



26th International Conference on Science and Technology Indicators
"From Global Indicators to Local Applications"

#STI2022GRX

Full paper

STI 2022 Conference Proceedings

Proceedings of the 26th International Conference on Science and Technology Indicators

All papers published in this conference proceedings have been peer reviewed through a peer review process administered by the proceedings Editors. Reviews were conducted by expert referees to the professional and scientific standards expected of a conference proceedings.

Proceeding Editors

Nicolas Robinson-Garcia
Daniel Torres-Salinas
Wenceslao Arroyo-Machado



Citation: Tintero-Cañadas, E., Sixto-Costoya, A., Aleixandre-Benavent, R., Castelló-Cogollos, L., & Valderrama-Zurián, J. C. (2022). Analysis of FAIR principles in datasets in the substance abuse field. A comparative study. In N. Robinson-Garcia, D. Torres-Salinas, & W. Arroyo-Machado (Eds.), *26th International Conference on Science and Technology Indicators*, STI 2022 (sti22148). <https://doi.org/10.5281/zenodo.6975666>

Copyright: © 2022 the authors, © 2022 Faculty of Communication and Documentation, University of Granada, Spain. This is an open access article distributed under the terms of the [Creative Commons Attribution License](#).

Collection: <https://zenodo.org/communities/sti2022grx/>

26th International Conference on Science and Technology Indicators | STI 2022

“From Global Indicators to Local Applications”

7-9 September 2022 | Granada, Spain

#STI22GRX

Analysis of FAIR principles in datasets in the substance abuse field. A comparative study¹

Eva Tendo-Cañadas^{*}, Andrea Sixto-Costoya^{*,****}, Rafael Aleixandre-Benavent^{*,**}, Lourdes Castelló-Cogollos^{*,****} and Juan Carlos Valderrama-Zurián^{*,***}

^{*}*etenca@alumni.uv.es; andrea.sixto@uv.es, Rafael.Aleixandre@uv.es; lourdes.castello@uv.es; Juan.Valderrama@uv.es*

Unidad de Información e Investigación Social y Sanitaria (Grupo UISYS), Universitat de València. Plaça de Cisneros, 4, 46003, Valencia (Spain)

^{**}Ingenio (CSIC-Universitat Politècnica de València), Edificio 8E, Cam. de Vera, 46022, Valencia (Spain)

^{***}Departamento de Historia de la Ciencia y Documentación (Universitat de València), Avendia Blasco Ibañez, 15, 46010, Valencia (Spain)

^{****}Departamento de Sociología y Antropología Social (Universitat de València), Av. Tarongers, 4b, 46021, Valencia (Spain)

^{*****}IUMPA, Instituto de Matemáticas Pura y Aplicada (Universidad Politécnica de Valencia), Camino de Vera, 46022, Valencia (Spain)

Introduction

Currently there is no debate about the usefulness of the FAIR principles, since for the own good of the scientific community research data must be Findable, Accessible, Interoperable and Reusable (Mons, Barend; Schultes, Erik; Fenghong, Liu; Jacobsen, 2020). FAIR principles can be applied over any digital object; it can be datasets, software, or any digital result of a research (Devaraju et al., 2020).

After the theoretical principles, several evaluation tools have been developed to check either manually or automatically whether the datasets follow the principles or not. Among them, there are two particularly interesting tools, as they allow the scientific community to automatically assess the degree of compliance with the FAIR principles based on a series of metrics. These tools are the *F-UJI Automated FAIR Data Assessment Tool* (<https://www.fairsfair.eu/f-uji-automated-fair-data-assessment-tool>) that is the result of the FAIRsFAIR Horizon 2020 project and the *FAIR-Checker* (<https://fair-checker.france-bioinformatique.fr/>) developed by the French Institute for Bioinformatics (IFB).

The objective of this study is to analyse the datasets related with the substance abuse field, comparing *F-UJI Automated FAIR Data Assessment Tool* and *FAIR-Checker*.

¹ Acknowledgement: Andrea Sixto-Costoya is beneficiary of a Margarita Salas grant for postdoctoral researchers (ref MS21-020).

Methodology

In the first phase, a bibliographic search related with the substance abuse field was conducted in the Data Citation Index of the Web of Science. The search was limited from 2017 to 2021 and filtered by document type “dataset”. 766 were obtained as a result.

In the second phase, the information was downloaded and the database created. No duplicated records were found. The information downloaded for each dataset was: title, ID, name of the repository. Subsequently 621 of the 766 datasets were analysed in both tools comparing the results obtained by each one. 145 datasets could not be analysed because they do not have a DOI.

In both tools each principle gets a final assessment. On one hand we have the *F-UJI Automated FAIR Data Assessment Tool*, which final assessments may vary between: incomplete, initial, moderate and advanced. On the other hand, we have the *FAIR-Checker* tool, in which the final assessment of each principle is expressed throughout a percentage. The *F-UJI Automated FAIR Data Assessment Tool* takes into consideration in their evaluation 16 metrics out of the 17 metrics recommended, whereas *FAIR-Checker* only evaluates 13.

Results

When comparing both tools there are some aspects important to mention, such as the score given to each of the metrics. In the *F-UJI Automated FAIR Data Assessment Tool* the score depends on each of the principles, in the Findable principle the score may vary between 0 and 7, in the Accessible principle the score may vary between 0 and 3, in the Interoperable principle the score may vary between 0 and 4, and in the Reusable principle the score may vary between 0 and 10. Meanwhile in the *FAIR-Checker* tool each metric values always are between: None, 0, 1, 2.

In the Tables 1 and 2 we can see the comparison between the final assessments obtained by all datasets analysed in each tool.

Table 1. Fair Checker final assessments

	0%	25%	33'3%	50%	60%	66'7%	75%	80%	100%
Findable	13	62	0	9	0	0	228	0	309
Accessible	0	0	0	0	0	0	0	0	621
Interoperable	75	0	0	0	227	0	0	310	9
Reusable	84	0	237	0	0	300	0	0	0

Table 2. F-UJI Automated FAIR Data Assessment Tool final assessments

	Incomplete	Initial	Moderate	Advanced
Findable	0	12	481	128
Accessible	17	293	226	85
Interoperable	49	298	273	1
Reusable	17	325	247	32

As we can see in the Findable principle, the first tool determines that 309 datasets comply 100% with this principle, while with the second tool only 128 are in the advanced category. In the case of the Accessible principle, we find the greatest difference, with *Fair Checker* it is determined that every dataset complies 100%, meanwhile with *F-UJI Automated FAIR Data*

Assessment Tool we find only 85 datasets to be in the advanced category and the majority of the datasets to be assigned the initial category. In Interoperable principle we can see how, even though in the first tool only 75 datasets comply 0% with the principle and the rest are found after 60% compliance; in the second tool 49 are incomplete and 298 initial. We see the most similar evaluations between both tools are found in the Reusable principle, since the first tool places most datasets between 33.3% and 66.7%, and the second tool between Initial and Moderate.

The Figures 1 and 2 show the visual comparison of both tools when evaluating the same dataset.

Figure 1. F-UJI Automated FAIR Data Assessment Tool’s analysis of a dataset

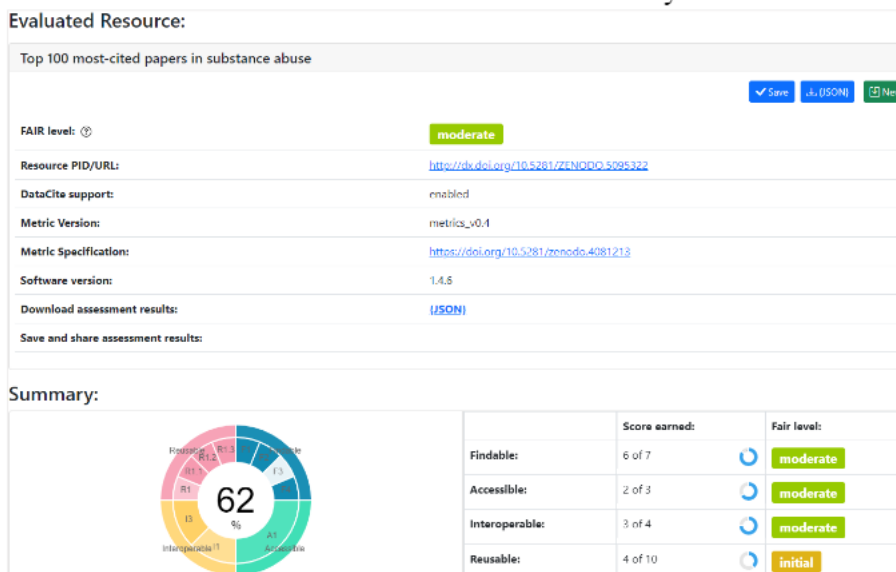
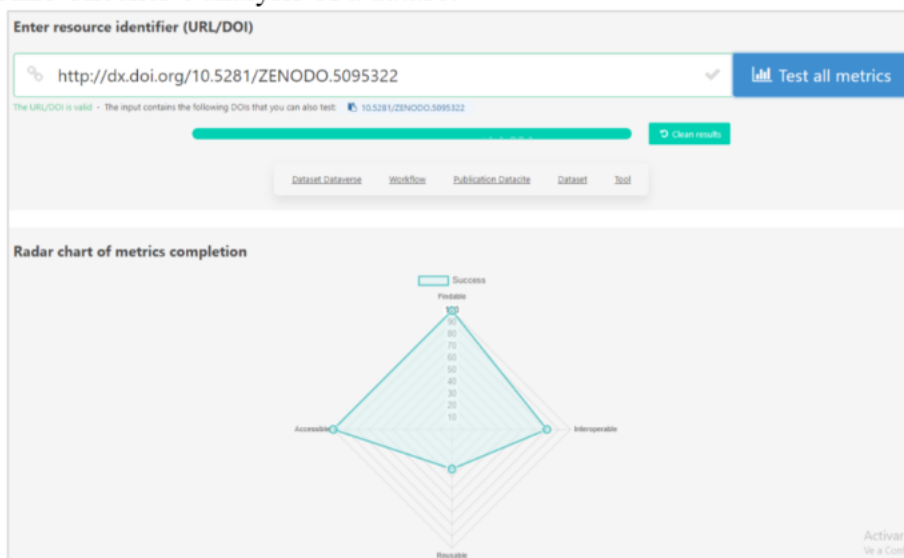


Figure 2. FAIR-Checker’s analysis of a dataset



Conclusions

When conducting the analysis in *FAIR-Checker*, the results obtained are binary, it either complies with the metric or it does not, meanwhile in the *F-UJI Automated FAIR Data Assessment Tool*'s analysis it determines a degree of compliance, this shows how the second tool is more accurate.

It is also important to take into consideration that the only tool that gives a final assessment of the dataset itself throughout the analysis, is the *F-UJI Automated FAIR Data Assessment Tool* and the time this tool takes carrying out the analysis of each dataset is considerably shorter.

Based on the results obtained we can conclude that *F-UJI Automated FAIR Data Assessment Tool* is more precise in the evaluation of datasets.

References

Devaraju, A., Huber, R., Mokrane, M., Herterich, P., Cepinskas, L., de Vries, J., L'Hours, H., Davidson, J., & White, A. (2020, October). FAIRsFAIR data object assessment metrics. Zenodo. Retrieved from <https://doi.org/10.5281/ZENODO.4081213>

Mons, B., Schultes, E., Liu, F., & Jacobsen, A. (2020). The FAIR Principles: First Generation Implementation Choices and Challenges. *Data Intelligence*, 2(1-2), 1-9. https://doi.org/10.1162/dint_e_00023