

Updating Linked Data practices for FAIR Digital Object principles

Stian Soiland-Reyes^{‡§}, Leyla Jael Castro^l, Daniel Garijo[¶], Marc Portier[#], Carole Goble[‡], Paul Groth[§]

‡ Department of Computer Science, The University of Manchester, Manchester, United Kingdom

§ Informatics Institute, Faculty of Science, University of Amsterdam, Amsterdam, Netherlands

l Informationszentrum Lebenswissenschaften (ZB Med), Cologne, Germany

¶ Ontology Engineering Group, Universidad Politécnica de Madrid, Madrid, Spain

Vlaams Instituut voor de Zee (VLIZ), Oostende, Belgium

Corresponding author: Stian Soiland-Reyes (soiland-reyes@manchester.ac.uk)

Abstract

Background

The *FAIR principles* (Wilkinson et al. 2016) are fundamental for data discovery, sharing, consumption and reuse; however their broad interpretation and many ways to implement can lead to inconsistencies and incompatibility (Jacobsen et al. 2020).

The European Open Science Cloud ([EOSC](#)) has been instrumental in maturing and encouraging FAIR practices across a wide range of research areas. Linked Data in the form of [RDF](#) (Resource Description Framework) is the common way to implement machine-readability in FAIR, however the principles do not prescribe RDF or any particular technology (Mons et al. 2017).

FAIR Digital Object

FAIR Digital Object (FDO) (Schultes and Wittenburg 2019) has been proposed to improve researcher's access to digital objects through formalising their metadata, types, identifiers and exposing their computational operations, making them actionable FAIR objects rather than passive data sources.

FDO is a set of principles (Bonino et al. 2019), implementable in multiple ways. Current realisations mostly use *Digital Object Interface Protocol* (DOIPv2) (DONA Foundation 2018), with the main implementation [CORDRA](#). We can consider DOIPv2 as a simplified combination of object-oriented (CORBA, SOAP) and document-based (HTTP, FTP) approaches.

More recently, the [FDO Forum](#) has prepared detailed [recommendations](#), currently open for comments, including a [DOIP endorsement](#) and updated [FDO requirements](#). These point out **Linked Data** as another possible technology stack, which is the focus of this work.

Linked Data

[Linked Data standards](#) (LD), based on the Web architecture, are commonplace in sciences like bioinformatics, chemistry and medical informatics – in particular to publish Open Data as machine-readable resources. LD has become ubiquitous on the general Web, the [schema.org](#) vocabulary is used by over 10 million sites for indexing by search engines – [43% of all websites](#) use [JSON-LD](#).

Although LD practices align to FAIR (Hasnain and Rebholz-Schuhmann 2018), they do not fully encompass active aspects of FDOs. The HTTP protocol is used heavily for applications (e.g. mobile apps and cloud services), with REST APIs of customised [JSON structures](#). Approaches that merge the LD and REST worlds include [Linked Data Platform](#) (LDP), [Hydra](#) and [Web Payments](#).

Meeting FDO principles using Linked Data standards

Considering the potential of FDOs when combined with the mature technology stack of LD, here we briefly discuss how FDO principles in Bonino et al. (2019) can be achieved using existing standards. The general principles (G1–G9) apply well: Open standards with HTTP being stable for 30 years, JSON-LD is widely used, FAIR practitioners mainly use RDF, and a clear abstraction between the RDF model with stable bindings available in multiple serialisations.

However, when considering the specific principles (FDOF1–FDOF12) we find that additional constraints and best practices need to be established – arbitrary LD resources cannot be assumed to follow FDO principles. This is equivalent to how existing use of DOIP is not FDO-compliant without additional constraints.

Namely, persistent identifiers (PIDs) (McMurry et al. 2017) (FDOF1) are common in LD world (e.g. using <http://purl.org/> or <https://w3id.org/>), however they don't always have a declared type (FDOF2), or the PID may not even appear in the metadata. URL-based PIDs are resolvable (FDOF3), typically over HTTP using redirections and content-negotiation. One great advantage of RDF is that all attributes are defined semantic artefacts with PIDs (FDOF4), and attributes can be reused across vocabularies.

While CRUD operations (FDOF6) are supported by native HTTP operations (GET/PUT/POST/DELETE) as in [LDP](#), there is little consistency on how to define operation interfaces in LD (FDOF5). Existing REST approaches like [OpenAPI](#) and [URI templates](#) are mature and good candidates, and should be related to defined types to support machine-actionable composition (FDOF7). HTTP error code *410 Gone* is used in tombstone pages for removed resources (FDOF12), although more frequent is *404 Not Found*.

Metadata is resolved to HTTP documents with their own URIs, but these frequently don't have their own PID (FDOF8). [RDF-Star](#) and nanopublications (Kuhn et al. 2021) give ways to identify and trace provenance of individual assertions.

Different metadata levels (FDOF9) are frequently developed for LD vocabularies across different communities (FDOF10), such as [FHIR](#) for health data, [Bioschemas](#) for bioinformatics and >1000 more specific [bioontologies](#). Increased declaration and navigation of *profiles* is therefore essential for machine-actionability and consistent consumption across FAIR endpoints.

Several standards exist for rich collections (FDOF11), e.g. [OAI-ORE](#), [DCAT](#), [RO-Crate](#), [LD P](#). These are used and extended heterogeneously across the Web, but consistent machine-actionable FDOs will need specific choices of core standards and vocabularies. Another challenge is when multiple PIDs refer to “almost the same” concept in different collections – significant effort have created manual and automated semantic mappings (Baker et al. 2013, de Mello et al. 2022).

Currently the FDO Forum has suggested the use of LDP as a possible alternative for implementing FAIR Digital Objects (Bonino da Silva Santos 2021), which proposes a novel approach of content-negotiation with custom media types.

Discussion

The Linked Data stack provides a set of specifications, tools and guidelines in order to help the FDO principles become a reality. This mature approach can accelerate uptake of FDO by scholars and existing research infrastructures such as the European Open Science Cloud (EOSC).

However, the amount of standards and existing metadata vocabularies poses a potential threat for adoption and interoperability. Yet, the challenges for agreeing on usage profiles apply equally to DOIP as LD approaches.

We have worked with different scientific communities to define RO-Crate (Soiland-Reyes et al. 2022), a lightweight method to package research outputs along with their metadata. While RO-Crate’s use of schema.org shows just one possible metadata model, it’s powerful enough to be able to express FDOs, and familiar to web developers.

We have also used FAIR Signposting (Van de Sompel et al. 2022) with HTTP *Link*: headers as a way to support navigation to the individual core properties of an FDO (PID, type, metadata, licence, bytestream) that does not require heuristics of content-negotiation and is agnostic to particular metadata vocabularies and serialisations.

We believe that by adopting Linked Data principles, we can accelerate FDO today – and even start building practical ways to assist scientists in efficiently answering topical questions based on knowledge graphs.

Keywords

FAIR Digital Object, FDO, FAIR, FAIR Signposting, Linked Data, RDF, standards, best practices

Presenting author

Stian Soiland-Reyes

Presented at

First International Conference on FAIR Digital Objects, presentation

Acknowledgements

We would like to acknowledge the [RO-Crate community](#) and the [WorkflowHub Club](#). Thanks to Rudolf Wittner for valuable comments.

Funding program

Stian Soiland-Reyes is supported by the European Union's Horizon 2020 research and innovation programme under grant agreement numbers H2020-INFRAEDI-02-2018 823830 (BioExcel-2), H2020-INFRAEOSC-2018-2 824087 (EOSC-Life) and the Horizon Europe programme under grant agreements HORIZON-INFRA-2021-EMERGENCY-01 101046203 (BY-COVID), HORIZON-INFRA-2021-EOSC-01 101057344 (FAIR-IMPACT).

Leyla Jael Castro is supported by a German Research Foundation DFG grant for NFDI4DataScience.

Daniel Garijo is supported by the Madrid Government (Comunidad de Madrid-Spain) under the Multiannual Agreement with Universidad Politécnica de Madrid in the line Support for R&D projects for Beatriz Galindo researchers, in the context of the V PRICIT (Regional Programme of Research and Technological Innovation)

Author contributions

Author contributions to this article according to the Contributor Roles Taxonomy [CASRAI](#) [CrEDiT](#):

- **Stian Soiland-Reyes:** Conceptualization, Formal Analysis, Investigation, Software, Writing – original draft, Writing – review and editing
- **Leyla Jael Castro:** Writing – original draft
- **Daniel Garijo:** Conceptualization, Writing – review and editing
- **Marc Portier:** Investigation, Writing – original draft, Writing – review and editing
- **Carole Goble:** Supervision
- **Paul Groth:** Conceptualization, Supervision

Conflicts of interest

References

- Baker T, Bechhofer S, Isaac A, Miles A, Schreiber G, Summers E (2013) Key choices in the design of Simple Knowledge Organization System (SKOS). *Journal of Web Semantics* 20: 35-49. <https://doi.org/10.1016/j.websem.2013.05.001>
- Bonino da Silva Santos LO (2021) FAIR Digital Object Framework Documentation. <https://fairdigitalobjectframework.org/>. Accessed on: 2022-7-10.
- Bonino L, Wittenburg P, Carroll B, Hardisty A, Leggott M, Zwölf C (2019) FAIR digital object framework v1.02. <https://github.com/GEDE-RDA-Europe/GEDE/blob/master/FAIR%20Digital%20Objects/FDOF/FAIR%20Digital%20Object%20Framework-v1-02.docx>. Accessed on: 2022-7-10.
- de Mello BH, Rigo SJ, da Costa CA, da Rosa Righi R, Donida B, Bez MR, Schunke LC (2022) Semantic interoperability in health records standards: a systematic literature review. *Health and Technology* 12 (2): 255-272. <https://doi.org/10.1007/s12553-022-00639-w>
- DONA Foundation (2018) Digital Object Interface Protocol specification, version 2.0. <https://doi.org/0.DOIP/DOIPV2.0>
- Hasnain A, Rebholz-Schuhmann D (2018) Assessing FAIR Data Principles Against the 5-Star Open Data Principles. *Lecture Notes in Computer Science* 11155: 469-477. https://doi.org/10.1007/978-3-319-98192-5_60
- Jacobsen A, de Miranda Azevedo R, Juty N, Batista D, Coles S, Cornet R, Courtot M, Crosas M, Dumontier M, Evelo C, Goble C, Guizzardi G, Hansen KK, Hasnain A, Hettne K, Heringa J, Hooft RW, Imming M, Jeffery K, Kaliyaperumal R, Kersloot M, Kirkpatrick C, Kuhn T, Labastida I, Magagna B, McQuilton P, Meyers N, Montesanti A, van Reisen M, Rocca-Serra P, Pergl R, Sansone S, da Silva Santos LOB, Schneider J, Strawn G, Thompson M, Waagmeester A, Weigel T, Wilkinson M, Willighagen E, Wittenburg P, Roos M, Mons B, Schultes E (2020) FAIR Principles: Interpretations and Implementation Considerations. *Data Intelligence* 2 (1-2): 10-29. https://doi.org/10.1162/dint_r_00024
- 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>
- McMurry J, Juty N, Blomberg N, Burdett T, Conlin T, Conte N, Courtot M, Deck J, Dumontier M, Fellows D, Gonzalez-Beltran A, Gormanns P, Grethe J, Hastings J, Hériché J, Hermjakob H, Ison J, Jimenez R, Jupp S, Kunze J, Laibe C, Le Novère N, Malone J, Martin MJ, McEntyre J, Morris C, Muilu J, Müller W, Rocca-Serra P, Sansone S, Sariyar M, Snoep J, Soiland-Reyes S, Stanford N, Swainston N, Washington N, Williams A, Wimalaratne S, Winfree L, Wolstencroft K, Goble C, Mungall C, Haendel M, Parkinson H (2017) Identifiers for the 21st century: How to design, provision, and reuse persistent identifiers to maximize utility and impact of life science data. *PLOS Biology* 15 (6). <https://doi.org/10.1371/journal.pbio.2001414>

- Mons B, Neylon C, Velterop J, Dumontier M, da Silva Santos LOB, Wilkinson M (2017) Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud. *Information Services & Use* 37 (1): 49-56. <https://doi.org/10.3233/isu-170824>
- Schultes E, Wittenburg P (2019) FAIR Principles and Digital Objects: Accelerating Convergence on a Data Infrastructure. *Communications in Computer and Information Science* 3-16. https://doi.org/10.1007/978-3-030-23584-0_1
- Soiland-Reyes S, Sefton P, Crosas M, Castro LJ, Coppens F, Fernández J, Garijo D, Grüning B, La Rosa M, Leo S, Ó Carragáin E, Portier M, Trisovic A, RO-Crate Community, Groth P, Goble C (2022) Packaging research artefacts with RO-Crate. *Data Science* 5 (2). <https://doi.org/10.3233/ds-210053>
- Van de Sompel H, Klein M, Jones S, Nelson M, Warner S, Devaraju A, Huber R, Steinhoff W, Tykhonov V, Boruta L, Meijers E, Soiland-Reyes S, Wilkinson M (2022) FAIR Signposting Profile. 20220727 . signposting.org URL: <https://signposting.org/FAIR/>
- 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>