

## PRACTICE PAPER

# The FAIR Data Maturity Model: An Approach to Harmonise FAIR Assessments

Christophe Bahim<sup>1</sup>, Carlos Casorrán-Amilburu<sup>2</sup>, Makx Dekkers<sup>3</sup>, Edit Herczog<sup>4</sup>, Nicolas Loozen<sup>1</sup>, Konstantinos Repanas<sup>2</sup>, Keith Russell<sup>5</sup> and Shelley Stall<sup>6</sup>

<sup>1</sup> PwC, BE

<sup>2</sup> European Commission, BE

<sup>3</sup> Independent, ES

<sup>4</sup> Vision and Values, BE

<sup>5</sup> Australian Research Data Commons, AU

<sup>6</sup> American Geophysical Union, US

Corresponding author: Christophe Bahim ([christophe.bahim@pwc.com](mailto:christophe.bahim@pwc.com))

In the past years, many methodologies and tools have been developed to assess the FAIRness of research data. These different methodologies and tools have been based on various interpretations of the FAIR principles, which makes comparison of the results of the assessments difficult. The work in the RDA FAIR Data Maturity Model Working Group reported here has delivered a set of indicators with priorities and guidelines that provide a ‘lingua franca’ that can be used to make the results of the assessment using those methodologies and tools comparable. The model can act as a tool that can be used by various stakeholders, including researchers, data stewards, policy makers and funding agencies, to gain insight into the current FAIRness of data as well as into the aspects that can be improved to increase the potential for reuse of research data. Through increased efficiency and effectiveness, it helps research activities to solve societal challenges and to support evidence-based decisions. The Maturity Model is publicly available and the Working Group is encouraging application of the model in practice. Experience with the model will be taken into account in the further development of the model.

**Keywords:** FAIR; research data; assessment; digital object; evaluation; maturity; data management

## 1 Introduction

Over the last couple of years, several methodologies and tools for the assessment of FAIRness have been developed and tested (Bahim, Dekkers, & Wyns, 2019). Nevertheless, as the FAIR principles (Wilkinson, Dumontier, & Aalbersberg, 2016) do not strictly define how to achieve a state of FAIRness, the methodologies and tools were developed based on a range of different interpretations. As a result, methodologies and tools produce results that do not allow benchmarking. In addition, research organisations and data infrastructures cannot develop or follow a minimum set of shared guidelines to improve FAIRness of data because of the heterogeneity of the available tools for assessment.

To address this situation, a working group was set up under the aegis of the Research Data Alliance (RDA), taking advantage of RDA’s global cross-sectoral impartial community. The RDA Working Group “FAIR Data Maturity Model”<sup>1</sup> was established with the objective to develop a set of assessment criteria to facilitate comparisons across assessment approaches and reach consensus among the over 200 members of the Working Group. This paper describes the process used by the Working Group that led to the agreement on the assessment criteria. The novelty in this work is that it was the first time that people representing such a wide

<sup>1</sup> <https://www.rd-alliance.org/groups/fair-data-maturity-model-wg>.

range of backgrounds and regions were brought together. The work was carried out in the period from January 2019 to June 2020, with editorial support provided by the European Commission, and was led by three co-chairs representing perspectives from Europe, the US and Australia. The resulting consensus was published as an RDA Recommendation (FAIR Data Maturity Model Working Group, 2020) in June 2020 which is openly available for use by anyone, anywhere in the world, allowing for broad adoption. Adoption of the Recommendation is underway in various places, and experience that is being gained will be fed back into the further development of the Recommendation in the coming years.

## 2 Approach

The Working Group used a methodology to develop a FAIR data maturity model in four phases (**Figure 1**).

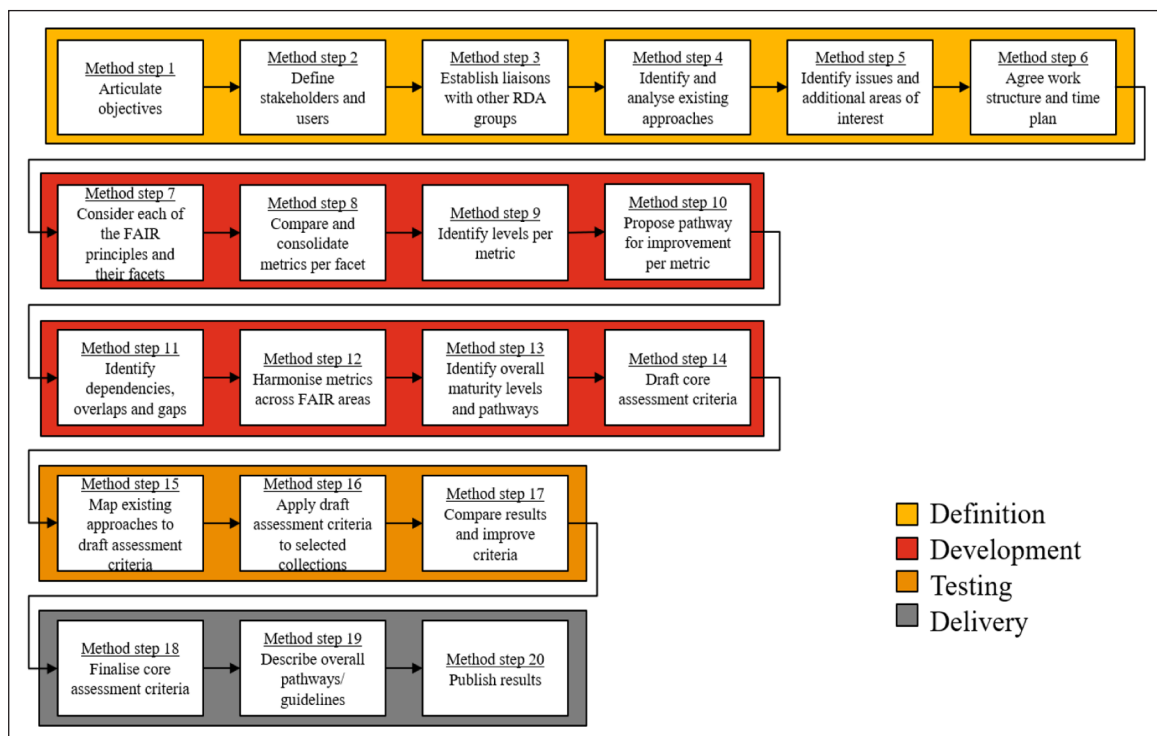
In the first phase, the definition phase, a scoping, planning and landscaping exercise were carried out. The landscaping exercise was published (Bahim, Dekkers, & Wyns, 2019) and served as a basis for the next phase.

Once the baseline was established, the second phase consisted in building the model. To this end, all aspects mentioned in the FAIR principles that could be assessed were identified. All input received from the Working Group members was clarified, standardised, and consolidated in a draft set of indicators, together with proposed priorities and draft guidelines for the application of the indicators in practical situations.

The purpose of the third phase, the testing phase, was to determine the soundness, completeness and usability of the indicators. After a pilot test, testers used a test plan with a template to record results and feedback. The results of the testing phase allowed the Working Group to revise the model and present a stable version as a draft RDA Recommendation. Issues that were brought up during the testing phase are included in the Discussion section further on.

The fourth and last phase was the delivery phase, to finalise and publish the FAIR Data Maturity Model (FAIR Data Maturity Model Working Group, 2020) as an RDA Recommendation.

Throughout these four phases, nine working meetings<sup>2</sup> were held to develop the model. True to the spirit of any RDA Working Group, during the meetings of this Working Group, consensus among our members was sought. The consensus process used several tools: in addition to discussion on GitHub,<sup>3</sup> a number of surveys



**Figure 1:** Development methodology.

<sup>2</sup> <https://www.rd-alliance.org/node/60731/file-repository>.

<sup>3</sup> <https://github.com/RDA-FAIR/FAIR-data-maturity-model-WG>.

were conducted to gather opinions on proposed decisions, and draft versions of the recommendation were made available as a Google document in which the members of the Working Group could make comments and suggest improvements. This ensured that all views and opinions were taken into account and that the resulting Recommendation truly represented consensus in the Working Group.

### 3 Results

The landscaping exercise brought many interesting points to light. For example, there were notable differences in the formats used in the approaches that were analysed: some used checklists, others used questionnaires. Not all the initiatives assessed the same type of object: some focused on datasets and their use, while others assessed the data management plan. Some elements of the FAIR principles were covered more often than others; for example, the area of findability was the most covered whereas the area of accessibility received less attention. Furthermore, some approaches included questions that did not directly refer to any of the FAIR principles. For example, there were questions about the data repository and curation practices, which are not explicitly mentioned in the FAIR principles. Another important difference was the way in which answers were to be provided, with some methodologies asking for yes/no answers and others giving a set of options to choose from.

The Working Group used the landscaping exercise as base material to develop the set of indicators and maturity levels – with respect to the level of FAIRness – which were discussed on GitHub.

The indicators that are used in the FAIR Data Maturity Model are derived from the FAIR principles and aim to formulate measurable aspects of each principle that can be used by evaluation approaches (**Table 1**).

**Table 1:** FAIR data maturity model indicators.

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

(Contd.)

FAIR ID	Indicator	Priority	
I1	RDA-I1-01M	Metadata uses knowledge representation expressed in standardised format	<input type="checkbox"/> <input type="checkbox"/> Important
I1	RDA-I1-01D	Data uses knowledge representation expressed in standardised format	<input type="checkbox"/> <input type="checkbox"/> Important
I1	RDA-I1-02M	Metadata uses machine-understandable knowledge representation	<input type="checkbox"/> <input type="checkbox"/> Important
I1	RDA-I1-02D	Data uses machine-understandable knowledge representation	<input type="checkbox"/> <input type="checkbox"/> Important
I2	RDA-I2-01M	Metadata uses FAIR-compliant vocabularies	<input type="checkbox"/> <input type="checkbox"/> Important
I2	RDA-I2-01D	Data uses FAIR-compliant vocabularies	<input type="checkbox"/> Useful
I3	RDA-I3-01M	Metadata includes references to other metadata	<input type="checkbox"/> <input type="checkbox"/> Important
I3	RDA-I3-01D	Data includes references to other data	<input type="checkbox"/> Useful
I3	RDA-I3-02M	Metadata includes references to other data	<input type="checkbox"/> Useful
I3	RDA-I3-02D	Data includes qualified references to other data	<input type="checkbox"/> Useful
I3	RDA-I3-03M	Metadata includes qualified references to other metadata	<input type="checkbox"/> <input type="checkbox"/> Important
I3	RDA-I3-04M	Metadata include qualified references to other data	<input type="checkbox"/> Useful
R1	RDA-R1-01M	Plurality of accurate and relevant attributes are provided to allow reuse	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Essential
R1.1	RDA-R1.1-01M	Metadata includes information about the licence under which the data can be reused	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Essential
R1.1	RDA-R1.1-02M	Metadata refers to a standard reuse licence	<input type="checkbox"/> <input type="checkbox"/> Important
R1.1	RDA-R1.1-03M	Metadata refers to a machine-understandable reuse licence	<input type="checkbox"/> <input type="checkbox"/> Important
R1.2	RDA-R1.2-01M	Metadata includes provenance information according to community-specific standards	<input type="checkbox"/> <input type="checkbox"/> Important
R1.2	RDA-R1.2-02M	Metadata includes provenance information according to a cross-community language	<input type="checkbox"/> Useful
R1.3	RDA-R1.3-01M	Metadata complies with a community standard	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Essential
R1.3	RDA-R1.3-01D	Data complies with a community standard	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Essential
R1.3	RDA-R1.3-02M	Metadata is expressed in compliance with a machine-understandable community standard	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Essential
R1.3	RDA-R1.3-02D	Data is expressed in compliance with a machine-understandable community standard	<input type="checkbox"/> <input type="checkbox"/> Important

**Table 2:** Priorities.**Indicator priorities**

**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.

In order to determine the relative importance of the indicators, the Working Group defined a set of priorities as shown in **Table 2**.

The indicators defined in the FAIR Data Maturity Model can be used to evaluate data resources and their metadata. The indicators are primarily intended to be used as the foundation, on top of which evaluation methodologies can be built. Each methodology can then define its own questions or metrics.

In addition to the assignment of priorities to the indicators, the Working Group also developed two ways of 'scoring' the evaluation: the first was to look at the progress made per indicator on a five-level scale, while the second assigned a yes/no score to each indicator.

1. **Measuring progress:** the emphasis lies on delivering a measure of the extent to which a resource under evaluation meets the requirements expressed in an indicator.
2. **Measuring pass-or-fail:** the emphasis lies on determining whether a resource under evaluation meets the requirement of an indicator on a binary, pass or-fail scale.

The model may be used during the development of Data Management Plans for research data, before any data resources have been produced, to specify the level of FAIRness that the resources are expected to achieve. It can also be used after the production of data resources to test what is the achieved level of FAIRness of the resources. Data producers, e.g. researchers, and data publishers can use the model to determine how to improve the FAIRness of their data, while project managers and funding agencies can use the model to determine whether the data resources achieve a pre-defined, expected level of FAIRness.

#### 4 Observations

It is important to note that the Working Group agreed that assessment should not be conceived as a value judgment but rather as a means to encourage improvement of the level of data FAIRness.

It also needs to be noted that the FAIR principles are aspirational in nature, focusing on a long-term view of improving the potential for reuse of research data. As such, they should not be interpreted as strict rules. Communities have different practices and therefore there should be some flexibility on how the FAIR principles should be implemented within a certain community, i.e. a professional domain or grouping of related organisations. Such flexibility allows different communities to adapt the speed of their 'journey' towards greater FAIRness of their data.

Although the word "community" is mentioned only once in the FAIR Principles (R1.3) (The FAIR Data Principles|FORCE11; FAIR Principles – GO FAIR; Wilkinson, Dumontier, & Aalbersberg, 2016), it was agreed that consensus within communities was a first crucial step towards FAIRness. Such consensus also helps to calibrate the priorities assigned to the indicators as these priorities depend on the way researchers in a particular community think about the digital objects that are relevant to them, as well as on existing community standards.

In the F2 principle "Rich metadata is provided to allow discovery", 'rich' allows for different interpretations. Because of this, the Working Group agreed that further discussion is necessary to narrow down the definition of 'rich', and to pinpoint its essential elements. This discussion will be taken forward in the future work being planned for the maintenance phase of the Working Group.

Identifying metadata and/or data is a practice that is different across disciplines. The Working Group decided that both metadata and data should be identified with a persistent identifier (PID) because both metadata and data were considered important in their own right. However, it was acknowledged that requiring separate PIDs to be assigned to both data and metadata might not align with existing practices where a PID resolves to a landing page that contains the metadata of the object and a link (e.g. URL) to the actual data contents. The main argument behind this was that machines should be able to automatically locate the actual data without needing a human to interpret the landing page. Assumptions on the way that data and metadata objects are identified in practice will influence the implementation of the FAIR principles.

#### 5 Discussion

The FAIR principles represent a collection of best practices regarding data management that is general enough to be valid across domains and has brought visibility on how poorly curated data can hinder data reuse and delay scientific advancement.

The work of the FAIR Data Maturity Model Working Group has strengthened the realisation that community involvement and input is crucial to identify and build the standards and approaches that are required to enable FAIR data in a range of disciplines. These are also the standards and approaches that are required down the track for automated assessment of the FAIRness of data. At present, there is a lack of such standards and approaches in most disciplines to make such automated assessment possible.

Specifically, the work carried out in the FAIR Data Maturity Model Working Group has been essential in identifying the relevant indicators when understanding the degree to which a digital object is FAIR. For any indicators to be accepted by the broad scientific community, they must come from the community itself and, as such, the outputs of this RDA Working Group are the result of extensive consultation of and dialogue between domain experts from across disciplines. They represent a solid community-endorsed foundation on which further work can be built.

From a policy maker's perspective, the indicators are essential in raising awareness of the FAIR principles, and in ensuring that the different research communities build comparable tools with which they can assess their respective FAIR data maturity.

From a funder's perspective, these community-approved indicators are a necessary first step towards building a sound methodology to assess the FAIRness of publicly funded research outputs. As data management and FAIR become increasingly important elements in the assessment of proposals, and in the periodic assessments of ongoing projects, funders can only benefit from effective, unbiased and inclusive methodologies with which to evaluate the FAIRness of the research data produced.

In North America, particularly within some private funders, the value of well-documented, shared data is promoted and sometimes required. Notably, the U.S. Geological Survey (USGS) has taken decisive steps toward making their data FAIR and also including the role of repositories in providing necessary FAIR services on behalf of researchers, such as registering datasets for Digital Object Identifiers (Lightsom, 2019). We are hopeful that as USGS progresses in their work, they will consider using the FAIR Data Maturity Model to evaluate their approach to FAIRness. Furthermore, work at the National Oceanic and Atmospheric Administration (NOAA) on a method to optimize the reuse of NOAA data through a various criteria of completed information positioned them to better support their researchers and help this Working Group as one of the adoption stories demonstrating that efforts made to make data FAIR benefit researchers and their communities.<sup>4</sup>

From the European Commission's perspective, with regards to policy, the FAIR principles, are instrumental in promoting the Open Science policy<sup>5</sup> of the Commission, which includes the wide and early sharing of knowledge (e.g. data) and tools by researchers within and across disciplines, and with society at large. The FAIR principles are at the heart of the European Open Science Cloud,<sup>6</sup> an initiative to build a trusted, open and distributed system for the scientific community, providing researchers with a seamless access to a web of FAIR data and services built on top of those data. With regards to the Commission's role as a funder, the FAIR principles are going to be very prominent under the upcoming framework programme for Research and Innovation, Horizon Europe (Commission, 2018). The Commission will advocate for "responsible research data management in line with the FAIR principles". Applicants will be evaluated at proposal stage on their plans with respect to data management, and specifically, on how they plan on making their data FAIR. This will continue to be relevant during the lifetime of a project, as beneficiaries will have to report on their data through a data management plan that will need to be updated during the project.

As the scientific community values datasets that underpin research findings, the need to ensure the quality, understanding, and consistency of how these datasets are prepared for others to discover and experience needs a method of measurement. The FAIR Data Maturity Model provides a way for these community-based FAIR assessments to have comparable results and provide consistent feedback as to how well communities are doing in making research data FAIR. This is of interest to funders, institutions, and publishers in support of openness, transparency, and integrity. We welcome all those interested in this work to join the RDA Working Group and continue to support adoption efforts and feedback on improvement.

## 6 Further work

With the publication of the RDA Recommendation, the RDA FAIR Data Maturity Model Working Group reached the end of its charter. As it was considered useful to continue the work, the Working Group has transitioned to a maintenance Working Group dedicated to maintaining the FAIR data maturity model. As part of the ongoing work, feedback will be solicited from the community and other interested parties to take care of the further development and revision of the model. The focus will be on (1) coordinating efforts to ensure the adoption of the FAIR data maturity model by developers of assessment approaches and tools, (2) maintaining the deliverables produced taking into account feedback from the communities, and (3) proposing a governance model for handling maintenance activities. The maintenance Working Group will operate without time limit and will be home to discussion forums to address specific problems related to the outputs and recommendations produced during the inception phase.

<sup>4</sup> <https://www.ncei.noaa.gov/products/ocean-exploration-and-research-oer-data-management>.

<sup>5</sup> <https://ec.europa.eu/research/openscience/>.

<sup>6</sup> <https://ec.europa.eu/digital-single-market/en/%20european-cloud-initiative>.

## Acknowledgements

Throughout the development of the FAIR data maturity model, many people significantly contributed to the discussions and meetings. The Chairs, the European Commission and the editorial team is truly grateful for the work they have done and would like to thank them. Below you can find the names of the contributors, in no particular order; Athanasios Karalopoulos, Alejandra González Beltrán, Alicia Fátima Gómez Sánchez, Andras Holl, Anusuriya Devaraju, Barbara Sierman, Carole Goble, Françoise Genova, Ge Peng, Gerry Coen, Helen Parkinson, Hervé L'hours, Keith Jeffery, Kerry Levett, Kevin Long, Jean-Eudes Hollebecq, Jolanda Strubel, Jonathan Petters, Leyla Garcia, Marco Molinaro, Maggie Hellström, Mark Wilkinson, Marta Teperek, Michel Dumontier, Nichola Burton, Nick Juty, Mustapha Mokrane, Oya Deniz Beyan, Peter McQuilton, Rob Hooft, Romain David, Susanna Sansone and Yann Le Franc.

The publication of this paper was supported by the RDA Europe 4.0 project that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 777388.

## Competing Interests

The authors have no competing interests to declare.

## References

- Bahim, C, Dekkers, M and Wyns, B.** 2019. Results of an Analysis of Existing FAIR Assessment Tools. *Research Data Alliance*. DOI: <https://doi.org/10.15497/RDA00035>
- Commission, E.** 2018. Establishing Horizon Europe – the Framework Programme for Research and Innovation, laying down its rules for participation and dissemination. Retrieved from EU law – EUR-Lex: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52018PC0435>.
- FAIR Data Maturity Model Working Group.** 2020. FAIR Data Maturity Model: specification and guidelines. *Research Data Alliance*. DOI: <https://doi.org/10.15497/RDA00050>
- FAIR Principles – GO FAIR.** n.d. Retrieved from GO FAIR: <https://www.go-fair.org/fair-principles/>.
- Lightsom, F.** 2019, May 30. Building a Roadmap for Making Data FAIR in the U.S. Geological Survey: A 2019 CDI Project. Retrieved from USGS: <https://my.usgs.gov/confluence/display/cdi/Building+a+Roadmap+for+Making+Data+FAIR+in+the+U.S.+Geological+Survey%3A+A+2019+CDI+Project>.
- The FAIR Data Principles|FORCE 11.** n.d. Retrieved from Force 11|The future of research communications and e-scholarship: <https://www.force11.org/group/fairgroup/fairprinciples>.
- Wilkinson, MD, Dumontier, M and Aalbersberg, IJ.** 2016. The FAIR Guiding Principles for scientific data management and stewardship. DOI: <https://doi.org/10.1038/sdata.2016.18>


**How to cite this article:** Bahim, C, Casorrán-Amilburu, C, Dekkers, M, Herczog, E, Loozen, N, Repanas, K, Russell, K and Stall, S. 2020. The FAIR Data Maturity Model: An Approach to Harmonise FAIR Assessments. *Data Science Journal*, 19: 41, pp. 1–7. DOI: <https://doi.org/10.5334/dsj-2020-041>

**Submitted:** 13 July 2020

**Accepted:** 19 October 2020

**Published:** 27 October 2020

**Copyright:** © 2020 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

 *Data Science Journal* is a peer-reviewed open access journal published by Ubiquity Press.

**OPEN ACCESS** 