FAIROs: Towards FAIR assessment in Research Objects

Esteban González[1 [0000−0003−4112−6825]], Alejandro Benítez[1], and Daniel Garijo[1 [0000−0003−0454−7145]]

Universidad Politécnica de Madrid, Spain
{esteban.gonzalez, daniel.garijo}@upm.es,
guillermoalejandro.benitez@alumnos.upm.es

Abstract. The FAIR principles have become a popular means to publish research outputs (i.e., data, software, ontologies, etc.) in a Findable, Accessible, Interoperable and Reusable manner. In fact, the scientific community has developed frameworks to assess the FAIRness of individual scientific outputs (like data and software), offering guidance and suggestions to researchers. However, scientific outputs are rarely published in isolation. Research Objects have been proposed as a framework to capture the relationships and context of all constituents of an investigation. In this paper we present FAIROs, a framework for assessing the compliance of a Research Object (and its constituents) against the FAIR principles. FAIROs reuses existing FAIR validators for individual resources and proposes i) two scoring methods for assessing the fairness of Research Objects, ii) an initial implementation of the scoring methods in the FAIROs framework, and iii) an explanation-based approach designed to visualize the obtained scores. We validate FAIROs against 165 Research Objects, and start a discussion towards assessing the advantages and limitations of different scoring systems.

Keywords: FAIR assessment · Research object · Aggregation methods.

1 Introduction

The Findable, Accessible, Interoperable and Reusable principles (FAIR) introduce a set of best practices to share data, make data more reusable and support the reproducibility of results in research [22]. FAIR addresses using persistent identifiers for resources, rich metadata to favour discovery, explicit licenses to understand usage terms and using well established vocabularies to facilitate interoperability.

Although FAIR was originally proposed for datasets, additional initiatives are appearing to apply these principles to other research outputs, such as

software [15] [16], ontologies [19], virtual research environments[1] or digital objects [4] among others.

Since research outputs are rarely produced in isolation, the scientific community has proposed Research Objects [2] [20] to capture the context around a scientific investigation. Research Objects also provide the means to pack all the resources within some research, easing its understandability and facilitating its dissemination. However, assessing the compliance of a Research Object against the FAIR principles (i.e., their FAIRness) is challenging, as Research Objects aggregate multiple resources which may be prompt to individual assessment.

In this work we describe FAIROs, a framework for assessing the FAIRness of a research investigation, modeled as Research Object. Our contributions include i) an approach to integrate the FAIRness scores of all resources in a Research Object, ii) an implementation of our approach in an executable tool, and iii) a visualization of the scoring system for helping users understand how to improve the FAIRness of their resources.

FAIROs integrates existing efforts developed to assess the FAIRness of individual datasets, software projects and ontologies. We have tested our framework by calculating the FAIRness scores of over 100 Research Objects available in public repositories, comparing two different aggregation metrics. Based on our results, we discuss the advantages and disadvantages of different scoring methods for Research Object FAIRness.

The rest of the paper is structured as follows. Section 2 describes existing efforts for assessing FAIR in different domains. Section 3 describes our approach, while Section 4 compares differences when using two scoring methods. Section 5 discusses our results, and Section 6 concludes the paper.

2 Related work

There are two main areas of related work: community efforts for guiding users when adopting FAIR, and quality assessment in Research Objects.

2.1 Adopting the FAIR principles

Checking whether a resource follows the FAIR principles has been a subject of discussion in the last years. For instance, in [23], the authors propose a community-driven framework to assess the FAIRness of digital objects. This framework is based on: i) a collection of maturity indicators, ii) principle compliance tests, and iii) a module to apply those tests to digital resources. The proposed indicators may be an starting point to define which tests are needed for each type of resource [18].

[1] https://rd-alliance.org/group/fair-virtual-research-environments-wg/case-statement/fair-virtual-research-environments-vres

Community groups like FAIRassist.org[2] have compiled lists of guidelines and tools for assessing the FAIRness of digital resources. These range from self-assessment tools like questionnaires and checklists to semi-automated validators [8]. Examples of automated validation tools include the F-UJI Automated FAIR Data Assessment Tool[3] [7], FAIR Evaluator[4] and FAIR Checker[5] for datasets or digital objects; HowFairIs[6] [21] for code repositories; and and FOOPS [11] to assess ontologies. In our work, we build and incorporate some of these validators when calculating the FAIRness of all resources contained within a Research Object.

However, different assessment tools may have different interpretations for each principle. In [9] the authors compare different FAIRness evaluation tools (F-UJI, FAIR Evaluator and FAIR Checker), obtaining different scores and different level of detail when assessing resources. The study highlights the importance of transparency in the evaluation, making each metric and score explainable. Hence, we follow this principle in our work.

2.2 Quality assessment models for Research Objects

Research Object are semantically rich aggregations of research resources designed to facilitate the publication, sharing and reuse of such entities[1]. In order to evaluate Research Objects, [10] proposes ROHUB[7], a platform with metadata checklists to assess Research Object *completeness*. This indicator is calculated by checking the inclusion of metadata such as title, description, creator, publisher information, etc.; as well as the inclusion of certain resources like datasets, software or publications. However, this approach only takes into account the presence of general metadata, implementing only a subset of the FAIR principles. Our work builds on this work, taking into account every resource in a Research Object and assessing their compliance against all FAIR principles.

Lastly, FAIR Digital Objects (FDO) have been recently proposed as a self-contained, typed, machine-actionable data package [6]. However, to the best of our knowledge, there are no existing works addressing FDO FAIRness.

3 FAIROs: A Research Object FAIR assessment framework

Figure 1 shows an overview of the architecture of FAIROs. Our framework has two main components:

[2] https://fairassist.org/
[3] https://www.fairsfair.eu/f-uji-automated-fair-data-assessment-tool
[4] https://fairsharing.github.io/FAIR-Evaluator-FrontEnd
[5] https://github.com/IFB-ElixirFr/fair-checker
[6] https://github.com/fair-software/howfairis/
[7] https://www.rohub.org

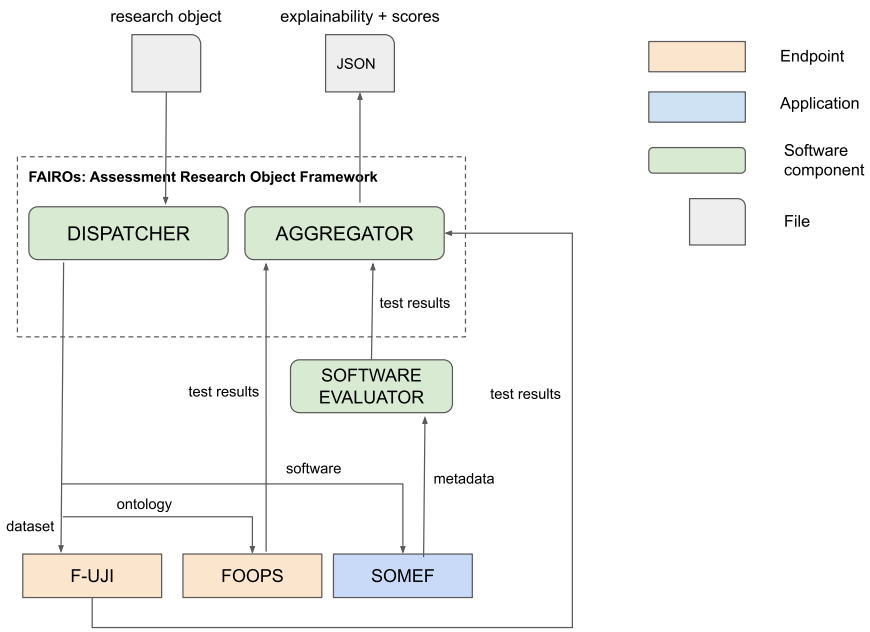

Fig. 1. FAIROs architecture

- Dispatcher. This component analyzes the Research Object metadata file and detects the different resources of the research object (datasets, software and ontologies). Once the resources are analyzed, the information is sent to the specific tool, as described in Section 3.1.
- Aggregator. This component collects the results from the tests executed and aggregates the information to calculate a FAIR score for the Research Object. Tests depend of resource's type and the FAIR principle, as described in Section 3.2.

FAIROs is open source and available on GitHub [8]. The source code of the version presented in this paper can be found in Zenodo [3].

3.1 Modules for individual resource FAIR assessment

In order to to measure the FAIRness of Research Objects, FAIROs integrates existing tools to assess individual datasets, software and ontologies. These tools have been selected based on: i) their ability to assess the analyzed resources automatically, without human intervention, ii) their accessibility, iii) ease of use and iv) they implement metrics described in scientific publications.

[8] https://github.com/oeg-upm/FAIR-Research-Object

Table 1. FAIR principles coverage by the FAIR assessment tools used in FAIROs

Service	F1	F2	F3	F4	A1	A1.1	A1.2	A2	I1	I2	I3	R1	R1.1	R1.2	R1.3
F-UJI	X	X	X	X	X				X		X	X	X	X	X
SOMEF	X	X											X	X	
FOOPS	X	X	X	X	X	X		X	X	X		X	X	X	

For **datasets**, we have deployed in our server an instance of the application F-UJI. This application provides a REST service to assess the FAIRness of datasets based on 16 of the 17 metrics defined in the FAIRsFAIR Data Objects Assessment Metrics.[9] The results of these tests are sent to the aggregator module.

For **ontologies**, we use the service provided by the Ontology Pitfall Scanner for FAIR (FOOPS!). This tool analyzes an ontology with several tests based on [12] [19] and returns a FAIRness score with an explanation of the results, categorized by each FAIR principle. The module dispatcher prepares the request and the response from FOOPS! is sent to the aggregator module.

Regarding **software** assessment, we have created a validator based on the principles identified in [16]. In order to extract metadata associated with a given code repository, we use the Software Metadata Extraction Framework (SOMEF) [17], which analyzes repository documentation to retrieve the license, description, installation instructions, requirements, versions, citation text or provenance information (authors, creation date). These metadata fields are particular relevant for reusability.

As shown in Figure 1, an additional component has been developed to execute SOMEF, run a set of tests to validate the extracted metadata and send the results to the aggregator component. Table 1 summarizes the FAIR principles [22] covered by the different components of our framework. For reference, we have enumerated the FAIR principles in Annex A.

The last component of the aggregator is in charge of evaluating the metadata of **Research Objects** themselves, by using RO-Crate specification [20] and the ro-crate-py library[5]. The tests we run to analyze the FAIR principles are detailed below:

- F1: We verify if the RO has a persistent identifier ['w3id.org', 'doi.org', 'purl.org', 'www.w3.org'].
- F2: We verify if the the following minimum metadata ['author', 'license', 'description'] are present in the ro-crate.
- F3: We verify that the hasPart elements exists and are described in the RO.

[9] https://www.fairsfair.eu/fairsfair-data-object-assessment-metrics-request-comments

- R1.1: We verify if there is at least one author, a datePublished and a citation in the root element.
- R1.2: We verify that all elements of the RO have the following fields: ['author', 'datePublished', 'citation']

3.2 Scoring Aggregations of resources in Research Objects

A Research Object is composed by resources, including the metadata of the Research Object itself. Each resource is evaluated against the FAIR principles (Findable, Accesible, Interoperable and Reusable), and each principle is evaluated by a set of tests (which depend on the type of the resource and the FAIR assessment tool used). The final FAIRness score associated with a Research Object may vary depending on the percentage of passed by each resource and on the design decisions used to aggregate the FAIR scores.

To illustrate this difference, we have defined two aggregation metrics in our framework to calculate the final FAIRness score of a Research Object. Note that the metrics are based on the tests defined for each principle and resource. Metrics do not change the tests executed, only how test results are dealt with in the aggregation formula. Our proposed metrics are:

- Global metric: calculated by formula (1). It represents the percentage of total passed tests. It doesn't take into account the principle to which a test belongs.
- FAIR average metric: calculated by formula (2). It represents the average of the passed tests ratios for each principle plus the ratio of passed tests used to evaluate the Research Object itself.

Both metrics are agnostic to the kind of resource analyzed. The score they produce ranges from [0 - 100].

$$total\_score_{global} = \frac{\#tests\_passed}{\#tests} \tag{1}$$

$$total\_score_{FAIR\_average} = \frac{\sum_{i \in G} \frac{\#test\_passed_i}{\#tests_i}}{\#G}, \tag{2}$$

where G represents the group of tests from the categories Findable (F), Accesible (A), Interoperable (I), Reusable (R) and Research Objects (RO).

3.3 Result format and visualization

FAIROs generates two outputs:

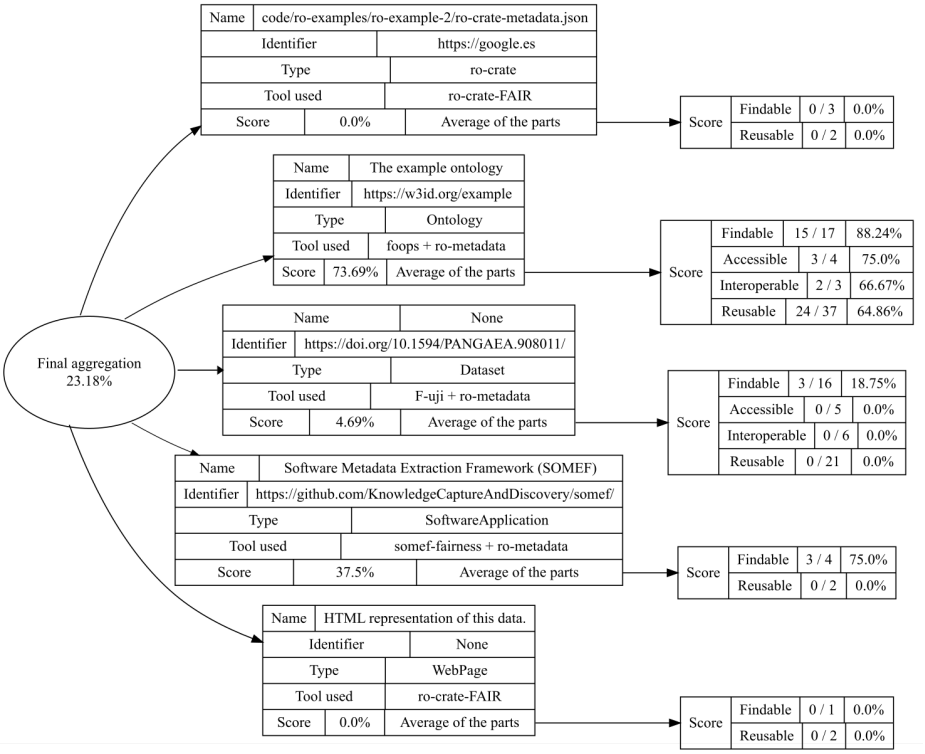

Fig. 2. Scores visualization diagram

- A JSON file with the executed tests and the final/intermediate scores generated by a specific metric. The final score is calculated based on the metric selected. Each resource is analyzed based on the tests defined for each FAIR principle. Also, the different tests executed for each principle are described.
- A visual diagram explaining the scores and each test. The diagram is generated in the Graphviz format.[10] An example can be seen in Figure 2.

The output JSON file format is composed by a list of components (resources of the research object), which defines the tests executed and the obtained scores. Inside each component, users can see the executed tests, and the explanation of their results. Listing 1.1 shows an example, showing some of the tests run for the Findable principle.

[10] https://graphviz.org/doc/info/lang.html

```
{"principle_id": "F3",
    "category_id": "Findable",
    "title": "Metadata clearly and explicitly include the identifier
        ↪ of the data they describe",
    "description": "This check verifies that the hasPart elements
        ↪ exists and are describe in the ro",
    "total_passed_tests": 1,
    "total_tests_run": 1,
    "status": "ok",
    "explanation": "All element identifiers exists"
}
```

Listing 1.1. Test execution with explanation

For each component, we include a summary of the tests categorized by principle. The final score calculated by the formula defined by the selected metric is included under the key *overall_score*, as shown in Listing 1.2.

```
"overall_score": {
        "description": "The score is calculated by adding all the
            ↪ scores of the different components together. All passed
            ↪ tests and all total tests are added together and then
            ↪ the percentage is calculated",
        "score": 37.5,
        "total_sum": {
            "total_passed_tests": 3,
            "total_run_tests": 8
        }
    }
```

Listing 1.2. Total score obtained for a sample Research Object

The output file format is the same for both of our proposed metrics.

4 Comparing FAIR Research Object Assessment Metrics

We have run a small experiment comparing the effect of our proposed metrics when assessing the FAIRness of more than 160 worflow-centric Research Objects. In all runs, we have used the release v0.0.1 of our framework [3].

The objective of our experiment was to analyze the behaviour and the impact of the two metrics defined in our framework on workflow-centric Research Objects. Therefore, we downloaded 168 Research Objects[13] from the workflowhub[11] platform. These Research Objects are composed by a workflow and other resources such as datasets and other files (images of

[11] https://www.workflowhub.eu

sketches in many cases). In total, 74 Research Objects contain one resource, 75 contain 2 resources, and 16 contain 3 or more resources.

As a result, each Research Object produces two JSON files, one for the global metric and another one for the FAIR_average metric.

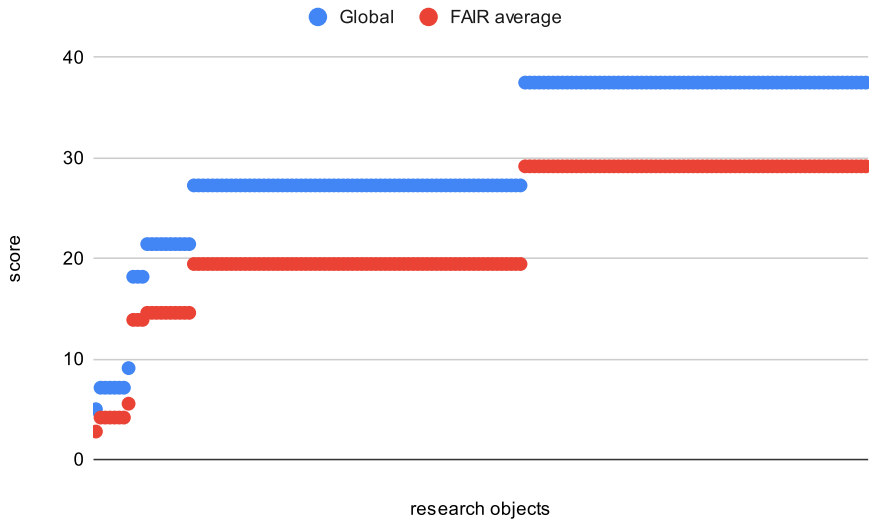

Fig. 3. Distribution of FAIR RO scores in over 160 Research Objects

The distribution of scores is depicted in Fig 3. The two scores are different, but consistent, i.e., when the score of a metric is increased, the scores of the other metric increase too. We also detect regular patterns in the results, influenced by the number of resources and the publication of the Research Objects in the same platform.

A histogram (See Fig. 4) with the frequency of the scores yields the same results. In summary:

- There are four clusters of scores among the analyzed Research Objects, due to their common number of resources.
- When the value of the score is low, data from the global metric are more dispersed than the FAIR average metric 1. This means, the scores of the first metric generates scores with more variance. In general, the standard deviation of metric 0 (6.31) is higher than the standard deviation of metric 1 (5.59).

All the files used and produced in our experiment are available online [14].

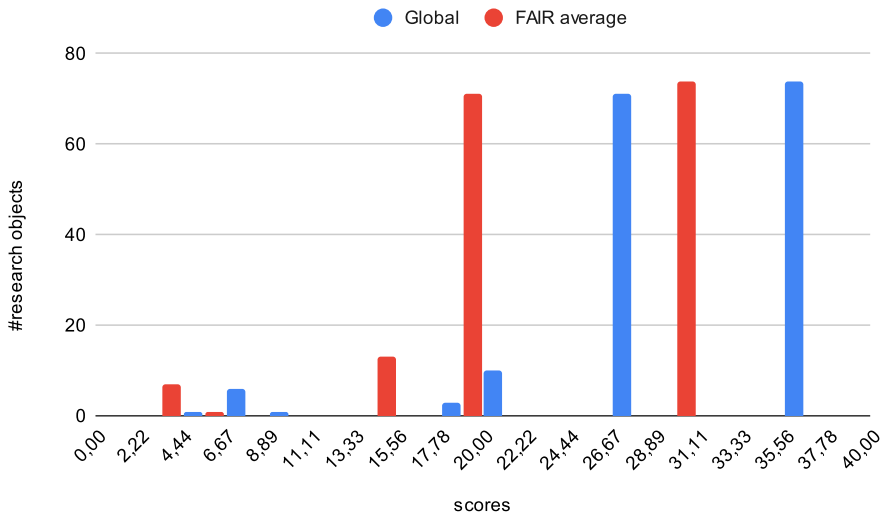

Fig. 4. Distribution of the number of ROs with a given score.

5 Discussion

Based on our initial results, we discuss several points we consider key for addressing the compliance of Research Objects against the FAIR principles:

Research Object resource metadata. As shown in Figure 3, many of the FAIRness scores obtained for the Research Objects analyzed are not high (scores are below 40 out of 100 for the best metric). This is due, in part, to the lack of metadata for the resources included in a Research Object. However, additional work is needed to identify whether this behavior is commonplace in Research Objects belonging to other common platforms or not.

Research Object FAIRness as a proxy of RO quality? Figure 3 shows how by using different aggregation metrics, the FAIRness score of a RO can change significantly (even if scores are consistent with each other). This has two main implications: 1) scores may be used as a guideline to improve the quality of a Research Object, but should not act as a replacement for its quality; and 2) it is key to explain the aggregation method used to produce any Research Object FAIRness score. The objective of the scoring system should not be to produce a ranking, but become a mechanism to improve the FAIRness of Research Objects.

Resource tests in FAIR scores. Each type of resource (datasets, ontologies, etc.) has different number of tests, so the weight in the final score will depend on the proportion of resources defined in the Research Object.

For example, imagine two Research Objects; one with 4 datasets and another with 6 datasets and 5 software packages. Since the number of tests to assess the FAIRness of software is much more lower than the number of tests for datasets, the score of the first Research Object will likely be much better than the second, even when the second Research Object will be more comply to more FAIR principles, in percentage.

FAIROs limitations. While our framework provides a first step towards assessing Research Object FAIRness, it also present some limitations. For example, FAIROs makes some assumptions about concrete resources for which there is no FAIR validator (e.g., instance images and sketches may be considered datasets, which may miss metadata embedded in the files)

6 Conclusions and Future Work

In this work we have presented FAIROs, a framework for assessing the compliance of Research Objects against the FAIR principles. FAIROs integrates existing tools for assessing the FAIRness of datasets, software and ontologies; and provides two different aggregation metrics for calculating Research Object FAIRness scores.

Our results show a potential imbalance in the FAIRness score obtained depending on the number of tests run by each principle and resource. For example, if a dataset assessment runs more tests than software, or if more tests are run for the findable principle than for interoperability .

As for future work, we plan to to increment the number of tests run in the Research Object to get a more detailed impression of the Research Object FAIRness. These tests will be accompanied with a collection of recommendations to improve the FAIRness of the target resource.

Also, we would like to open the aggregation metrics to users. Users could define their own metrics based on their preferences. For example, they can assign more weight to some resources/principles than others, according to their use cases.

From the architecture point of view, we want to transform our framework into a service, giving support to the community. Our plan considers integrating FAIROs in the RELIANCE Research Object platform,[12] in order to test our work against hundreds of Research Objects in the Geosciences domain.

References

1. Bechhofer, S., De Roure, D., Gamble, M., Goble, C., Buchan, I.: Research objects: Towards exchange and reuse of digital knowledge. Nature Precedings pp. 1–1 (2010). https://doi.org/https://doi.org/10.1038/npre.2010.4626.1

[12] https://reliance.rohub.org/

2. Belhajjame, K., Zhao, J., Garijo, D., Gamble, M., Hettne, K., Palma, R., Mina, E., Corcho, O., Gómez-Pérez, J.M., Bechhofer, S., Klyne, G., Goble, C.: Using a suite of ontologies for preserving workflow-centric research objects. Journal of Web Semantics **32**, 16–42 (2015). https://doi.org/https://doi.org/10.1016/j.websem.2015.01.003
3. Benitez, A., González, E., Garijo, D.: FAIROs: A framework to assess FAIR principles in research objects (May 2022). https://doi.org/10.5281/zenodo.6599423, https://doi.org/10.5281/zenodo.6599423
4. Collins, S., Genova, F., Harrower, N., Hodson, S., Jones, S., Laaksonen, L., Mietchen, D., Petrauskaitė, R., Wittenburg, P.: Turning fair into reality: Final report and action plan from the european commission expert group on fair data (2018)
5. De Geest, P., Droesbeke, B., Eguinoa, I., Gaignard, A., Huber, S., Leo, S., Pireddu, L., Rodríguez-Navas, L., Sirvent, R., Soiland-Reyes, S.: ro-crate-py (5 2022). https://doi.org/10.5281/zenodo.3956493
6. De Smedt, K., Koureas, D., Wittenburg, P.: Fair digital objects for science: From data pieces to actionable knowledge units. Publications **8**(2) (2020). https://doi.org/10.3390/publications8020021
7. Devaraju, A., Huber, R.: F-uji - an automated fair data assessment tool (Oct 2020). https://doi.org/10.5281/zenodo.4063720
8. Devaraju, A., Mokrane, M., Cepinskas, L., Huber, R., Herterich, P., de Vries, J., Akerman, V., L'Hours, H., Davidson, J., Diepenbroek, M.: From conceptualization to implementation: Fair assessment of research data objects. Data Science Journal **20**(1), 1–14 (2021)
9. Dumontier, M.: A comprehensive comparison of automated fairness evaluation tools (2022), http://ceur-ws.org/Vol-3127/paper-6.pdf
10. Garcia-Silva, A., Gomez-Perez, J.M., Palma, R., Krystek, M., Mantovani, S., Foglini, F., Grande, V., De Leo, F., Salvi, S., Trasatti, E., Romaniello, V., Albani, M., Silvagni, C., Leone, R., Marelli, F., Albani, S., Lazzarini, M., Napier, H.J., Glaves, H.M., Aldridge, T., Meertens, C., Boler, F., Loescher, H.W., Laney, C., Genazzio, M.A., Crawl, D., Altintas, I.: Enabling fair research in earth science through research objects. Future Generation Computer Systems **98**, 550–564 (2019). https://doi.org/https://doi.org/10.1016/j.future.2019.03.046
11. Garijo, D., Corcho, O., Poveda-Villalón, M.: FOOPS!: An ontology pitfall scanner for the fair principles. International Semantic Web Conference (ISWC) 2021: Posters, Demos, and Industry Tracks **2980** (2021), http://ceur-ws.org/Vol-2980/paper321.pdf
12. Garijo, D., Poveda-Villalón, M.: Best practices for implementing fair vocabularies and ontologies on the web. In: Giuseppe Cota, M.D., Pozzato, G.L. (eds.) Applications and Practices in Ontology Design, Extraction, and Reasoning. IOS Press, Netherlands (2020). https://doi.org/https://doi.org/10.3233/SSW200034
13. Gonzalez, E., Benitez, A., Garijo, D.: Tpdl2022 - experiment - research objects data (May 2022). https://doi.org/10.5281/zenodo.6595409, https://doi.org/10.5281/zenodo.6595409
14. Gonzalez, E., Benitez, A., Garijo, D.: Tpdl2022 - experiment - results (May 2022). https://doi.org/10.5281/zenodo.6595466, https://doi.org/10.5281/zenodo.6595466

15. Katz, D.S., Gruenpeter, M., Honeyman, T.: Taking a fresh look at fair for research software. Patterns **2**(3), 100222 (2021). https://doi.org/https://doi.org/10.1016/j.patter.2021.100222
16. 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., et al.: Towards fair principles for research software. Data Science **3**(1), 37–59 (2020)
17. Mao, A., Garijo, D., Fakhraei, S.: Somef: A framework for capturing scientific software metadata from its documentation. In: 2019 IEEE International Conference on Big Data (Big Data). pp. 3032–3037 (2019). https://doi.org/10.1109/BigData47090.2019.9006447, http://dgarijo.com/papers/SoMEF.pdf
18. de Miranda Azevedo, R., Dumontier, M.: Considerations for the conduction and interpretation of fairness evaluations. Data Intelligence **2**(1-2), 285–292 (2020)
19. Poveda-Villalón, M., Espinoza-Arias, P., Garijo, D., Corcho, O.: Coming to terms with FAIR ontologies: A position paper. In: Proceedings of the 22nd International Conference on Knowledge Engineering and Knowledge Management (EKAW 2020). pp. 255–270 (2020). https://doi.org/10.1007/978-3-030-61244-3_18
20. Soiland-Reyes, S., Sefton, P., Crosas, M., Castro, L.J., Coppens, F., Fernández, J.M., Garijo, D., Grüning, B., La Rosa, M., Leo, S., et al.: Packaging research artefacts with ro-crate. Data Science pp. 1–42 (2022). https://doi.org/https://doi.org/10.3233/DS-210053
21. Spaaks, J.H., Kuzak, M., Martinez-Ortiz, C., van Werkhoven, B., Etuk, E., Saladi, S., Holding, A., Tjong Kim Sang, E., Diblen, F., Verhoeven, S.: howfairis (Mar 2021). https://doi.org/10.5281/zenodo.4591110, https://doi.org/10.5281/zenodo.4591110
22. Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.W., da Silva Santos, L.B., Bourne, P.E., et al.: The fair guiding principles for scientific data management and stewardship. Scientific data **3**(1), 1–9 (2016)
23. Wilkinson, M.D., Dumontier, M., Sansone, S.A., Bonino da Silva Santos, L.O., Prieto, M., Batista, D., McQuilton, P., Kuhn, T., Rocca-Serra, P., Crosas, M., et al.: Evaluating fair maturity through a scalable, automated, community-governed framework. Scientific data **6**(1), 1–12 (2019)

A FAIR principles

Table 2. FAIR principles

Principle	Description
F1	(meta)data are assigned a globally unique and persistent identifier
F2	data are described with rich metadata (defined by R1 below)
F3	metadata clearly and explicitly include the identifier of the data it describes
F4	(meta)data are registered or indexed in a searchable resource
A1	(meta)data are retrievable by their identifier using a standardized communications protocol
A1.1	the protocol is open, free, and universally implementable
A1.2	the protocol allows for an authentication and authorization procedure, where necessary
A2	metadata are accessible, even when the data are no longer available
I1	(meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2	(meta)data use vocabularies that follow FAIR principles
I3	(meta)data include qualified references to other (meta)data
R1	meta(data) are richly described with a plurality of accurate and relevant attributes
R1.1	(meta)data are released with a clear and accessible data usage license
R1.2	(meta)data are associated with detailed provenance
R2	(meta)data meet domain-relevant community standards