

On the Design of a FAIR Data Maturity Model for OneAquaHealth

Adriano TRAMONTANO ^{a,1}, Arriel BENIS ^{b,1}, Catherine CHRONAKI ^c,
Gora DATTA ^c, Anne MOEN ^d, Silje Havrevold HENNI ^d, Maria JOÃO FEIO ^c
and Oscar TAMBURIS ^{a,2}

^aNational Research Council of Italy, Naples, Italy

^bHolon Institute of Technology, Holon, Israel

^cHL7 Europe, Bruxelles, Belgium

^dUniversity of Oslo, Oslo, Norway

^eUniversity of Coimbra, Coimbra, Portugal

ORCID ID: Adriano Tramontano 0009-0003-8460-3844; Arriel Benis 0000-0002-9125-8300; Catherine Chronaki 0000-0001-6638-8448; Gora Datta 0000-0002-7783-296X; Anne Moen 0000-0002-3825-9355; Silje Havrevold Henni 0000-0002-9309-8726; Maria João Feio 0000-0003-0362-6802; Oscar Tamburis 0000-0002-0130-7915

Abstract. The scope of this work is to describe the overall process of assessing the compliance of the main Digital Objects produced in the OneAquaHealth project with the FAIR principles (Findability, Accessibility, Interoperability, and Reusability) via a custom FAIR Data Maturity Model. The model was designed and developed according to the project features, and in according with the One Digital Health Framework. Its goal is also to provide a tool characterized by a solid educational ground, so as to set the foundation of a timely FAIRification process for the next project steps.

Keywords. One Health, One Digital Health, OneAquaHealth, FAIRness

1. Introduction

In the OneAquaHealth project (OAH, here onwards), One Digital Health (ODH) embodies the convergence of diverse perspectives, emphasizing the need for integrative, systemic, and syndemic decision-making processes within a multidisciplinary framework. Maximizing its potential in the OneAquaHealth project is as well critical to drive innovative, sustainable, and impactful solutions for urban aquatic ecosystem management and public health [1,2]. In this regard, the FAIR framework serves as both a strategic guide and a practical toolset for ensuring that the project's digital outputs—spanning from the Environmental Monitoring Data and the Human Health and Wellbeing Indicators, up to citizen science App and the AI-based DSS—can be effectively shared, reused, and integrated within and beyond the scope of the project [3]. To operationalize these principles, OAH has adopted a dedicated FAIRification strategy, which includes (i) the definition of the Digital Objects (DOs) produced in the project, and (ii) the assessment of the compliance of such DOs with the FAIR principles (Findability,

¹ Share First Co-authorship

² Corresponding Author: Oscar Tamburis, IBB CNR, Via Tommaso de Amicis 95, 80145 Naples, Italy.
E-mail: oscar.tamburis@cnr.it

Accessibility, Interoperability, and Reusability). To this end, a custom FAIR Data Maturity Model (FDMM) was designed and developed within the project. Its goal is to offer a structured and repeatable method for assessing the FAIRness of the project's identified DOs as well as to support ongoing improvement of their FAIR attributes. While the FAIR principles provide high-level guidance, the FDMM translates these into concrete, measurable criteria for OneAquaHealth datasets. This model serves both as an evaluation tool and as a roadmap for the next FAIRification activities across the data lifecycle, from planning and documentation to storage, dissemination, and reuse.

2. Methods

A Digital Object is a formally defined, discrete digital entity that can be stored, accessed, and exchanged via computational systems. Such definition encompasses a wide spectrum of digital artifacts: datasets, documents, code, images, workflows, protocols, annotations, simulation outputs, dashboards, and more. Those can be considered as living nodes in a data ecosystem, capable of supporting scientific inquiry, decision-making, and long-term knowledge retention [4]. An FDMM consists instead of a set of assessment indicators, categorized by FAIR principle [5,6]. It is also applied in three phases:

- *Preparation*: Collection of metadata and documentation for each DO;
- *Assessment*: Evaluation of FAIRness using the model's rubric, involving both the data stewards and FAIR experts;
- *Improvement Planning*: Identification of gaps and recommendations for enhancing FAIR compliance.

Although standing on these aspects, the OAH FDMM was also designed by specifically drawing upon: (i) the broader One Digital Health perspective, emphasizing multi-domain interoperability and sustainability and (ii) the pursue an educational purpose as well, since such integration represents moreover a crucial step towards equipping future researchers (not only for computer science-related disciplines) with essential skills for navigating a data-driven approach [7,8].

2.1. Data Sources

The OAH project has so far produced a range of important DOs, in the form of Excel files hosted on the project's Microsoft Teams (MS-T) platform, which are foundational for the whole project.

One DO derives from in-situ sources, capturing dynamic aspects of urban aquatic ecosystems, and from the following analysis of the literature to describing and enhancing the results accordingly. This DO serves not only ecological research and trend analysis but also underpin an early warning system to detect signs of ecosystem stress, degradation, or recovery. Two other DOs comprise health-related indicators gathered alongside their data, which reflect the wellbeing of urban populations in relation to environmental quality. The indicators and their data are collected or derived from Public Health Databases and Literature Review.

2.2. OAH FDMM structure and rationale

The proposed model enables systematic assessment across all four FAIR categories using a well-defined scoring scheme and indicator hierarchy, articulated as follows:

- Categories and Principles (#): The model follows the canonical structure of the FAIR principles, with each of the four categories (F, A, I, R) broken down into sub-principles (e.g., F1, A1.1, R1.2).
 - According to the literature, each sub-principle is assigned a level of priority, to which a score is accordingly related – 1/Useful; 2/Important; 3/Essential [9];
- Assessment Sentences (S): Each sub-principle is described through a concise qualitative sentence
 - The sentence expresses the intent of the criterion, to make understanding by both technical and non-technical stakeholders easier;
- Tags (Q): A set of conceptual tags (e.g., <what>, <which>, <how sufficient>) help group and differentiate the nature of the attributes being assessed, such as the type of identifier used, the type of access protocol, or the sufficiency of metadata.
 - The combination S+Q is renamed Indicator (I).
- Assessment Levels (A): Each sentence is accompanied by three levels of achievement, expressing the maturity of implementation:
 - Level 1: Basic or minimal compliance.
 - Level 2: Intermediate quality or adoption of community-aligned practices.
 - Level 3: Optimal alignment with FAIR-compliant, machine-actionable, or community-standard protocols.
- Evaluation method, i.e. the way that the results of the evaluation of the indicators can be given a value.

This multi-tiered approach ensures the model remains both comprehensive and flexible, allowing assessors to capture progress and gaps in FAIRness in a nuanced way. A FAIR Score for digital resources was eventually defined, as expressed in the formula:

$$FAIR\ Score = \frac{\sum w_{P_i} * S_{P_i}}{\sum w_{P_i}} \in [1; 3]. \quad (1)$$

Once measured the FAIRness of a number of DOs quantitatively against a set of defined metrics, the FAIRification process includes identifying weaknesses (e.g., metadata missing, unclear license), taking specific actions to improve those weak areas, and to re-evaluating after changes, in a loop.

3. Results

The datasets assessed play a crucial role for either the documentation of the biological conditions of aquatic ecosystems, and for getting valuable information as to human health, across three out of five of the project's pilot sites – namely Benevento (Italy), Oslo (Norway), and Coimbra (Portugal). The following main results are summarized according to the FAIR principles. References to specific sub-principles are reported in (...). Results are also summarized in Fig.1.

Findability: all three assessed datasets use internal row-based identifiers that are not globally unique or persistent, limiting external discoverability (F1). Metadata are generally present but lack richness and machine-actionability (F3). Although MS–T offers internal organization and basic search, none of the datasets are indexed in external repositories or registries, reducing visibility and findability outside the project environment (F4). *Average Score*: 1,7.

Accessibility: it is the most mature principle across the datasets. All files are retrievable via the open HTTP protocol and benefit from MS–T’s version control, enabling access to current and previous versions with associated metadata (e.g., author, date) (A1, A1.1). Authentication and authorization are in place, but manual download is required, and access remains limited to project members, preventing open accessibility to the broader community (A1.2, A2). *Average Score*: 2,0.

Interoperability: it is the weakest area for all datasets. The use of proprietary Excel formats, free-text metadata, and the absence of machine-readable formats or formal ontologies prevents automated data integration (I1). While some datasets reference scientific terms or include informal cross-references, none employ standardized vocabularies or linked data principles, significantly hindering semantic alignment and reuse in external systems (I2, I3). *Average Score*: 1,2.

Reusability: this is partially supported through contextual metadata and informal provenance tracking, primarily via MS–T’s version history (R1). However, the absence of formal usage licenses, limited methodological documentation, and lack of alignment with domain-specific standards restrict the datasets’ potential for responsible reuse beyond the project (R1.1, R1.2). Improving licensing clarity, provenance structure, and compliance with established community standards will be essential for future FAIR maturity (R1.3). *Average Score*: 1,5.

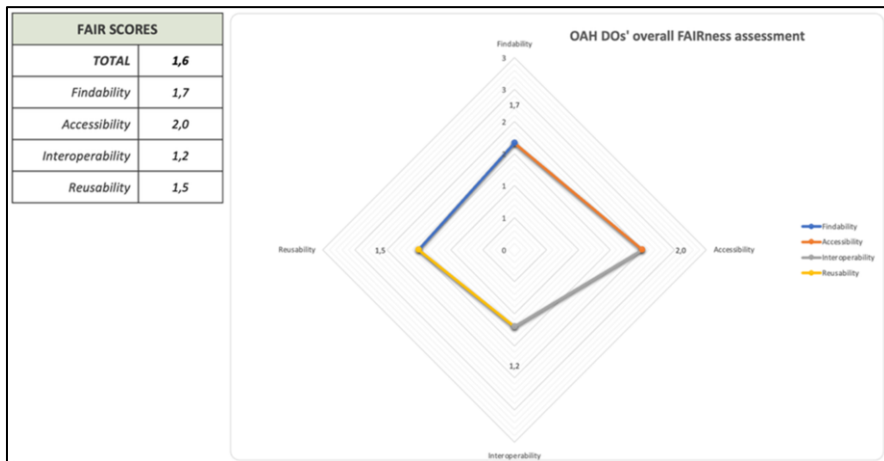


Figure 1: Overall FAIR scores for the OAH assessed Digital Objects

4. Discussion and Conclusions

Literature indicates rising adoption of FAIR principles in health and environmental sciences, with domain-specific adjustments taking place [10,11]. However, specific development of the FAIR Data Maturity Model is still very limited in practice in a broader way both in health and ecological science. In such a scenario, the OAH FDMM has been conceived with a distinct set of features and constraints in mind, tailored to the interdisciplinary and applied nature of the project, which include:

- balanced Depth and Usability: While thorough in its coverage of FAIR principles, the model avoids over-complexity by using clear descriptors and

progression levels – corresponding to understandable, realistic steps toward FAIRness – that can be scored semi-quantitatively;

- emphasis on Practical Implementation: Instead of focusing exclusively on theoretical compliance, the model considers the real-world technical and policy constraints faced by the project’s environmental and health datasets:
- inclusivity of stakeholders: its structure supports communication across diverse teams—from data managers and technical implementers to researchers and domain experts:
- community Standards and Interoperability Awareness: Several indicators (e.g., I1, R1.3) emphasize adherence to open formats, community standards, and machine-actionable structures, also critical for enhancing the efficiency of a health-related Natural Language Processing application. This reflects the project’s ambition to enable long-term data integration and reuse across domains.

The main contributions of the OAH FDMM do not really lie in outperforming existing models in technical precision, but rather in complementing them with a context-sensitive, human-centered, operationally feasible and educationally grounded tool for a twofold aim: (i) evaluating and improving FAIRness within a real-world, cross-domain research environment; and (ii) pursuing a gradual FAIRification within One Digital Health ecosystems.

Acknowledgments

This work was partially funded by the EU Horizon Europe Research and Innovation Program (HORIZON-CL6-2022-GOVERNANCE-01) grant agreement No. 101086521 - OneAquaHealth (Protecting urban aquatic ecosystems to promote One Health).

References

- [1] Home - OneAquaHealth, (2022). <https://www.oneaquahealth.eu/> (accessed April 22, 2025).
- [2] A. Benis, O. Tamburis, C. Chronaki, and A. Moen, One Digital Health: A Unified Framework for Future Health Ecosystems, *J Med Internet Res.* **23** (2021) e22189. doi:10.2196/22189.
- [3] M.D. Wilkinson, M. Dumontier, et al., The FAIR Guiding Principles for scientific data management and stewardship, *Sci Data.* **3** (2016) 160018. doi:10.1038/sdata.2016.18.
- [4] P. Wittenburg, From Persistent Identifiers to Digital Objects to Make Data Science More Efficient, *Data Intelligence.* **1** (2019) 6–21. doi:10.1162/dint_a_00004.
- [5] Research Data Alliance FAIR Data Maturity Model Working Group, FAIR Data Maturity Model: specification and guidelines, (2020). doi:10.15497/RDA00050.
- [6] A. Devaraju, and R. Huber, An automated solution for measuring the progress toward FAIR research data, *Patterns.* **2** (2021) 100370. doi:10.1016/j.patter.2021.100370.
- [7] O. Tamburis, and A. Benis, One Digital Health for more FAIRness, *Methods Inf Med.* **61** (2022) e116–e124. doi:10.1055/a-1938-0533.
- [8] R. Gonzalez Soltero, D. Pino García, et al., FAIR data management: a framework for fostering data literacy in biomedical sciences education, *BMC Med Res Methodol.* **24** (2024) 284. doi:10.1186/s12874-024-02404-1.
- [9] G. Peng, G. Berg-Cross, et al., Harmonizing quality measures of FAIRness assessment towards machine-actionable quality information, *International Journal of Digital Earth.* **17** (2024) 2390431. doi:10.1080/17538947.2024.2390431.
- [10] M. Cannon, A. Kelly, et al., Implementing an Open & FAIR DATA SHARING policy—A case study in the earth and environmental sciences, *Learned Publishing.* **35** (2022) 56–66. doi:10.1002/leap.1442.
- [11] R. Nault, M.C. Cave, et al., A Case for Accelerating Standards to Achieve the FAIR Principles of Environmental Health Research Experimental Data, *Environ Health Perspect.* **131** (2023) 065001. doi:10.1289/EHP11484.