FAIR Digital Twins of protein variants for Pandemic Preparedness

Max Van den Boom, Lilli Schuckert, Erik Schultes
Metabolomics and Analytics Center, Leiden Academic Center for Drug Research
Contact: e.a.schultes@lacdr.leidenuniv.nl

Universiteit Leiden
The Netherlands

More Info

1. Introduction

Presently, pandemic-related data are diverse, distributed, and heterogeneous in their representation, prohibiting data discovery, reuse and timely analysis for global trends that are critical to an informed pandemic response. Here, we present results from the Health-Holland funded STAYAHEAD project (2023-2025) [1] aiming to develop methods that make pandemic-related data FAIR (*Findable, Accessible, Interoperable, Reusable*) [2], focusing on the human immunology of the SARS-CoV-2 Spike protein.

2. Making pandemic-related data FAIR

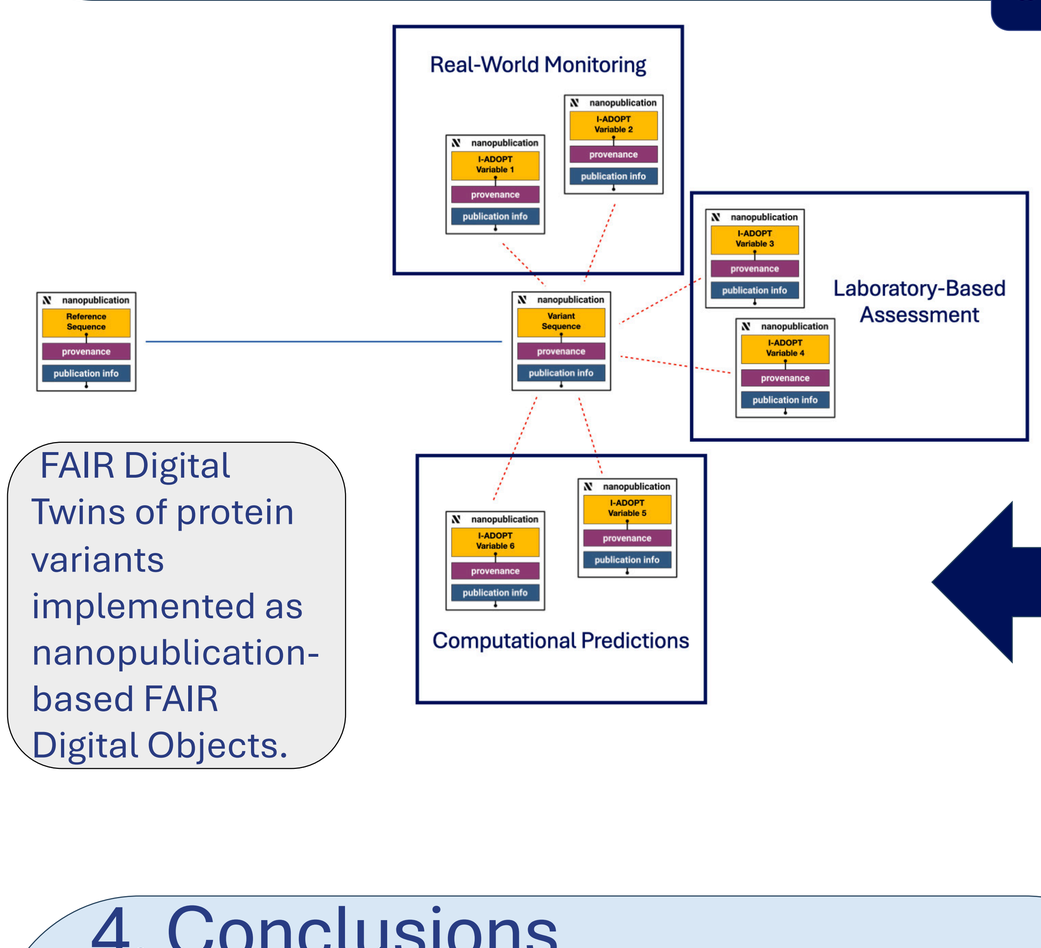
In order to make pandemic-related data FAIR, we leveraged recent findings that when nanopublication assertions refer to digital object types (and type specific metadata) they adhere to the emerging specifications of FAIR Digital Objects (FDOs) [3]. Hence, appropriately constructed nanopublications and the existing nanopublication server network can already provision a general and universal FDO ecosystem. In this study, our FDO approach involves:

1) the conceptual modeling of nanopublication templates representing Spike protein variants and observations made on these variants;
2) the publication (or re-publication) of data using these nanopublication templates onto the open and decentralized nanopublication server network.

3. Methods

We employ pandemic-related data from three disparate knowledge domains:

- **Real-World Monitoring:** Spike protein variant detection and metabolic profiling of disease progression using high-throughput mass spectrometry.
- **Laboratory-Based Assessment:** Evaluation of immunological escape of Spike protein variants with in vitro assays.
- **Computational Predictions:** Machine-Learning on predicted structural features of large numbers of Spike protein variants (both documented and theoretical).



FAIR Digital Twins of protein variants implemented as nanopublication-based FAIR Digital Objects.

We employ a novel combination of three innovative technology frameworks to render these pandemic-related data FAIR:

- **The Nanopublication Framework:** Provides an open standard for publication of unitary, stand-alone semantic assertions, as well as a mature technology stack supporting publication support, storage, access, search functionality, and provenance metadata (e.g., time/date stamps, authentication keys, globally unique identifiers) [4]. We use Nanopublications to (1) identify large numbers of individual Spike protein variants, and (2) to describe a variety of data about them.
- **The I-ADOPT Framework:** Provides an emerging, open, and minimal standard for describing scientific observations and measurements. We use I-ADOPT to describe observations made on individual Spike protein variants [5].
- **FAIR Digital Object Framework:** Provides an emerging, open and minimal standard for semantically describing any digital object. Recent results confirm that appropriately constructed Health-Holland nanopublication assertions are conformant to the FAIR Digital Object Framework [6]. We use nanopublication-based FDOs to demonstrate a FAIR data space for managing pandemic-related data.

Taken together, this method demonstrates an open, light-weight and reusable platform supporting the decentralized yet trust-worthy provision of real-time, province-rich FAIR Digital Twins of Spike protein variants [7].

4. Conclusions

- We have implemented FAIR Digital Twins for protein variants using a novel combination of existing, low-cost, easy-to-use and thus rapidly deployable technologies.
- However, future directions must include the development of access control mechanisms for privacy sensitive (e.g., patient) data.
- FAIR Digital Twins can significantly impact pandemic preparedness and outbreak management by enabling real-time data sharing and AI-driven insights.
- Our FAIR Digital Twin can be extended with data types from additional knowledge domains, enhancing its utility for real-time monitoring and AI-assisted prediction of viral-immune system evolution.
- The ultimate scalability and sustainability of this approach depend on a balance of public investments (e.g., infrastructure) and market-driven solutions (e.g., services).

Acknowledgments

This work is part of the Health-Holland funded STAYAHEAD project (LSHM22038), project PI Dr. Thomas Hankemeier. This work was also supported by the GO FAIR Foundation, Leiden, The Netherlands. We thank Tobias Kuhn and Barbara Magagna for expert guidance.

References

[1] Staying ahead of the virus https://www.health-holland.com/project/2023/2022/staying-ahead-virus
[2] Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.18
[3] Magagna, B., Bonino, L., Kuhn, T., Pires, L.F., Schultes, E., Can FAIR Digital Objects be implemented as nanopublications? Preprint May 2024, https://osf.io/p3khc
[4] Nanopublications: FAIR data containers for scientific results, and more https://nanopub.net/
[5] Magagna, B., Moncoiffé, G., Devaraju, A., Stoica, M., Schindler, S., Pamment, A., & RDA I-ADOPT WG. (2022). InteroperAble Descriptions of Observable Property Terminologies (I-ADOPT) WG Outputs and Recommendations (1.1.0). Zenodo. https://doi.org/10.15497/RDA00071
[6] L. O. Bonino da Silva Santos, T. P. Sales, C. M. Fonseca, and G. Guizzardi, "Towards a conceptual model for the FAIR Digital Object framework," in Formal Ontology in Information Systems, IOS Press, 2023, pp. 227–24
[7] Schultes E, Roos M, Bonino da Silva Santos LO, Guizzardi G, Bouwman J, Hankemeier T, Baak A and Mons B (2022) FAIR Digital Twins for Data-Intensive Research. Front. Big Data 5:883341. https://doi.org/10.3389/fdata.2022.883341

eosc