The Comparative Anatomy of Nanopublications and FAIR Digital Objects

Erik Anthony Schultes[‡,§], Barbara Magagna[|], Tobias Kuhn[¶], Marek Suchánek[#], Luiz Olavo Bonino da Silva Santos[¤], Barend Mons[«]

‡ Leiden University, Leiden, Netherlands
§ Leiden Center for Data Science, Leiden, Netherlands
| GO FAIR Foundation, Leiden, Netherlands
¶ Vrije Universiteit Amsterdam, Department of Computer Science, Amsterdam, Netherlands
# Czech Technical University in Prague, Faculty of Information Technology, Prague, Czech Republic
¤ University of Twente, Enschede, Netherlands
« Leiden University Medical Center, Leiden, Netherlands

Corresponding author: Barbara Magagna (barbara@gofair.foundation)

Abstract

Beginning in 1995, early Internet pioneers proposed Digital Objects as encapsulations of data and metadata made accessible through persistent identifier resolution services (Kahn and Wilensky 2006). In recent years, this Digital Object Architecture has been extended to include the FAIR Guiding Principles (Wilkinson et al. 2016), resulting in the concept of a FAIR Digital Object (FDO), a minimal, uniform container making any digital resource machine-actionable. Intense effort is currently underway by a global community of experts to clarify definitions around an FDO Framework (FDOF) and to provide technical specifications (FAIR DO group 2020, FAIR Digital Object Forum 2020 , Bonino da Silva Santos (2021)) regarding their potential implementation.

Beginning in 2009, nanopublications were independently conceived (Groth et al. 2010) as a minimal, uniform container making individual semantic assertions and their associated provenance metadata, machine-actionable. They represent minimal units of structured data as citable entities (Mons and Velterop 2009). A nanopublication consists of an *assertion*, the *provenance* of the assertion, and the provenance of the nanopublication (*publication info*). Nanopublications are implemented in and aligned with Semantic Web technologies such as RDF, OWL, and SPARQL (World Wide Web Consortium (W3C) 2015) and can be permanently and uniquely identified using resolvable Trusty URIs (Groth et al. 2021). The existing Nanopublication Server Network provides vital services orchestrating nanopublications (Kuhn et al. 2021) including identifier resolution, storage, search and access. Nanopublications can be used to expose quantitative and qualitative data, as well as hypotheses, claims, negative results, and opinions that are typically unavailable as structured data or go unpublished altogether. The first practical application of nanopublications occurred in 2014, with the publication of millions of nanopublications as part of the FANTOM5 Project (The FANTOM Consortium and the RIKEN PMI and CLST

©
.

(DGT) 2014, Lizio et al. 2015). Since then, millions of real-world examples spanning diverse knowledge domains are now available on the nanopublication server network.

Like nanopublication, the FDOF also posits an ultra-minimal approach to structured, self-contained, machine-readable data and metadata. An FDO consists of: the *object* itself (subsequently referred to here as the *resource* to avoid confusion with other meanings of the term "object"); the *metadata* describing the resource; and a globally unique and persistent *identifier* with predictable resolution behaviors.

These two technologies share the same vision of a data infrastructure, and act as instances of Machine-Actionable Containers (MACs) that make use of minimal uniform standards to enable FAIR operations. Here, we compare the structure and computational behaviors of the existing nanopublication infrastructure, to those in the proposed FAIR Digital Object Framework. Although developed independently there are clear parallels between the vision and the approach of nanopublication and FDOF. Each aspires to minimal standards for the encapsulation of digital information into free-standing, publishable (citable, referenceable) entities. The minimal standards involve globally unique and persistent identifiers that resolve to standardized semantically enabled metadata descriptions that include machine actionable paths to the resource itself.

At the same time, there are also differences. The scope of nanopublications is limited to the assertional data type and, as the name suggests, nanopublications should remain small in size (limited to single assertions as individual triples or small RDF graphs). In contrast FDOs are unlimited in their scope, accommodating digital resources of arbitrarily large size, type and complexity, so long as their type can be ontologically described. Furthermore, whereas nanopublications represent a moderately mature technology, the FDOF is a specification still under development. If it were possible to formally draw points of contact between the two approaches, then it would be possible to leverage the vast practical experience gained in the nanopublishing of assertions for the FDO community.

Here, inspired by recent applications of nanopublications in the FIP Wizard tool (Schultes et al. 2020), and their extension to research claims (Kuhn 2022, McNamara 2022) and data using Schultes (2022a), Schultes (2022b), we attempt a point-by-point comparison of the specifications between nanopublication and FDOs. We find a remarkable congruence between the currently proposed FDO requirements and the existing nanopublication infrastructure, including several FDO-like qualities already embodied in the nanopublication ecosystem.

Keywords

FAIR Principles, Nanopublication, Nanopublication Ecosystem, Machine-Actionable Containers, FIP Wizard, FAIR Wizard of Leiden

Presenting author

Erik Anthony Schultes

Presented at

First International Conference on FAIR Digital Objects, presentation

Conflicts of interest

References

- Bonino da Silva Santos LO (2021) FAIR Digital Object Framework Documentation. https://fairdigitalobjectframework.org/. Accessed on: 2022-7-10.
- FAIR Digital Object Forum (Ed.) (2020) FAIR Digital Objects Forum . https://fairdo.org. Accessed on: 2022-8-27.
- FAIR DO group (2020) Joint Statement on FAIR Digital Object Framework. https://docs.google.com/document/d/11FmDxgncy-LynQqTlvxFProW-i5II7JBFtp7ELyztlg/edit. Accessed on: 2022-7-10.
- Groth P, Gibson A, Velterop J (2010) The anatomy of a nanopublication. Information Services & Use 30: 51-56. https://doi.org/10.3233/isu-2010-0613
- Groth P, Schultes E, Thompson M, Tatum Z, Dumontier M, Kuhn T, Chichester C (2021) Nanopublication Guidelines. URL: https://nanopub.org/guidelines/working_draft/
- Kahn R, Wilensky R (2006) A framework for distributed digital object services. International Journal on Digital Libraries 6 (2): 115-123. https://doi.org/10.1007/s00799-005-0128-x
- Kuhn T, Taelman R, Emonet V, Antonatos H, Soiland-Reyes S, Dumontier M (2021) Semantic micro-contributions with decentralized nanopublication services. PeerJ Computer Science 7 https://doi.org/10.7717/peerj-cs.387
- Kuhn T (2022) The Future of Science Publishing - Personal Views. IOS Press. IOS Press 35 Anniversary meeting. URL: https://www.iospress.com/ios-press-35
- Lizio M, , Harshbarger J, Shimoji H, Severin J, Kasukawa T, Sahin S, Abugessaisa I, Fukuda S, Hori F, Ishikawa-Kato S, Mungall CJ, Arner E, Baillie JK, Bertin N, Bono H, de Hoon M, Diehl AD, Dimont E, Freeman TC, Fujieda K, Hide W, Kaliyaperumal R, Katayama T, Lassmann T, Meehan TF, Nishikata K, Ono H, Rehli M, Sandelin A, Schultes EA, 't Hoen PA, Tatum Z, Thompson M, Toyoda T, Wright DW, Daub CO, Itoh M, Carninci P, Hayashizaki Y, Forrest AR, Kawaji H (2015) Gateways to the FANTOM5 promoter level mammalian expression atlas. Genome Biology 16 (1). https://doi.org/10.1186/s13059-014-0560-6
- McNamara C (2022) The Future of Science Publishing – Focus on Nanopublications and Formalization Papers. IOS Press. URL: https://labs.iospress.com/news-blog/future-science-publishing-focus-on-nanopublications

- Mons B, Velterop J (2009) Nano-publication in the e-science era. *Workshop on Semantic Web Applications in Scientific Discourse (SWASD 2009)*, Washington, DC.
- Schultes E, Magagna B, Hettne KM, Pergl R, Suchánek M, Kuhn T (2020) Reusable FAIR Implementation Profiles as Accelerators of FAIR Convergence. Lecture Notes in Computer Science138-147. https://doi.org/10.1007/978-3-030-65847-2_13
- Schultes E, et al. (2022a) FAIR Wizard of Leiden. https://bit.ly/FWLnano. Accessed on: 2022-7-10.
- Schultes E (2022b) The FAIR Wizard of Leiden. IOS Press. IOS Press 35 Anniversary meeting. URL: https://www.youtube.com/watch?v=y79t6f3i4Ow
- The FANTOM Consortium and the RIKEN PMI and CLST (DGT) (2014) A promoter-level mammalian expression atlas. Nature 507 (7493): 462-470. https://doi.org/10.1038/nature13182
- Wilkinson M, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten J, da Silva Santos LB, Bourne P, Bouwman J, Brookes A, Clark T, Crosas M, Dillo I, Dumon O, Edmunds S, Evelo C, Finkers R, Gonzalez-Beltran A, Gray AG, Groth P, Goble C, Grethe J, Heringa J, 't Hoen PC, Hooft R, Kuhn T, Kok R, Kok J, Lusher S, Martone M, Mons A, Packer A, Persson B, Rocca-Serra P, Roos M, van Schaik R, Sansone S, Schultes E, Sengstag T, Slater T, Strawn G, Swertz M, Thompson M, van der Lei J, van Mulligen E, Velterop J, Waagmeester A, Wittenburg P, Wolstencroft K, Zhao J, Mons B (2016) The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data 3 (1). https://doi.org/10.1038/sdata.2016.18
- World Wide Web Consortium (W3C) (2015) Semantic Web . https://www.w3.org/standards/semanticweb/. Accessed on: 2022-8-27.