International FAIR Digital Objects Implementation Summit 2024
Full Papers
https://doi.org/10.52825/ocp.v5i.1417

Published: 18 Mar. 2025

Nanopublications as FAIR Digital Object Implementations

Barbara Magagna[1,2,*], Luiz Olavo Bonino da Silva Santos[1,3], Tobias Kuhn[4], Luís Ferreira Pires[1], and Erik Schultes[2,5]

[1]Semantics, Cybersecurity and Services, University of Twente, Enschede, the Netherlands
[2]GO FAIR Foundation, Leiden, the Netherlands
[3]BioSemantics, Leiden University Medical Center, Leiden, the Netherlands
[4]Knowledge Pixels, Zurich, Switzerland / Vrije Universiteit Amsterdam, Amsterdam, the Netherlands
[5]Leiden Academic Center for Drug Research, Leiden University, Leiden, the Netherlands
*Correspondence: Barbara Magagna, barbara@gofair.foundation

Abstract. Numerous implementations of FAIR Digital Objects (FDOs) are actively emerging and are being evaluated against FDO specifications. Here, we focus on the Nanopublication Framework as a possible implementation of FDOs. Nanopublications are unitary, standardised, and self-contained RDF-based knowledge graphs consisting of three subgraphs: an assertion graph that includes the main content, a provenance metadata graph, and a publication information graph. In a thorough analysis of the FDO specifications (represented by the efforts of the FDO Forum) and based on Semantic Web (represented by the FDO Framework) we demonstrate that the existing Nanopublication Framework aligns with the overall FDO definitions and requirements. We conclude that appropriately designed nanopublications can indeed be deployed as a general purpose FDO implementation. This nanopublication-based FDO approach is initially applied through the extension of a commonly used nanopublication template, now aligned with the FDO Framework. The original template is currently used in real-world applications where communities of practice describe the FAIR Supporting Resources that constitute FAIR Implementation Profiles.

Keywords: FAIR Principles, FAIR Digital Object, Nanopublication Network

1. Introduction

The explosion of research data created over the past three decades presents challenges to data reuse. The FAIR Guiding Principles, originally conceived at the Lorentz conference in 2014 [1] and published in 2016 [2] address this problem by defining the characteristics digital objects should have to make them findable, accessible, interoperable and reusable by computational agents. Key findings from the second European Open Science Cloud (EOSC)'s High-Level Expert Group report *Turning FAIR into Reality* [3] were that FAIR practices require a major shift of research culture and practice, and the presence of a FAIR data ecosystem, where FAIR Digital Objects (FDOs) would play a central role.

The FDO concept took shape during the Moving Forward on Data Infrastructure Technology Convergence workshop in Paris in 2019 (see https://tinyurl.com/FDO-Paris), involving a global community of experts covering different backgrounds. The workshop aimed to combine the FAIR Principles with the extensive work of several RDA working groups, such as Data Foundation and Terminology WG which resulted in the development of a Core Data Model known as the Digital Object Model [4].This builds on the Digital Objects (DO) introduced by Robert Kahn [5] in the early 1990s and addresses the need to support information management and enable interoperability between information systems. The Paris event inspired the formation of the FAIR Digital Object Forum (FDO Forum, see https://fairdo.org/) which in recent years has developed specifications for the implementation of FDOs [6]. In parallel, Bonino developed the FDO Framework [7] which is also implemented as an ontology (see https://raw.githubusercontent.com/utwente-scs/fdof-o/main/owl/fdof-o.ttl).

While the promise of FDOs is evident, a consensus on how to implement FDOs remains challenging. This has resulted in a variety of implementations [8], each focusing on specific aspects of the problem without providing a complete implementation solution. One of these approaches is based on the Digital Object Architecture (DOA) that leverages Handles as persistent identifiers and Digital Objects as a core data model. FDO prototypes include components like the Digital Object Interface Protocol, the Data Type Registry, and the Typed Kernel Attributes with specific operations in projects like DiSCCo and CLARIN. Furthermore, Research Object Crates (see https://www.researchobject.org/ro-crate/) and FAIR Signposting (see https://signposting.org/FAIR/) have emerged as promising technologies supporting the FDO concept. Despite the various implementations demonstrating specific components of FDOs, no single FDO infrastructure is yet fully operational.

Recently, a systematic comparative investigation of the FDOs applied in typical approaches of Linked Data implementation was performed [9], analyzing different dimensions including data, metadata, service, access, operations and computation. An important conclusion of this study is that the diverse and open deployment of Linked Data makes the alignment to the FDO requirements challenging. Linked Data have typically been underconstrained leaving too many degrees of freedom that, ironically, impede interoperability and make it difficult for machines to consistently orchestrate multiple, machine-actionable resources across repositories and data providers. The study proposes that an "FDO profile for Linked Data" should be established.

One particular technology that emerged on top of Linked Data called nanopublication is noteworthy however, in that it provides a consistent container format to attach provenance and metadata to an atomic RDF statement realized as named graphs. A brief comparative analysis [10] of the Nanopublication Framework with early versions of the FDO Framework (FDOF) [7] led to the identification of a close match.

This paper examines in more technical detail the extent to which nanopublications comply with FDOs based on the set of FDO recommendations defined by the FDO Forum, and the FDO Framework.

The paper is structured as follows: Section 2 takes a closer look into nanopublications and how they are applied for describing resources used in FAIR Implementation Profiles; Section 3 analyses the compliance of Nanopublication Framework with the FDO Forum FDO Requirements specifications; Section 4 discusses the conformity of the Nanopublication Framework with the FDO Framework; Section 5 presents our conclusions.

2. The form and function of nanopublications

As early as 2009, nanopublications had been conceptualized as minimal, uniform containers for machine-interpretable information, employing Linked Data principles and semantic specifications [11]. Originally they were designed specifically for scientific assertions, providing researchers with the format to express statements in their research domain, as stand-alone (scholarly) publications. The first practical application of nanopublications was demonstrated in 2015 in the publication of millions of instances as part of the FANTOM5 Project [12]. Soon after, the first steps were taken to deploy a distributed server network offering essential services to orchestrate nanopublications [13], including identifier resolution, creation, storage, search and access. The nanopublication server network currently hosts millions of nanopublications representing a diverse spectrum of knowledge domains, as well as approximately 40 services that support routine, FAIR-related operations on nanopublication templates and instances.

Essentially, a nanopublication is a small RDF knowledge graph consisting of three sub-graphs: (1) the assertion graph, which constitutes the main content of the nanopublication, (2) the provenance metadata graph describing how the information was generated, and (3) the publication information graph including metadata about the creator, date and time of the nanopublication creation, the metadata template used and the usage license. Nanopublications are permanently and uniquely identified using a service that provisions resolvable Trusty URIs [14] which use cryptographic hash functions to create persistent, globally unique and content-addressable identifiers. Trusty URIs ensure the integrity and immutability of nanopublications and allow for decentralized and redundant retrieval.

While nanopublications were originally conceived to have a single RDF statement in the assertion to express a fact, observation or hypothesis [11], this paradigm has shifted towards a more general approach that allows for multiple RDF triples in order to express more complex statements. Furthermore, since their original conception, nanopublications have been broadened in their scope of use, to cover not only scientific statements, but also data points and datasets, workflows, social relations, reviews, annotations, and in general, all kinds of statements on the domain, meta, and social levels. For example, efforts around the FAIR Implementation Profile (FIP) have developed methods to list declarations of technological choices made by a community to meet the FAIR principles [15]. Here nanopublications are used to reference and describe not only the community itself but also the diverse array of FAIR Supporting Resources (FSRs) that compose the FAIR ecosystem. This approach builds on relevant classes and object properties defined in the FIP ontology that includes a FSR typology reflecting the GO FAIR Foundation's Interpretations of the FAIR Principles (see https://www.gofair.foundation/interpretation). Depending on the FSR type (for which 22 types have been defined), various minimal metadata schemas have been defined and implemented as nanopublication templates by reusing standard metadata elements.

In general, nanopublications can be used to make objects FAIR by providing formally precise statements about their nature and contexts. In addition, the Nanopublication Framework provides stable and persistent identifiers for the addressed objects but also for nanopublications serving as their metadata records. Whether the Nanopublica-

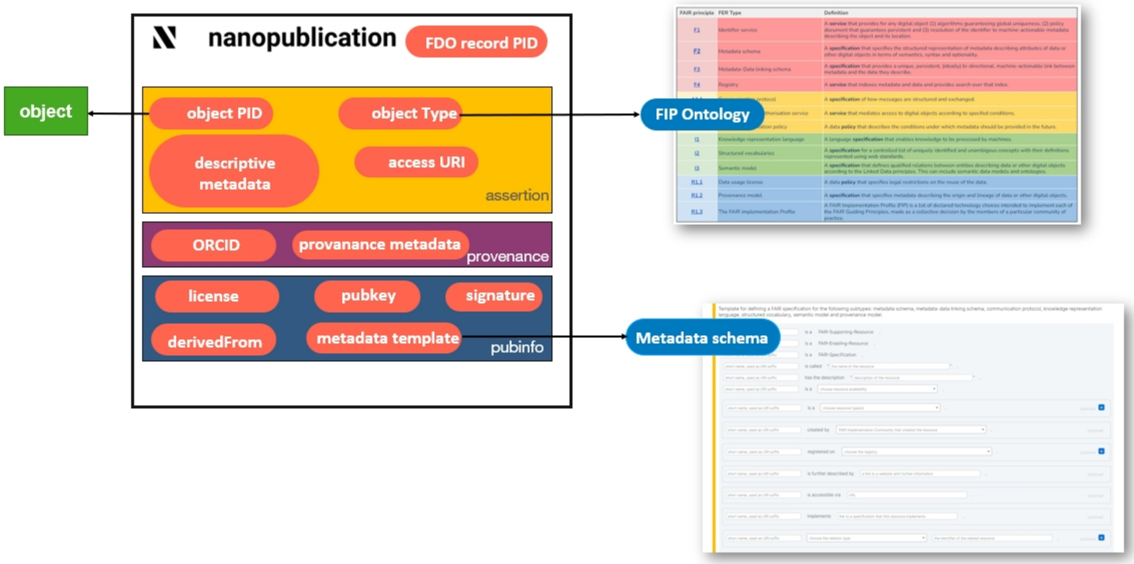

Figure 1. *Structure of an FSR nanopublication highlighting components compliant with the FDO Forum specifications.*

tion Framework might be helpful for implementing FDOs is further explored in section 3 and 4.

3. Nanopublication compliance to FDO Forum FDO Requirements Specifications

The working groups of the FDO Forum (see https://fairdo.org/wg/) have produced detailed requirements documents, which outline the technical path towards the realisation of FDOs. The *FAIR Digital Object Overview* [16] provides a full technical definition for FDOs that essentially states that a FDO is a unit composed of machine interpretable data having (1) an assigned globally unique and persistent identifier (PID), (2) a type definition for the object and (3) associated metadata about the object making it FAIR for both humans and computers. Stated more succinctly, the PID should resolve to a standard unit of information, further referred to as FDO record. Based on this definition, Figure 1 shows the mapping of the required FDO elements onto the nanopublication model used for FSR descriptions. Here, the FSR nanopublication as a whole can be taken as the FDO record; the FSR nanopublication is referenced by its Trusty URI which corresponds to the FDO record PID; the FSR nanopublication contains the persistent identifier to the object itself (object PID), the object type specification (as defined in the FIP Ontology) and the descriptive metadata based on the attributes defined in the metadata schema for the object it represents, depending on its type.

The FDO overview document [16] points to the *FDO Forum Requirements Specification* [6] and 11 other related documents that detail the specifications. We assessed the compliance of FSR nanopublications to requirements listed in [6], namely the *Generic Guidelines (GG)*, the *General Requirements (GR)*, the *PID Layer Specifications (PR)*, the *FDO Layer Specifications (FR)* and the *Resource Layer Specifications (RR)*. Here we focuses on only the most relevant requirements (see for more details https://osf.io/qndu8) grouped by topic, namely:

- *Long-term infrastructure investments (GG1, GG2)*: Long-term infrastructure investment should provide confidence in data survival over time. Nanopublications are supported by the nanopublication server network that guarantees their persistence. Long term infrastructure investments are being explored with the FAIR Connect Foundation but a definitive commitment has not yet been made.
- *Technology independence (GG1, GG8)*: The implementation of FDOs should not be tied to any particular technology stack to avoid vendor lock in. Nanopublications reuse non-proprietary RDF technology according to the W3C standards and support all applicable RDF serialization formats.
- *Abstraction principle (GG5, GR11)*: Abstracting away details should be possible, as well as creating FDOs that constitute FDO collections. Nanopublications can follow the abstraction principle because details can be separated from the main information by dividing them into separate nanopublications, and because they can be grouped in collections. The latter are represented as an nanopublication index, which is also a nanopublication itself.
- *FAIR Principle F1 (globally unique, persistent and resolvable identifier) (GG3, GG6, GR1, PR4)*: Every FDO is assigned one or more PIDs. Nanopublications are implemented in alignment with Semantic Web technologies using RDF triple statements and are permanently and uniquely identified using resolvable Trusty URIs.
- *Encapsulation and Typing system (GG7, GR7)*: FDOs should be linked to their supported operations according to their type. FAIR Supporting Resources are typed in the FIP Ontology [17]. Types can be defined, published, and structurally retrieved with properly formed nanopublications, and associated to specific operations and their definitions. We have not modeled nanopublications yet to precisely follow the FDO specifications, but the formats and the system for doing so are available and ready.
- *FDO profile and Data Type Registry (GR2, GR4, GR13, PR2, PR3, FR1, FR2)*: A PID resolves to a FDO record compliant with a machine-actionable FDO Profile specifying community-specific FDO attributes in order to lead to predictable results. The profile and the attributes should be defined in a Data Type Registry. FSR nanopublications are described based on type-specific metadata schemas that might be interpreted as minimal profiles. Their definition is based on Linked Data Principles and not on symbolic grammar, but both approaches allow for machine-actionability.
- *CRUD operations (GR6)*: It should be possible to Create, Read, Update and Delete FDOs. While nanopublications can be created, read and updated, they cannot be fully deleted once they have been published. Instead, they can be retracted while retaining their traceability, even if they are no longer valid.

Although the FDO Overview [16] acknowledges the possibility to implement FDOs as web-based implementation approach, the specifications are kept technology independent and do not suggest specific alignments for Linked Data solutions. For conformance, attributes of FDO nanopublications must be registered in a FAIR registry, and profiles with links to machine-actionable specifications should be provided. This would ensure that the Nanopublication Framework meets all of the FDO Forum's requirements.

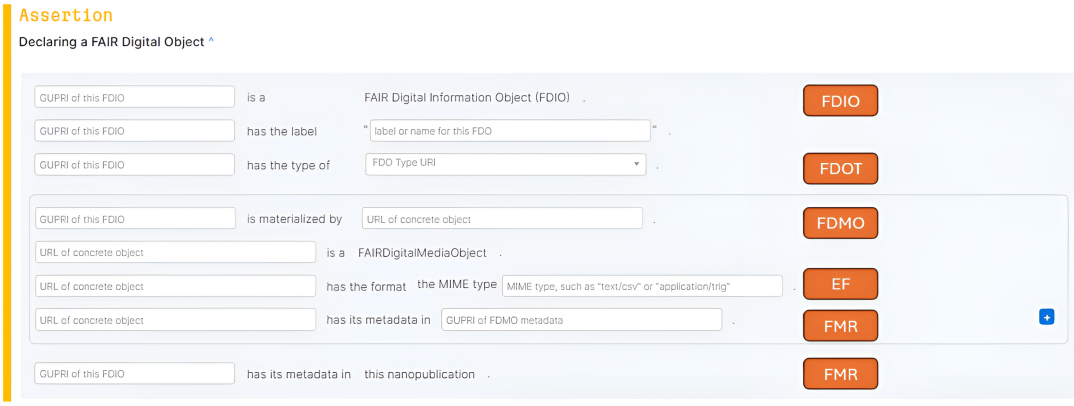

Figure 2. *Template for creating an FDO compliant nanopublication annotated with FDOF concepts.*

4. Nanopublication compliance to the FDO Framework

We evaluated the FSR nanopublications also against the requirements proposed in the FDOF [7] because this framework uses Semantic Web and just like nanopublications it explicitly requires RDF as the representation language to support semantic interoperability and machine actionability at the same time. The FDOF aims for compliance with the FAIR principles by leveraging URIs as PIDs and Linked Data as the core data model. An essential features of this framework is a predictable identifier resolution behaviour which is a method to access more information about the object (metadata) given its identifier and an object typing system. The FDOF aims to provide the minimal requirements for answering the following questions in a machine-actionable way: "What is the object that is identified by this identifier? How can I get more information (e.g., how to handle it?, who can handle it?, what is it allowed to do with it?) about this object? What can be done with the object? What am I allowed to do with this particular object?" [7]. To answer these questions we need to identify the type of object with respect to both its encoding format and its informational content.

Based on the conceptual model behind the FDOF [18] a FAIR Digital Object can be either a FAIR Digital Media Object (FDMO) or a FAIR Digital Information Object (FDIO), whereby the latter is materialized by at least one FDMO. The FDMO is connected to its characteristic encoding format. A given FDO is generally characterised by one sub-type of FDO Type (FDOT) which aggregates the set of properties that charaterise FDIOs of that type. The FAIR Metadata Record (FMR) as a specialization of FDIO describes FDOs with a number of metadata attributes based on its type and the specific instance being described. To enable a predictable identifier resolution behaviour and conformance with the FAIR principle F3, the object property `fdof:isMetadataOf` should connect the FMR with its described FDO.

Nanopublications can be structured according to this Framework by using the object properties defined in the FDOF ontology. We created a nanopublication template (Figure 2) that includes references to the FDOT, FDMO and FMR (see `https://nanodash.knowledgepixels.com/publish?285&template=https://w3id.org/np/RAJHSYGX-R2BcGQj-p1aT2Qaw44Yk7Bmco2nLBjyfZEfA`). We also drafted an implementation of the required properties in the FSR nanopublication templates to comply with the FDOF requirements by adding these RDF triples in the assertion graph (see `https://nanodash.knowledgepixels.com/publish?289&template=https://w3id.org/np/RAgtyrSb4TTffyfB24u0o2x9lnPcd2DGRzUUyjE5iuC6g`).

5. Conclusions

We have conducted a detailed technical comparison of the Nanopublication Framework with two FDO specifications: the FDO Forum Requirements and the FDO Framework. Our analysis shows that the existing Nanopublication Framework is consistent with the general definitions and requirements of both approaches. Any outstanding requirements can be accommodated with trivial modifications to the nanopublication schema. A key challenge for the future would be the design and development of a typing system for machine-actionable specifications for FDOs. We fully agree that an FDO profile for Linked Data as suggested by Soiland et al [9] would be beneficial.

Data availability statement

This submission is not using any data resources and does not generate any data.

Author contributions

B.M.: Conceptualization and writing - review and editing (lead). E.S.: Conceptualization (supporting). L.F.P: Supervision (equal). L.O.B.: Supervision (equal). T.K.: Software.

Competing interests

The authors declare that they have no competing interests.

Acknowledgement

The authors gratefully acknowledge Shuai Wang for his valuable suggestions on editing and graphical improvements. The collaboration project is co-funded by PPP Allowance awarded by Health Holland, Top Sector Life Sciences and Health, to stimulate public-private partnerships.

References

[1] S. Lusher and B. Mons, *Data FAIRPORT Lorentz workshop*, 2014. [Online]. Available: https://www.lorentzcenter.nl/jointly-designing-a-data-fairport.html.

[2] M. M. D. Wilkinson et al., "The FAIR Guiding Principles for scientific data management and stewardship," *Scientific data*, vol. 3, p. 160 018, 1 Mar. 2016, ISSN: 20524463. DOI: 10.1038/sdata.2016.18. [Online]. Available: https://doi.org/10.1038/sdata.2016.18.

[3] European Commission and Directorate-General for Research and Innovation, *Turning FAIR into reality – Final report and action plan from the European Commission expert group on FAIR data*. Publications Office, 2018. DOI: doi/10.2777/1524.

[4] G. Berg-Cross, R. Ritz, and P. Wittenburg, *Data Foundation and Terminology Work Group Products*, Feb. 2019. DOI: 10.15497/06825049-8CA4-40BD-BCAF-DE9F0EA2FADF.

[5] R. Kahn and R. Wilensky, "A framework for distributed digital object services," *International Journal on Digital Libraries*, vol. 6, pp. 115–123, 2 Apr. 2006, ISSN: 14325012. DOI: 10.1007/s00799-005-0128-x.

[6] I. Anders et al., *FDO forum FDO requirement specifications*, en, 2023. DOI: 10.5281/ZENODO.7782262. [Online]. Available: https://zenodo.org/record/7782262.

[7] L. O. Bonino da Silva Santos, *FAIR Digital Object framework documentation*. [Online]. Available: https://web.archive.org/web/20240413094951/https://fairdigitalobjectframework.org/.

[8] M. Thompson, K. Burger, R. Kaliyaperumal, M. Roos, and L. O. Bonino da Silva Santos, "Making FAIR easy with FAIR tools: From creolization to convergence," *Data Intelligence*, vol. 2, pp. 87–95, Nov. 2019. DOI: 10.1162/dint_a_00031.

[9] S. Soiland-Reyes, C. Goble, and P. Groth, "Evaluating FAIR Digital Object and Linked Data as distributed object systems," *PeerJ Computer Science*, vol. 10, e1781, 2024.

[10] E. Schultes, B. Magagna, T. Kuhn, M. Suchánek, L. Bonino da Silva Santos, and B. Mons, "The comparative anatomy of nanopublications and FAIR Digital Objects," *Research Ideas and Outcomes*, vol. 8, Oct. 2022, ISSN: 2367-7163. DOI: 10.3897/rio.8.e94150.

[11] B. Mons and J. Velterop, "Nano-publication in the e-science era," vol. 523, 2009.

[12] "A promoter-level mammalian expression atlas," *Nature*, vol. 507, no. 7493, pp. 462–470, Mar. 2014, ISSN: 1476-4687. DOI: 10.1038/nature13182.

[13] T. Kuhn, R. Taelman, V. Emonet, H. Antonatos, S. Soiland-Reyes, and M. Dumontier, "Semantic micro-contributions with decentralized nanopublication services," *PeerJ Computer Science*, vol. 7, 2021, ISSN: 23765992. DOI: 10.7717/peerj-cs.387.

[14] T. Kuhn and M. Dumontier, "Trusty uris: Verifiable, immutable, and permanent digital artifacts for linked data," vol. 8465 LNCS, 2014. DOI: 10.1007/978-3-319-07443-6_27.

[15] B. Magagna, E. Schultes, M. Suchánek, and T. Kuhn, "FIPs and practice," *Research Ideas and Outcomes*, vol. 8, Oct. 2022. DOI: 10.3897/rio.8.e94451.

[16] I. Anders et al., *FAIR Digital Object technical overview*, en, 2023. DOI: 10.5281/ZENODO.7824714. [Online]. Available: https://zenodo.org/record/7824714.

[17] B. Magagna, E. Schultes, and T. Kuhn, *FAIR Implementation Profile (FIP) ontology*, 2020. [Online]. Available: https://web.archive.org/web/20240527195631/https://peta-pico.github.io/FAIR-nanopubs/fip/index-en.html.

[18] 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. 227–241.