Making and assessing FAIR biomedical data

Chang Sun[a], Vincent Emonet[a], and Michel Dumontier[a]
[a] *Institute of Data Science, Faculty of Science and Engineering, Maastricht University*

Abstract. The FAIR Guiding Principles have achieved global pre-eminence in the context of modern research data management, with over 20 studies concerned with FAIR have been published at MIE in the past two years alone. However, the technology-agnostic nature of the FAIR principles leaves uncertainty in the implementation and assessment of FAIR digital resources. This tutorial will shed light on the intent of the FAIR principles, outline approaches to create biomedical data, and guide participants to automatically assess the FAIRness of digital resources. Participants will be able to compare and contrast existing metrics and testing frameworks, and further develop a custom test suite based on domain-specific needs. Participants will gain a critical understanding of the nature and application of the FAIR principles in biomedicine.

1. Topic Introduction

The FAIR Guiding Principles (Findable, Accessible, Interoperable, and Reusable) [1] have achieved global pre-eminence with respect to modern research data management, in which biomedical community and the healthcare sector is no exception. Over the past two years, MIE has featured over 20 studies that specifically address aspects of the FAIR principles. These pertain to the creation of new data, algorithms, models, and workflows, the description of experiences in the *FAIRification* medical or clinical resources, and the development of new tools for making resources FAIR [2-10]. Making resources FAIR is expected to facilitate their discovery, reuse, and have important implications in the development of trustworthy AI and responsible decision-making systems.

However, making FAIR data is not straightforward, and many questions arise in biomedicine and health. A part of the challenge lies in the fact that the FAIR principles are largely aspirational – they articulate functional requirements for long term reusability, but do not provide specific technical blueprints needed for precise implementations, instead leaving this to communities to solve [11,12]. Additionally, there is need to objectively assess whether any FAIRification efforts really do improve resource *FAIRness* along general and community-specific technical standards. A number of manual, semi-automated, and fully automated FAIRness assessment frameworks and tools (e.g. FAIR Maturity Indicators [13], FAIR Evaluator [14], FAIR Enough [15], F-UJI [16], FAIR-Checker [17]) have emerged to give communities the means to assess the quality of their efforts and to provide additional guidance. However, these tools and frameworks vary widely in their interpretation of the FAIR principles and what guidance they provide [18], causing additional confusion and ambiguity for funders, researchers, and developers alike.

This tutorial will provide clear instruction to participants on what FAIR is really about, what steps are needed to create FAIR biomedical/health data, how to assess the FAIRness of online digital resources, and how to create and use domain-specific FAIRness tests. Participants will be able to explain the differences between automatic FAIRness assessment tools, correctly interpret their outputs, and devise strategies to address identified shortcomings. Participants will discuss the limitations of current metric collections and explore the development of a domain-specific (rare

disease) FAIR metric test as a FAIR API using a Python library. Finally, participants will execute a new FAIRness assessment using these custom tests.

2. Tutorial Format (3 hours)

Part 1: (10mins + 40mins) Welcome + Introduction

- Introduce the FAIR Guiding Principles
- Discuss FAIR data recipes and corresponding implementation details
- Describe FAIR assessment in terms of approaches, metrics, and tools
- Present FAIR evaluation tools and services

Part 2: (10mins break + 60mins) Assess FAIRness of select biomedical resources

- Guide participants in assessing selected biomedical digital resources using particular tools and metric tests
- Jointly analyze and discuss the significance, advantages/limitations of the assessments
- Discuss concrete approaches to improve the FAIRness of their resources, at least as it pertains to the implemented metric tests

Part 3: (10mins break + 45 mins) Create a custom FAIR metrics test

- Discuss the need and potential of community-based/domain-specific metric tests and collections, with a focus on emergent standards in the rare disease community
- Describe how custom evaluation tests can be created using fair-test library [19]
- Guide participants to create, register, and execute a domain-specific FAIRness test and metric collection

Part 4: (15mins) Discussion + Closing

- Participants share their thoughts and experience
- Future of FAIR evaluation in the biomedical informatics community

3. Target Group

Researchers, developers, students, data managers, data stewards, and anyone who is interested in research data management, FAIR guiding principles, making their digital resources FAIR, FAIRness assessment and improvement, and building biomedical domain-specific FAIRness evaluation tests for the community.

4. Learning Objectives

1. To learn how to make digital resources FAIR and to improve their FAIRness
2. To understand automated FAIRness evaluation and executable metrics
3. To use existing tools to perform FAIRness evaluation
4. To modify or develop a custom collection of evaluation metrics
5. To create and publish new FAIR tests based on domain-specific requirements

5. Expected outcome

1. To raise awareness of the approaches and importance of applying FAIR
2. To raise awareness of existing metrics, tests, and tools for FAIRness evaluation

3. To guide attendants to evaluate their own resources and improve the FAIRness of existing resources
4. To guide attendants to publish their domain-specific FAIR Evaluation tests
5. To encourage shared efforts and domain requirements from the community to evaluate their domain-specific data

6. Speakers Bio

Chang Sun, a PhD candidate, research in privacy-preserving personal health data mining, personal data management tools for health research, FAIRness Evaluation, synthetic data generation for health and medical data. Contribution: Part 1-2.
Vincent Emonet, a research developer, is working on building biomedical ontologies and FAIR knowledge graphs. He also worked on the development of various tools to evaluate the FAIRness of published resources Contribution: Part 2-3.
Dr. Michel Dumontier, a Distinguished Professor of Data Science, co-Founder of the FAIR Principles, an expert in bioinformatics, biomedical informatics, knowledge graphs, and semantic technologies including ontologies and linked data. Contribution: Part 1-4.

References

[1] Wilkinson, Mark D., et al. "The FAIR Guiding Principles for scientific data management and stewardship." *Scientific data* 3.1 (2016): 1-9.
[2] Löbe M, et al. Problems in FAIRifying Medical Datasets. Stud Health Technol Inform. 2020 Jun 16;270:392-396. doi: 10.3233/SHTI200189. PMID: 32570413.
[3] Meloni, Vittorio, et al. "FAIRifying Clinical Studies Metadata: A Registry for the Biomedical Research." *Public Health and Informatics*. IOS Press, 2021. 779-783.
[4] Frexia, Francesca, et al. "openEHR is FAIR-Enabling by Design." *Public Health and Informatics*. IOS Press, 2021. 113-117.
[5] Jauer, Malte-Levin, and Thomas M. Deserno. "Data provenance standards and recommendations for FAIR data." *Digital Personalized Health and Medicine* (2020): 1237-1238.
[6] Gencturk, Mert, et al. "End User Evaluation of the FAIR4Health Data Curation Tool." *Public Health and Informatics*. IOS Press, 2021. 8-12.
[7] Stausberg, Jürgen, and Sonja Harkener. "Metadata of Registries: Results from an Initiative in Health Services Research." *Public Health and Informatics*. IOS Press, 2021. 18-22.
[8] Looten, Vincent, and Marie Simon. "Impact Analysis of the Policy for Access of Administrative Data in France: A Before-After Study." *Digital Personalized Health and Medicine*. IOS Press, 2020. 1133-1137.
[9] Johns, Marco, et al. "A Comprehensive Portal for Clinical and Translational Data Warehouses." *Public Health and Informatics*. IOS Press, 2021. 462-466.
[10] Bonacina, Stefano, et al. "Can the European EHR Exchange Format Support Shared Decision Making and Citizen-Driven Health Science?." *Public Health and Informatics*. IOS Press, 2021. 1056-1060.
[11] Mons, Barend, et al. "Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud." *Information Services & Use* 37.1 (2017): 49-56.
[12] Ammar, Ammar, et al. "A semi-automated workflow for FAIR maturity indicators in the life sciences." *Nanomaterials* 10.10 (2020): 2068.
[13] M. D. Wilkinson, S.-A. Sansone, E. Schultes, P. Doorn, L. O. B. da Silva San- tos, and M. Dumontier, "A design framework and exemplar metrics for fairness," Scientific data, vol. 5, no. 1, pp. 1–4, 2018.
[14] Wilkinson, Mark D., et al. "Evaluating FAIR maturity through a scalable, automated, community-governed framework." *Scientific data* 6.1 (2019): 1-12.
[15] FAIR Enough - FAIRness evaluation tool https://fair-enough.semanticscience.org/
[16] Devaraju, Anusuriya, et al. "From conceptualization to implementation: FAIR assessment of research data objects." *Data Science Journal* 20.1 (2021): 1-14.
[17] "Fair-checker." https://fair-checker.france-bioinformatique.fr/ 2021. Accessed: 2022-02-18.
[18] R. de Miranda Azevedo and M. Dumontier,"Considerations for the conduction and interpretation of fairness evaluations," Data Intelligence, vol. 2, no. 1-2, pp. 285– 292, 2020.
[19] "FAIR test - a library to build and deploy FAIR metrics tests APIs." https://github.com/MaastrichtU-IDS/fair-test 2022. Accessed: 2022-02-18.