FAIRsoft - A practical implementation of FAIR principles for research software

Eva Martín del Pico*, Salvador Capella-Gutiérrez† Josep Lluís Gelpí Buchaca‡,
*†‡Barcelona Supercomputing Center, Barcelona, Spain
‡Universitat de Barcelona, Barcelona, Spain
Email: *eva.martin@bsc.es, †salvador.capella@bsc.es, ‡gelpi@ub.edu

***Keywords*—FAIR, Open Science, Research Software**

I. Extended Abstract

Computational tools are increasingly becoming constitutive parts of scientific research, from experimentation and data collection to the dissemination and storage of results. Unfortunately, however, research software is not subjected to the same requirements as other methods of scientific research: being peer-reviewed, being reproducible and allowing one to build upon another's work. This situation is detrimental to the integrity and advancement of scientific research, leading to computational methods frequently being impossible to reproduce and/or verify [1]. Moreover, they are often opaque, direcly unavailable or impossible to use by others [2]. One step to address this problem could be formulating a set of principles that research software should meet to ensure its quality and sustainability, resembling the FAIR (Findable, Accessible, Interoperable and Reusable) Data Principles [3]. The FAIR Data Principles were created to solve similar issues affecting scholarly data, namely great difficulty of sharing and accessibility, and are currently widely recognized accross fileds. We present here FAIRsoft, our initial effort to assess the quality of research software using a FAIR-like framework, as a first step towards its implementation in OpenEBench [4], the ELIXIR benchmarking platform.

A. Proposal for a FAIRsoft scoring system

We analysed the FAIR principles for research software [5] to formulate an initial strategy to obtain a quantitative evaluation. The result is a set of measurable indicators generated following a two steps approach. In a first step, we derived a number of requirements that software must fulfil in order to be *Findable, Accessible, Interoperable* and *Re-Usable*, respectively. We call these properties high-level indicators. The second step took us to the desired degree of granularity through low-level indicators. A low-level indicator is one condition that contributes to a software meeting a high-level indicator. To allow for a practical evaluation, each low-level indicator is associated with a well defined evaluation procedure.

B. Measurement of software FAIRness

For each tool, FAIRsoft indicators can be measured using, in most cases, metadata from more than one reference source, which must be accessible and findable by any user in order to be valid. These sources, that include software registries and repositories, e-Infrastructures, software homepages and journal publications, can be extended to increase the indicators coverage for individual entries. We integrated metadata from Bio.tools, Bioconda, Bioconductor, Galaxy ToolShed, SourceForge , and Galaxy Europe as primary sources to discover tools and retrieve an initial collection of metadata. This collection was enriched mining secondary sources: Github, Bitbucket, OpenEBench, PubMed, Europe PMC and Wikidata. The resulting set of metadata was restructured to fit a common data model and then integrated by software instance.

Finally, FAIRsoft scores were computed for each software instance. To this end, low-level indicators were calculated and subsequently combined to generate high-level indicator scores. A weighting scheme was designed and implemeted to reflect the varying importance of individual low-level indicators.

C. Results and Discussion

We obtained a collection of 43,973 unique software instances, 71.4% of which were enriched with metadata from more than one source. The number of instances with available metadata varied greatly among sources, as well as the type of information they provided.

Results for the general four FAIR principles for research software are heterogeneous, as shown in Fig 1. However, patterns can be easily identified among them. Indicator scores for the *Findability* of research software are higher than for other principles, with a remarkable 44.0 % of published software. A lack of structured metadata is the main reason for lower *Findability* scores. The actual usability of the software includes indicators from both *Accessibility* and *Re-Usability* principles. We found that 15.6% of analysed instances score optimally for the main indicator of *Accessibility* (*Existance of a downloadable, buildable or accessible working version of the software*) and 47.6% for the main indicator of *Re-usability* (*Existance of License*), with an impressive predomination of Open Source Licenses (86.0%). Finally, *Interoperability* is the lowest scoring principle, partially due to the absense of agreed standards on how to represent software interoperability, making it difficult to define automatically measurable indicators. Nonetheless, exceptional cases exist, and there is abundant literature on what software interoperability is and how it can be measured both in terms of working with other research software as part of analytical workflows [6], and in terms of interoperating with underlying software components such as software libraries. Indeed, we found information about data types and formats, as well as about dependencies, is only structured when it is required to be machine-readable as in packages repositories (instances enriched with metadata from Bioconda and Bioconductor).

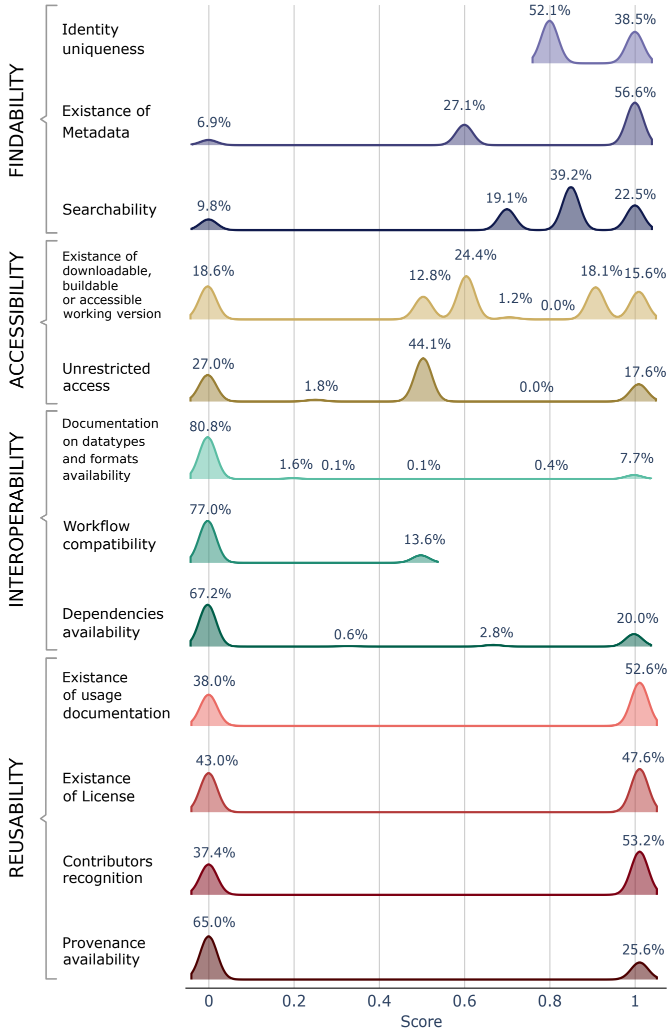

Fig. 1. High-level indicators scores of instances. Each possible score is labelled with the percentage of instances scoring it. Although scores are discrete values, they are shown as density plots for clarity.

D. Final remarks

The proposed FAIRsoft indicators can definitely contribute towards the consolidation of the FAIR principles for research software, driven by community initiatives, e.g. FAIR4RS, a joint work by the Research Data Alliance (RDA), Research Software Alliance (ReSA) and FORCE11. Indeed, this first generation of indicators should serve to improve automated measurement algorithms as well as to reflect the contribution of specific indicators to the four general principles.

This work should be considered an initial effort for having a quantitative overview of the common practices for developing research software in the Life Sciences. The ultimate goal is to contribute to the reproducibility and reliability of scientific outcomes by putting the focus on one of the key elements for success: Research Software. Periodical assessment of research software FAIRness would allow researchers to understand existing and emerging trends and can serve to identify areas of potential improvement for facilitating the generation of high-quality research software.

II. Acknowledgment

We thank the members of the INB Computational and Coordination Nodes at BSC-CNS as well as the members of the ELIXIR Tools Platform Best Practices group. This work was supported by the IMI FAIRplus Project, funded by the Innovative Medicines Initiative Joint Undertaking under grant agreement No 802750.

References

[1] D. B. Allison, A. W. Brown, B. J. George, and K. A. Kaiser, "Reproducibility: A tragedy of errors," *Nature*, vol. 530, no. 7588, pp. 27–29, Feb. 2016. [Online]. Available: http://www.nature.com/articles/530027a

[2] A. Morin, J. Urban, P. D. Adams, I. Foster, A. Sali, D. Baker, and P. Sliz, "Shining Light into Black Boxes," *Science*, vol. 336, no. 6078, pp. 159–160, Apr. 2012, publisher: American Association for the Advancement of Science. [Online]. Available: https://www.science.org/doi/10.1126/science.1218263

[3] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, J. Bouwman, A. J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C. T. Evelo, R. Finkers, A. Gonzalez-Beltran, A. J. Gray, P. Groth, C. Goble, J. S. Grethe, J. Heringa, P. A. t Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S. J. Lusher, M. E. Martone, A. Mons, A. L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S.-A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M. A. Swertz, M. Thompson, J. van der Lei, E. van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao, and B. Mons, "The FAIR Guiding Principles for scientific data management and stewardship," *Sci Data*, vol. 3, no. 1, p. 160018, Dec. 2016. [Online]. Available: http://www.nature.com/articles/sdata201618

[4] S. Capella-Gutierrez, D. d. l. Iglesia, J. Haas, A. Lourenco, J. M. Fernndez, D. Repchevsky, C. Dessimoz, T. Schwede, C. Notredame, J. L. Gelpi, and A. Valencia, "Lessons Learned: Recommendations for Establishing Critical Periodic Scientific Benchmarking," bioRxiv, Tech. Rep., Aug. 2017, section: New Results Type: article. [Online]. Available: https://www.biorxiv.org/content/10.1101/181677v1

[5] A.-L. Lamprecht, L. Garcia, M. Kuzak, C. Martinez, R. Arcila, E. Martin Del Pico, V. Dominguez Del Angel, S. van de Sandt, J. Ison, P. A. Martinez, P. McQuilton, A. Valencia, J. Harrow, F. Psomopoulos, J. L. Gelpi, N. Chue Hong, C. Goble, and S. Capella-Gutierrez, "Towards FAIR principles for research software," *DS*, vol. 3, no. 1, pp. 37–59, Jun. 2020. [Online]. Available: https://content.iospress.com/articles/data-science/ds190026

[6] C. Goble, S. Cohen-Boulakia, S. Soiland-Reyes, D. Garijo, Y. Gil, M. R. Crusoe, K. Peters, and D. Schober, "FAIR Computational Workflows," *Data Intelligence*, vol. 2, no. 1-2, pp. 108–121, 2020.

Eva Martín del Pico received her BSc degree in Biochemistry from Autonomous University of Madrid (UAM), Spain in 2016. She completed her MSc degree in Bioinformatics for Health Sciences from Pompeu Fabra University, Spain in 2019. Since then, she is a PhD student at the Coordination Node of the Spanish National Bioinformatics Institute (INB) at Barcelona Supercomputing Center (BSC).