed by the EOSC FAIR Working Group with the EOSC Architecture Working Group aims at setting a foundation for an efficient machine-enabled exchange of digital objects within EOSC and between EOSC and the outside world.

### 3.2.2. Consider FAIR as a journey

FAIR should be seen as a journey, and EOSC capacities will build up progressively. If FAIR is not seen as a continuum, we risk alienating communities who are not well advanced in sharing their data in a FAIR way, or indeed in data sharing at all. We also risk losing advanced communities for whom the effort to attain optional indicators does not outweigh the expected benefit. In addition to avoiding “mandatory” criteria as far as possible, using multi-step maturity scales to measure the FAIRness level of a resource, instead of a yes/no, pass/fail evaluation for each criterion, would provide an inclusive system, and a way to set up goals and measure progress. This can be visualised as radar charts, as proposed in the FAIR Data Maturity Model recommendation for their “measuring progress” evaluation (Figure 8).

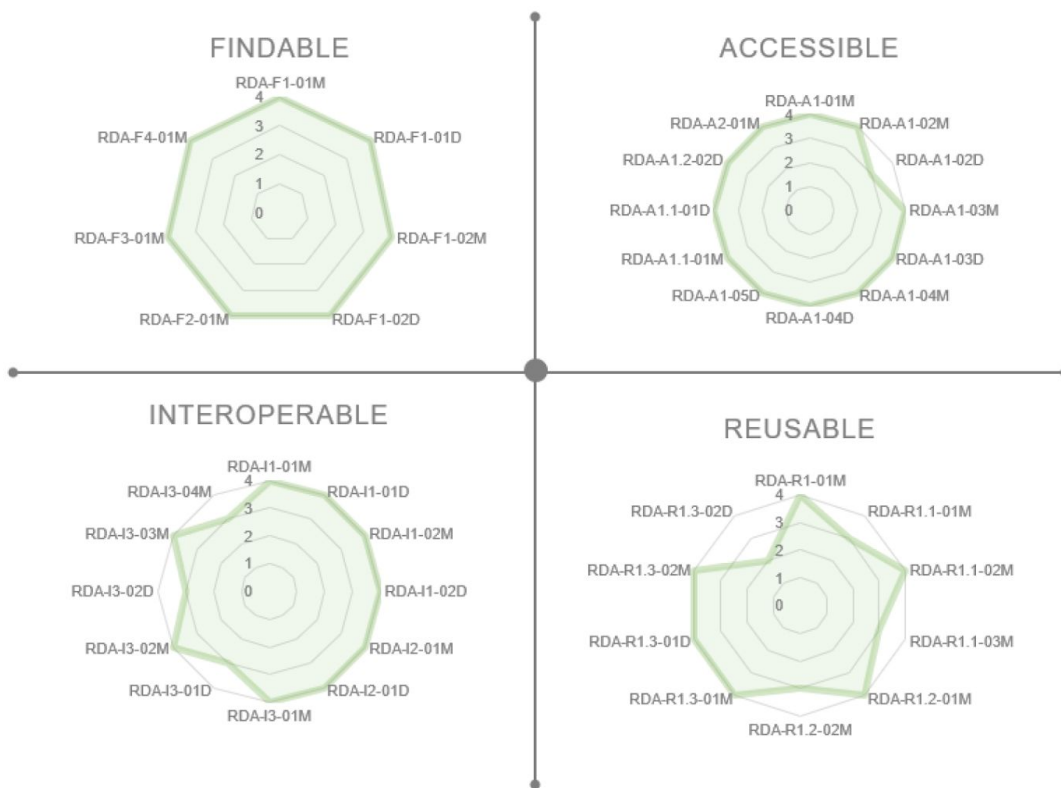


Figure 3 'Measuring progress' visualisation

Figure 8: Radar chart visualisation of the compliance with the criteria proposed by the RDA FAIR Data Maturity Model Working Group (Figure 3 of their Specification and Guidelines)

Recommendation 25 in *Turning FAIR into reality* notes how uptake in the implementation of FAIRness should be monitored over time; considering FAIRness as a continuum allows for progress along this continuum to be measured over time, and community progress to be measured diachronically.

Communities can mature their FAIRness levels as they develop more capabilities such as trained data stewards, tools to support FAIR data generation, the definition of data exchange standards, and underlying IT infrastructure. What has to be remembered is: FAIR transformation is a process and investing in it has a cost. Therefore targeting the highest maturity level might not be feasible for all cases all the time and is something we have to avoid in EOSC to ensure broad uptake and inclusivity. A value-based FAIR transformation can be considered as a perspective.

### 3.3. Not reinventing the wheel

The RDA FAIR Data Maturity Model Working Group worked transparently for 18 months to establish its list of criteria, including testing and community comments. Their starting point was based on throughout examination of a significant number of existing evaluation tools.

Our recommendation is that the RDA FAIR Data Maturity Model criteria should be used as a toolbox for further progress, to be adapted, in particular with respect to the priority level of the criteria, to the specific cases in which it will be used. This is the approach we take in Section 4.2, when discussing possible

elements of target FAIR Metrics for EOSC. People and projects using this model should liaise with the RDA Maintenance Group to provide their comments and support the work on possible necessary evaluation of the current version of the model.

### 3.4. Evaluation methods

The choice of the methods used to evaluate FAIRness is a key point which may have critical counter-productive consequences. Scalability would require automated evaluation, and the authors of the FAIR principles emphasise that machine actionability of the FAIR data resources was an original aim, and they push for automated metrics. Funders are also keen to get automated evaluations.

However, generalising automated evaluation of FAIRness as soon as the criteria have been defined but not yet fully tested would be a serious issue, especially in the current state of community preparedness, even if it is not used for pass/fail evaluation (Section 3.2.2). Even with a careful definition of the criteria, one will have to understand progressively how to perform the reviews, and how to take disciplinary differences and the results of their cost/benefit analysis into account. Another main risk is that algorithms introduce their own biases, as well as human evaluators in particular if a single person provides the evaluation. Also, automated evaluation is only capable of evaluating published data, at a stage where it is often too late to make easy changes to increase FAIRness: even if predictions on FAIRness can be done when processes are defined to obtain and prepare the data, a FAIR rating can be assigned only when the data exist and you can confirm that it has been done. What we really need is to influence decisions at the project planning stage and during project execution that will increase FAIRness after the project.

It is too early to recommend “the software solution” for FAIRness evaluation. Evaluation methods should be validated and tested before they are employed. Current automated methods are only applicable to already FAIRified data sets which follow certain standards. As FAIR transformation will be widespread, automated evaluators will be required to ensure scalability, but they will never be able to evaluate data sets still to be developed. To prevent unintended bias more research should be carried out on automated methods.

Many groups are working on defining FAIR assessment frameworks and/or to provide FAIR assessment tools. The first product of the RDA FAIR Data Maturity Model Working Group was an RDA Supporting Output *Results of an Analysis of Existing FAIR Assessment Tools*<sup>38</sup> which provides an overview of twelve existing FAIR assessment tools, listing the indicators used in these tools to assess the FAIRness of a data set. FAIRassist<sup>39</sup> is a list of tools provided by FAIRSharing.org. FAIRsFAIR M2.7 *Assessment Report on ‘FAIRness of services’*<sup>40</sup> also lists and analyses FAIR assessment frameworks and tools, building on and extending the RDA Working Group as well as the FAIRassist tool lists. One can also note the FAIRware initiative<sup>41</sup> launched by the Research on Research Institute (RoRI) to build open source

---

<sup>38</sup>

<https://www.rd-alliance.org/group/fair-data-maturity-model-wg/outcomes/results-analysis-existing-fair-assessment-tools>

<sup>39</sup> <https://fairassist.org/#/>

<sup>40</sup> <https://doi.org/10.5281/zenodo.3688762>

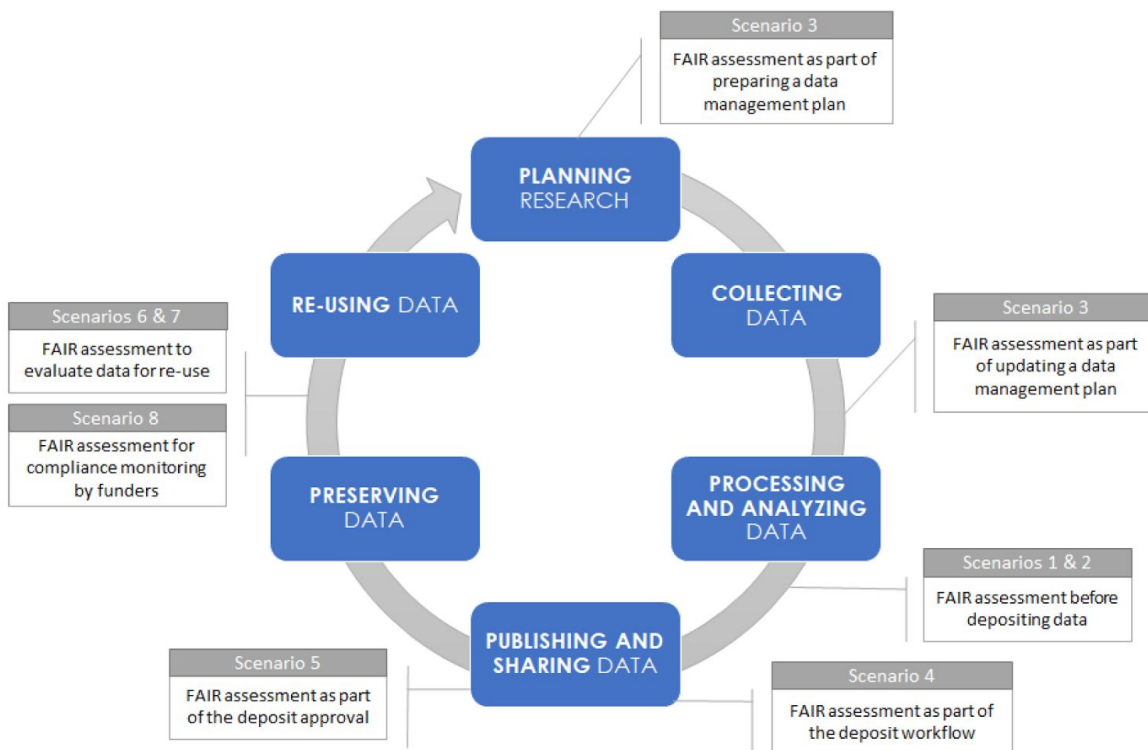
<sup>41</sup> <http://researchonresearch.org/announcement-partners-projects>. FAIRsFAIR will partner with the RoRI on FAIRware

<https://www.fairsfair.eu/news/rori-and-fairsfair-partner-fairware-initiative-funders-across-uk-europe-and-canada-join-forces>



software tool(s) to allow researchers, institutions and funders to assess and improve the FAIRness of the research outputs they produce.

There are many different points in a project lifecycle where FAIR evaluation could be useful (Figure 9). Each of these locations may require a different evaluation tool.



**Figure 9:** Research data lifecycle (figure adapted from Mosconi et al., 2019) and scenarios of FAIR assessment of datasets therein. Figure 3 from FAIRsFAIR M4.1, *Draft recommendations for FAIR datasets in certified repositories*.<sup>42</sup>

Taking into account all the aforementioned considerations, the Task Force believes it is currently premature to recommend any tools for (automatic) assessment of FAIRness.

### 3.5. FAIR metrics for other digital objects

The 5th recommendation of the *Six Recommendations for Implementation of FAIR Practices* is “Translate FAIR guidelines to other digital objects”, with the following rationale: “Applying FAIR principles to the context of specific communities requires adoption/translation. This need is more obvious in case of other (non-data) digital research objects where a direct mapping of the FAIR guiding principles may not be appropriate. The importance of each principle may depend on the priorities and maturity of the community in their use of certain research objects. This translation will need to be agreed in appropriate community

<sup>42</sup> FAIRsFAIR M4.1: <https://doi.org/10.5281/zenodo.3678716>; Mosconi, G., Li, Q., Randall, D. et al. Three Gaps in Opening Science. *Comput Supported Coop Work* 28, 749–789 (2019). et al., 2019, <https://doi.org/10.1007/s10606-019-09354-z>

fora, and such efforts should be incentivised financially (e.g. the costs and time required to organise community consultations).”

### 3.6. Providing guidelines

It is important to provide guidance, both to data providers and reviewers, as well as to the providers of evaluation systems. Metrics themselves do not include clear definitions for key concepts such as “metadata”, “automated”, “standardised, free and open source protocol”, “persistency”. The definition might vary by community or evaluator, which is by itself an issue with automated evaluation which removes nuances, specialist knowledge and context (Section 3.4). Community data and metadata standards should be referenceable via a community resource such as FAIRsharing.org<sup>43</sup> for ensuring a common understanding. Evaluators need guidelines with examples. The Guidelines produced by the RDA FAIR Data Maturity Model Working Group are a welcome step in that direction. It would be useful to have guidance tailored to communities, explaining how the FAIR criteria fit with their own community data sharing practices.

### 3.7. Cross-domain aspects

Interoperability comes in steps, from the level of the individual researcher to the disciplinary, then interdisciplinary ones. Each successive step requires a better documentation of the data, as new potential users share less of the context in which the dataset was obtained. Preparing data for cross-domain usage, which means a generic interpretation of data without prior context, can require enormous efforts, even for data which are FAIR for their community<sup>44</sup>. This can be justified when cross-domain use-cases are identified, but cannot be a mandatory requirement for all data. The concept of “Designated community” as defined by the Open Archival Information System (OAIS) and used by CoreTrustSeal is useful here: you need to know who you are trying to make content available to and for what purpose in order to provide the rich metadata they would need for reuse. Cross-domain requirements require careful study, and one should make sure that there is no unintended impact on hindering existing good data sharing practices within communities. Metrics, when they aim to be generic, have to be carefully defined in that respect.

Cross-disciplinary usage is in general better built on solid “disciplinary pillars” which ensure data and metadata quality. Brokering between existing thematic frameworks allows one to preserve the capabilities developed by the communities to fulfill their own needs.

## 4. Summary of the Status of FAIR Metrics, Draft Metrics for EOSC, Gaps and Priorities

### 4.1. Summary of the Status of FAIR Metrics

The state of definition of Metrics for FAIR Data is the most advanced to date. The RDA FAIR Data Maturity Model Working Group has published a model with 41 criteria allowing one to assess compliance

---

<sup>43</sup> <https://fairsharing.org>

<sup>44</sup> In communities with fast-changing technologies, even reuse inside the community after a few years may require additional annotations about e.g. known limitations of the data.

of data with the FAIR principles, using degrees of priority for each criterion - Essential, Important, Useful, which is “intended as a first version of the model”. The Working Group has moved to maintenance mode, which will work on promoting and improving the FAIR Data Maturity Model, and more generally FAIR assessments.

The model is being implemented, for instance in FAIRsFAIR, which is progressively defining criteria to deal with use cases. FAIRsFAIR Data Assessment Metrics is an on-going work. V0.3 (released July 2020) proposes 15 criteria which encompass the four FAIR principles and most of the 15 FAIR “guiding principles”. FAIRsFAIR Metrics do not retain all of the RDA “Essential” criteria, and some of the ones they propose are only deemed as “Important” in the RDA model, in particular because they include criteria for Interoperability, for which the RDA model has no ‘Essential’ criterion. They also group some of the RDA criteria together, and in some cases use different wording.

**As argued above, the RDA FAIR Data Maturity Model provides an essential toolbox for FAIR data metrics that has to be examined with a critical eye when defining criteria to evaluate the FAIRness of data in a given context. The degree of priority of the criteria should in particular be adapted to specific requirements.**

A tentative draft for recommended target and transitory FAIR Metrics for the EOSC, based on the RDA FAIR Data Maturity Model, is discussed in Section 4.2. This is by no way a definitive recommendation on FAIR Metrics for EOSC; instead, it proposes initial target aims for FAIRness in EOSC, which will be considered for future implementation, and it identifies intermediate steps in what must be considered a multi-step “journey” towards FAIR.

With respect to the other elements of the FAIR ecosystem, activities on the definition of FAIR for software are converging in the RDA/FORCE11/ReSA Working Group FAIR for Research Software, which was created in July 2020. Its aim is to define the FAIR principles for research software and provide guidelines on how to apply them. This WG should bring another key component to the FAIR ecosystem. They have 18 months starting at the date of creation of the WG to provide a draft recommendation. There is also on-going work on FAIR Semantics in FAIRsFAIR and the RDA Vocabulary Services Interest Group.<sup>45</sup>

## 4.2. Recommended FAIR Metrics for EOSC

The EOSC is defined as “the Web of FAIR data and services”, thus it is essential to progress the definition of criteria for FAIR in the EOSC. The FAIR Working Group is reluctant to propose a set version of metrics for EOSC at this stage: the wide participation in the development of the RDA Data Maturity Model and series of tests provide a good proof of concept, but to define general rules for inclusion in EOSC requires a considerably broader scope of tests that explores potential problems and fine-tunes the recommendations. Different communities attach different weights to the criteria, in particular but not only to interoperability, which has to be fully taken into account, and which lists no “Essential” priority in the RDA Maturity Model, but is nevertheless essential for EOSC.

To define a possible draft list of FAIR Metrics criteria for EOSC, we have used the criteria of the FAIR Data Maturity Model, checking also the more reduced list from FAIRsFAIR Data Object Assessment Metrics (V0.3). The latter provide an example of how to use the RDA Maturity Model as a starting point, but have been defined with repository use cases in mind, which may not fully reflect the EOSC use case.

---

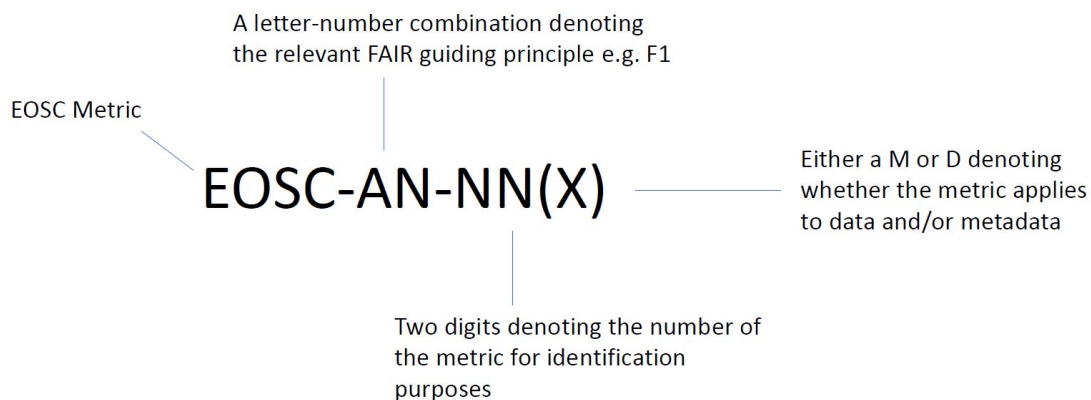
<sup>45</sup> <https://www.rd-alliance.org/groups/vocabulary-services-interest-group.html>

Our selection is done with EOSC requirements on data findability, accessibility, interoperability and reusability in mind, as well as machine actionability as a goal.

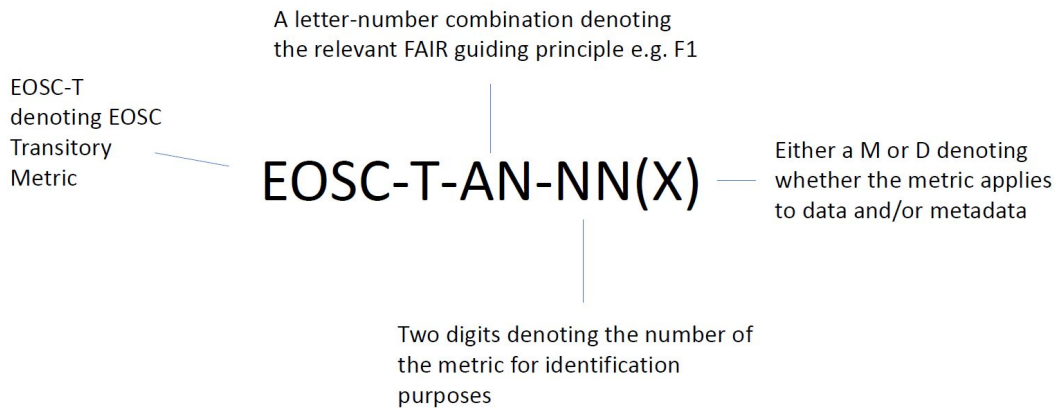
The definition of criteria has potentially very significant consequences if they are used to decide on participation or funding. Also metrics are not meant to be a punitive method for direct comparison between datasets from different areas, because communities will arrive at optimal FAIRness in different ways. These risks are well understood by the community, as shown during the SRIA consultation which gave a low priority to Metrics and Certification. Metrics therefore need to be implemented inclusively and progressively, taking fully into account that FAIR is a Journey, that community FAIR practices are diverse and that communities are at highly different stages of preparedness. **The criteria should be seen as a draft of a target for EOSC metrics, which will be met progressively by EOSC components, and should be tested thoroughly. It is essential to examine their applicability and to gather feedback in a wide range of contexts, in particular disciplinary contexts.**

We propose to take the need for progressiveness into account by a stepped process, with an evolving list of desirable criteria. Some criteria will appear at a given point of the process and remain, others, the “transitory” ones, will be in the list only temporarily, as a basic first stage, or basic first and second stages, towards a more mature status. The transitory criteria concern knowledge representation for discovery, licences and community standards. **The proposed timeline should also be extensively assessed during the tests of the proposed criteria.**

The naming schemes shown in Figures 10 and 11 are used for the proposed criteria and proposed transitory criteria respectively.



**Figure 10:** Naming scheme for the proposed EOSC criteria



**Figure 11:** Naming scheme for EOSC Transitory criteria

The proposed list of FAIR Metrics for EOSC is provided in Table 1. The criterion name in the first column is from the RDA FAIR Data Maturity Model WG list. The second column provides the RDA WG identifier/priority (E, I and U respectively for Essential, Important and Useful), and the third column the EOSC Metric name as defined. The “EOSC Timeline” columns provide the possible list of criteria to be considered in 2021, 2024 and 2028 respectively. “Transitory” criteria are tagged by their “EM-T” acronym and shown in italics in the list. A few comments are displayed in the last column.

Name (RDA WG)	RDA WG ID / Priority	EOSC Name	EOSC Timeline			Comments
			2021	2024	2028	
Data is identified by a persistent identifier	RDA-F1-01D/E	EOSC-F1-01D	EOSC-F1-01D	EOSC-F1-01D	EOSC-F1-01D	
Data is identified by a globally unique identifier	RDA-F1-02D/E	EOSC-F1-02D		EOSC-F1-02D	EOSC-F1-02D	
Rich metadata is provided to allow discovery	RDA-F2-01M/E	EOSC-F2-01M	EOSC-F2-01M	EOSC-F2-02M	EOSC-F2-02M	The expected content of "rich" will evolve with time.
Metadata includes the identifier for the data	RDA-F3-01M/E	EOSC-F3-01M	EOSC-F3-01M	EOSC-F3-01M	EOSC-F3-01M	
Metadata is offered in such a way that it can be harvested and indexed	RDA-F4-01M/E	EOSC-F4-01M		EOSC-F4-01M	EOSC-F4-01M	
Metadata contains information to enable the user to get access to the data	RDA-A1-01M/I	EOSC-A1-01M	EOSC-A1-01M	EOSC-A1-01M	EOSC-A1-01M	
<i>Metadata can be accessed manually (i.e. with human intervention)</i>	<i>RDA-A1-02M/E</i>	<i>EOSC-T-A1-03M</i>	<i>EOSC-T-A1-03M</i>			
<i>Data can be accessed manually (i.e. with human intervention)</i>	<i>RDA-A1-02D/E</i>	<i>EOSC-T-A1-03D</i>	<i>EOSC-T-A1-03D</i>			
Data identifier resolves	RDA-A1-03D/E	EOSC-A1-02D	EOSC-A1-02D	EOSC-A1-01M	EOSC-A1-01M	

to a digital object						
Metadata is accessed through standardized protocol	RDA-A1-04M/E	EOSC-A1-03M		EOSC-A1-03M	EOSC-A1-03M	
Data is accessible through standardised protocol	RDA-A1-04D/E	EOSC-A1-03D			EOSC-A1-03D	
Data can be accessed automatically (i.e. by a computer program)	RDA-A1-05D/I	EOSC-A1-04D		EOSC-A1-04D	EOSC-A1-04D	
Metadata is guaranteed to remain available, after data is no longer available	RDA-A2-01M/E	EOSC-A2-01M			EOSC-A2-01M	
<i>Metadata uses knowledge representation expressed in standardised format</i>	<i>RDA-I1-01M/I</i>	<i>EOSC-T-I1-01M</i>	<i>EOSC-T-I1-01M</i>	<i>EOSC-T-I1-01M</i>		This refers to knowledge representation in community formats.
<i>Data uses knowledge representation expressed in standardised format</i>	<i>RDA-I1-01D/I</i>	<i>EOSC-T-I1-01D</i>	<i>EOSC-T-I1-01D</i>	<i>EOSC-T-I1-01D</i>		This refers to knowledge representation in community formats.
Metadata uses machine understandable knowledge representation	RDA-I1-02M/I	EOSC-I1-01M			EOSC-I1-01D	The EOSC Interoperability Framework will build on community standards..
Data uses machine-understandable knowledge representation	RDA-I1-02D/I	EOSC-I1-01D			EOSC-I1-01D	
Metadata use FAIR-compliant vocabularies	RDA-I2-01M/I	EOSC-I2-01M			EOSC-I2-01M	
Plurality of accurate and relevant attributes are provided to allow reuse	RDA-R1-01M/E	EOSC-R1-01M		EOSC-R1-01M	EOSC-R1-01M	The expected content of "plurality" will evolve with time.
<i>Metadata includes information about the license under which the data can be reused</i>	<i>RDA-R1.1-01M/E</i>	<i>EOSC-T-R1.1-01M</i>	<i>EOSC-T-R1.1-01M</i>			
<i>Metadata refers to a standard reuse license</i>	<i>RDA-R1.1-02M/I</i>	<i>EOSC-T-R1.1-02M</i>		<i>EOSC-T-R1.1-02M</i>		
Metadata refers to a machine-understandable reuse licence	RDA-R1.1-03M/I	EOSC-R1.1-01M			EOSC-R1.1-01M	
Metadata includes provenance information according to community-specific standards	RDA-R1.2-01M/I	EOSC-R1.2-01M			EOSC-R1.2-01M	This criterion comes later because it is about provenance.
<i>Metadata complies with a community standard</i>	<i>RDA-R1.3-01M/E</i>	<i>EOSC-T-R1.3-01M</i>		<i>EOSC-T-R1.3-01M</i>		Progressive implementation when community standards are available.

Data complies with a community standard	RDA-R1.3-01D/E	EOSC-T-R1.3-01M		EOSC-T-R1.3-01M		Progressive implementation when community standards are available.
Metadata is expressed in compliance with a machine understandable community standard	RDA-R1.3-02M/E	EOSC-R1.3-01M			EOSC-R1.3-01M	

**Table 1:** Proposed timeline for EOSC FAIR Metrics, to be extensively assessed and tested

A detailed timeline for the aspects with “transitory” criteria are shown in Table 2.

Aspect	2021	2024	2028
<b>Discovery</b>	<ul style="list-style-type: none"> <li>• Metadata is provided to allow discovery</li> <li>• Metadata includes the identifier for the data</li> </ul>	<ul style="list-style-type: none"> <li>• Metadata is offered in such a way that it can be harvested and indexed</li> </ul>	<ul style="list-style-type: none"> <li>• Metadata is guaranteed to remain available, after data is no longer available</li> </ul>
<b>Licence</b>	<ul style="list-style-type: none"> <li>• Metadata includes information about the licence under which the data can be reused</li> </ul>	<ul style="list-style-type: none"> <li>• Metadata refers to a standard reuse licence</li> </ul>	<ul style="list-style-type: none"> <li>• Metadata refers to a machine understandable reuse licence</li> </ul>
<b>Standards</b>		<ul style="list-style-type: none"> <li>• Data/Metadata complies with a community standard</li> </ul>	<ul style="list-style-type: none"> <li>• Data/metadata is expressed in compliance with a machine understandable community standard</li> <li>• Metadata use FAIR-compliant vocabularies</li> </ul>

**Table 2:** Timeline for the “transitory” criteria

There are a number of issues with the implementation of the criteria with respect to existing practices. For instance, we focus the F1 criteria on data, with no separate criteria for metadata, because they are often presented together in existing systems. There are also different community practices for the usage of PIDs, which is also recognised in the *Persistent Identifier (PID) Policy for the European Open Science Cloud*<sup>46</sup> produced by the EOSC FAIR and Architecture Working Groups. The wording of several criteria remains vague, with terms such “knowledge representation”, “FAIR vocabularies”, “plurality of accurate and relevant attributes”, which come from the FAIR guiding principles themselves. We consider here the “cross-community language” relevant in the EOSC context to be the EOSC Interoperability Framework. The guidelines included in the RDA Recommendation are in general useful and should be improved if needed with application to EOSC in mind. The main issues are likely that some communities do not yet have the required standards, or they are not fully implemented in the community data holdings.

When using such metrics a large measure of flexibility is needed, which is a real issue for automated evaluation, and the information about the relevant community and community standards should be

<sup>46</sup> The pre-typeset version of the document is available at <https://docs.google.com/document/d/1IHFaei35j1ecGc--gaSrYeq5qMkclHVZhfuyCPjjax8/edit>.

available. The information gathered in particular in FAIRSharing.org is an asset for FAIRness evaluation in that respect.

One should remember here that FAIR transformation is a process, that care should be taken to ensure broad uptake and inclusiveness, and that these metrics are a draft proposed for extensive discussion and testing.

### 4.3. Gaps and potential opportunities for extension

Existing work on FAIR metrics should be extended under Horizon Europe to ensure applicability across disciplines and support implementation. FAIR assessment should be inclusive and progressive, and its usage should take the specific context and needs into account. Metrics should be combined with a FAIR assessment framework, which reflects the needs of different communities while offering comparable methods to assess FAIRness.

The questions of authorisation and authentication and openness should be discussed further. The EOSC business models suggest access to the EOSC ecosystem through authorisation and authentication, whereas some communities' practices are based on the usage of open resources. In addition, in some cases a usage licence is not attached to the data but the usage relies on community ethics. The implementation of explicit usage licences is of course desirable, but it will likely take time. EOSC take-up by well-established communities could be jeopardised if they lose some of the capacities they are accustomed to use.

The proposed set of target FAIR metrics criteria for EOSC is aimed as a starting point for thorough assessment. One key issue is that while some of the criteria we see as straightforward, already widely applied or easily implementable, implementation of some other criteria might be problematic with some or many communities. We emphasise the need for a much broader community consultation before application, and acknowledging the need for progressive implementation. If this cross community consultation is skipped or applied partially, we risk closing the doors of EOSC to some disciplines.

The proposed criteria, and more generally the RDA FAIR Data Maturity Model can already be used to measure progress on the path to FAIRness, but lots of care should be taken before applying any model for pass/fail measurements. The need to develop automated evaluation tools for scalability is recognised but there are risks associated to the tool biases. More work is required on criteria and tools before providing "the software solution" for FAIR evaluation.

The individual assessment models and metrics should be aligned with RDA core metrics and should not hinder a comparative evaluation.

Other critical elements of EOSC include software, PID services, semantics and registries, for which assessment frameworks have yet to be defined. The alignment of the repository certification schema with FAIR is underway but needs to be further developed and tested. This point will be discussed in the companion document on certifying the services required to enable FAIR research outputs within EOSC.

All the assessment frameworks have to be maintained over time, taking into account feedback from implementation and evolving requirements; the FAIR guiding principles themselves may have to be maintained.



We propose to update the proposed metrics first after two years, then every three years. This proposal can be adjusted depending on the situation, for instance if the RDA criteria evolve with time. We propose that an RDA Working Group is created in the framework of the RDA Global Open Research Commons Interest Group, which gathers EOSC and other similar initiatives from other regions, to seek international agreement. The Working Group recommendations should be ratified at EOSC level, for instance by a Working Group of the EOSC Stakeholder Forum. Engaging the communities and gathering their feedback will be essential. This WG should be created immediately with participation of EOSC stakeholders as a neutral, international forum to assess and test the proposed metrics.

Metrics and certification provide an a posteriori approach to quality control. To achieve sustainable and robust progress, FAIRness should be built into the whole research process: from planning through observation or data creation till publication. We emphasise two essential components: the human (researcher and data support roles) and software/instrumental (instruments and software which implements data documentation).

#### 4.4. Priorities for future work

Significant progress has been made on defining FAIR metrics and certification schema for repositories. This should continue to be built on rather than reinventing the wheel, particularly given the global input and consensus fostered via the Research Data Alliance on these topics. Priorities for FAIR metrics lie in implementation and robustly testing across research communities.

**Priority 1:** Support the assessment and improvement of the RDA FAIR Data Maturity Model.

Priority 1.1: Support disciplinary communities to clarify their requirements with respect to FAIR and identify cross-community use cases.

Priority 1.2: Test the FAIR data maturity model in a wide range of communities, in a neutral forum and seeking for international agreement, to fine-tune and customise the recommendations and guidance, assess the degree of priorities, identify adverse consequences and apply corrections.

**Priority 2:** Assess and test the proposed EOSC FAIR data metrics in a neutral forum, which could be a Working Group set up by the RDA Global Open Research Commons Interest Group to seek for international agreement with the international EOSC counterparts.

**Priority 3:** Support the definition of evaluation tools; their thorough assessment and evaluation including inclusiveness; comparison of tools (manual, automated); identification of their biases and applicability in different many contexts, including thematic ones.

**Priority 4:** Support the definition of FAIR for software and of the assessment framework for key elements of the FAIR ecosystem, in particular PID services and semantics.

**Priority 5:** Define and implement governance of the principles, assessment frameworks and metrics, adapted to each specific case.

**Priority 6:** Guidance for and support to implementation: support data and service providers to progress in the FAIRness of their holdings.

Success will be strongly dependent on take-up by communities, especially the thematic ones. Three among the *Six Recommendations for Implementation of FAIR Practices* (FP), which detail support which

should be provided, are fully endorsed as necessary for the implementation and take-up of FAIR Metrics, namely:

**FP Recommendation 1** Fund awareness-raising, training, education and community-specific support.

**FP Recommendation 2** Fund development, adoption and maintenance of community standards, tools and infrastructures.

**FP Recommendation 3** Incentivise development of community governance.

EOSC is one of the stakeholders identified for all three recommendations. Their Recommendation 4, “Translate FAIR Guidelines for other digital objects”, is included in our Priority 4.