FAIR Connect 1 (2023) 29–34
DOI 10.3233/FC-230001
IOS Press

FSR Thumbnail Article

FAIR digital objects, persistent identifiers and machine actionability

Sharif Islam *
Naturalis Biodiversity Center, Leiden, The Netherlands
E-mail: sharif.islam@naturalis.nl; ORCID: https://orcid.org/0000-0001-8050-0299

Editor: Barbara Magagna (https://orcid.org/0000-0003-2195-3997)
Solicited reviews: André Ferreira (https://orcid.org/0000-0002-7838-0061); Shuai Wang (https://orcid.org/0000-0002-1261-9930)

Received 6 January 2023
Accepted 9 January 2023

Abstract. Based on the work of the Research Data Alliance and FAIR Digital Objects Forum, and examples from three different domains, this article highlights the importance and effectiveness of PID level metadata for FAIR implementation and machine actionability.

Keywords: Digital objects, machine readable, persistent identifiers, metadata, PID Kernel Record

Persistent identifiers (PID) and associated metadata are essential components of FAIR implementation [16]. By uniquely identifying digital objects and metadata, PID systems, services and associated policies ensure content accessibility and resolution. Therefore, providing the foundation for findability and accessibility – important aspects of FAIR Principle F1 that states both data and metadata must have a globally unique and persistent identifier [23]. Over the past several years, initiatives such as the PID Kernel Working Group of the Research Data Alliance (RDA) [18] and the Freya project [2] have identified and outlined the benefit of metadata in the context of data findability and linking. Some of these discussions have been continued and evolved within the conceptual framework of FAIR Digital Object [7]. The FAIR Digital Objects Forum [12] is now facilitating further specification work by looking into PID attributes and specific aspects of machine readability, interpretability, and actionability. The reason for these three distinctions is that structured content (such as JSON, JSON-LD, RDF, XML) is essential for machine readability but the machine also requires semantic artefacts to interpret the nature of the digital object (such as object type and relationship with other entities). And actionability here means machines can perform automated processing of digital objects [20].

The metadata elements linked to the PID record must enable operations at a higher abstraction level because FAIR Digital Objects are domain agnostic. The clear distinction between data and metadata, as outlined in FAIR principles F1 and F3, is adhered to by this higher abstraction. The abstraction also permits a further degree of separation

*Corresponding author: Darwinweg 2, 2333 CR Leiden, Nederland. Tel.: +310717519600; E-mail: sharif.islam@naturalis.nl.

Handle.Net®

Handle Values for: 10.1038/sdata.2016.18

Index	Type	Timestamp	Data
1	URL	2021-12-02 17:06:06Z	http://www.nature.com/articles/sdata201618
700050	700050	2019-11-12 05:24:20Z	20191112061430537
100	HS_ADMIN	2021-12-02 17:06:06Z	handle=0.na/10.1038; index=200; [delete hdl,read val,modify val,del val,add val,modify admin,del admin,add admin,list]

Fig. 1. Handle PID record: https://hdl.handle.net/10.1038/sdata.2016.18?noredirect of the FAIR principle article. These metadata can also be harvested via an API.

between metadata at the PID record level and that at the content/bitstream level. The PID record may include information such as the type of digital object (whether it is a dataset or a scholarly article), a list of permitted actions (read, write, and update), version data, and details about license and usage. The content level metadata for instance can contain other elements like a document's author, the file size, and image metadata, details about the variables in the datasets. While the PID level metadata can specify common components (either at the domain or specific use case level), the content level metadata can differ among digital objects. In other words, the PID level metadata provides a generic mechanism to describe different types of digital objects prior to accessing or resolving the content. The PID Kernel Working group defined the collection of these PID metadata as the PID Kernel Information [18]. Based on the RDA working group's output, a working draft from the FDO forum is proposing the term FDO Record to highlight that there could be possible implementation of FDO without explicitly relying on the attributes stored in a PID record [12]. DOI kernel [10] from the DOI Foundation and DataCite Metadata Kernel [5] also provide similar minimum metadata which are records needed to describe the digital object. These PID level metadata can be implemented in the PID system or a separate database if the PID system does not allow the functionality of adding extra metadata during the PID creation and registration. This article highlights the importance of PID records and PID level metadata for FAIR implementation and recommends adding PID records for all digital object types.

Along with the precise object definition, semantics [1], vocabularies [3], and provenance [13], different types of digital objects (e.g., data, software, concepts, algorithms, workflows etc.) and metadata are playing a more and more significant role in data intensive research. These building blocks are important for data and information to be "Fully AI-ready" (another interpretation of the FAIR acronym) where the concept of machine actionability is central. A 2022 paper entitled "FAIR Digital Twins for Data-Intensive Research" by Schultes et al. [19] proposes a modular approach to building systems based on FAIR Digital Objects where best practices adoption around identifiers and associated metadata will help.

There are currently several types of PID implementation available, including DOI, Handle, ARK, and several registration agencies that allow the use of various metadata schema. For example, in Fig. 1, the PID record of the 2016 FAIR principle journal article [23] shows only the URL and the timestamp rather than details about the type and nature of this digital object. Machine actionability is here limited due to a lack of information and specification, which prevents machines from inferring possible actions. Further information can be found via the Crossref API record [4] as Crossref was the DOI registration agency for this object. However, the DataCite API [6] returns a 404 error when someone searches for this article in DataCite. Additionally, the PID record in the Handle server (Fig. 1) does not provide the information that Crossref was used to register the DOI. Users can undoubtedly get this material from a multitude of sources. The point is that because programmatic operation support is currently dispersed, machine actionable operations are difficult to create. Adopting minimal metadata at the PID level, which is based on the FAIR Digital Object framework, is one option to improve this.

Below, I provide three examples of PID records from three different domains that show a few common elements and possibilities of different operations (two examples are borrowed from [21]).

1. DiSSCo PID record (Fig. 2)

The Distributed System of Scientific Collections (DiSSCo [9]) is a new Research Infrastructure (RI) for Natural Science Collections that uses the FAIR Digital Object concept as the core element for the data architecture [15]. In this example, the digitalObjectType states that this is a "Digital Specimen". The object types and type definitions can

Handle.Net®

Handle Values for: 20.5000.1025/KMX-F7Q-6UX

Index	Type	Timestamp	Data
1	**pid**	2022-09-12 06:32:59Z	https://hdl.handle.net/20.5000.1025/KMX-F7Q-6UX
2	**pidIssuer**	2022-09-12 06:32:59Z	{"id":"https://doi.org/10.22/10.22/2AA-GAA-E29","pidType":"DOI","primaryNameFromPid":"RA Issuing DOI"}
3	**digitalObjectType**	2022-09-12 06:32:59Z	{"id":"http://hdl.handle.net/21...","pidType":"Handle","primaryNameFromPid":"Digital Specimen"}
4	**digitalObjectSubtype**	2022-09-12 06:32:59Z	{"id":"https://hdl.handle.net/21...","pidType":"Handle","primaryNameFromPid":"ZoologyVertebrateSpecimen"
5	10320/loc	2022-09-12 06:32:59Z	<locations><location href="https://sandbox.dissco.tech/api/v1/specimens/20.5000.1025/KMX-F7Q-6UX" id="0" weight="0"/></locations>
6	**issueDate**	2022-09-12 06:32:59Z	2022-09-12
7	**issueNumber**	2022-09-12 06:32:59Z	1
8	**pidStatus**	2022-09-12 06:32:59Z	ACTIVE
11	**pidKernelMetadataLicense**	2022-09-12 06:32:59Z	https://creativecommons.org/publicdomain/zero/1.0/
14	**digitalOrPhysical**	2022-09-12 06:32:59Z	physical
15	**specimenHost**	2022-09-12 06:32:59Z	{"id":"https://ror.org/0566bfb96","pidType":"ROR","primaryNameFromPid":"Needs to be fixed!"}
100	HS_ADMIN	2022-09-12 06:32:59Z	handle=300:0.NA/20.5000.1025; index=200; [create hdl,delete hdl,create derived prefix,delete derived prefix,read val,modify val,del val,add val,modify admin,del admin,add admin,list]

Handle Proxy Server Documentation
Handle.net Web Site

Fig. 2. DiSSCo Digital Specimen https://hdl.handle.net/20.5000.1025/KMX-F7Q-6UX?noredirect.

be also stored as a FAIR Digital Object in a data type registry (see example [11]). Besides the type, the other metadata elements in this example describe the version of the PID record and the status of the record – as PID lifecycle information is important to understand the current status of the object. These elements are still under discussion as DiSSCo needs to accommodate different types of specimen records and metadata profiles (from zoology and botany to rocks and fossils). There are also similar discussions going on with the IGSN implementation [22] that deals with geological and earth science samples.

2. Persistent identification of instruments (Fig. 3)

The Persistent Identification of Instrument (PIDINST) schema came out of the RDA working group discussion with an interdisciplinary group. This schema provides a mechanism to register PID level records and it is PID system independent [17]. The example record shows specific information about a sensor. Each index also points to a Handle record (21.T11148/709a23220f2c3d64d1e1, [14]) in a data type registry. The goal is to describe instrument instances, including relations to other entities such as the model or manufacturing details in a more standard way as instruments are essential for different types of research activities. These metadata are important for reusability, reproducibility, and experiment validation. PID level records can help different workflows to figure out what the instrument is for instance or what measurements were taken in what form.

3. DARIAH dataset example (Fig. 4)

The Digital Research Infrastructure for the Arts and Humanities (DARIAH, [8]) aims to enhance and support digitally enabled research and teaching. The example uses key-value pairs with human readable data for the index. The metadata also includes information about who is responsible for the dataset and has pointers to the landing zone and the bit stream. Before needing to access the actual information, a machine entity can examine these structured records and choose a resolution or content negotiation processes.

4. Conclusion

The examples above show the importance of the separation between PID records and content level metadata. The examples also illustrate the possibility of describing digital object types using a type registry and finding a

Handle.Net®

Handle Values for: 21.T11998/0000-001A-3905-1

Index	Type	Timestamp	Data
1	URL	2021-04-14 14:02:37Z	https://linkedsystems.uk/system/instance/TOOL0022_2490/current/
2	21.T11148/8eb858ee0b12e8e463a5	2021-04-15 16:41:02Z	{"identifierValue":"http://hdl.handle.net/21.T11998/0000-001A-3905-1","identifierType":"Handle"}
3	21.T11148/9a15a4735d4bda329d80	2021-04-14 14:02:37Z	https://linkedsystems.uk/system/instance/TOOL0022_2490/current/
4	21.T11148/709a23220f2c3d64d1e1	2021-04-14 14:02:37Z	Sea-Bird SBE 37-IM MicroCAT C-T Sensor
5	21.T11148/4eaec4bc0f1df68ab2a7	2021-04-14 14:11:28Z	[{"owner": {"ownerName":"National Oceanography Centre","ownerContact":"louise.darroch@bodc.ac.uk","ownerIdentifier": {"ownerIdentifierValue":"http://vocab.nerc.ac.uk/collection/B75/current/ORG00009 /","ownerIdentifierType":"URL"}}}]
6	21.T11148/1f3e82ddf0697a497432	2021-04-14 14:05:56Z	[{"manufacturer":{"manufacturerName":"Sea-Bird Scientific","manufacturerIdentifier": {"manufacturerIdentifierValue":"http://vocab.nerc.ac.uk/collection/L35/current/MAN0013 /","manufacturerIdentifierType":"URL"}}}]
7	21.T11148/c1a0ec5ad347427f25d6	2021-04-14 14:02:37Z	[{"modelName":"Sea-Bird SBE 37 MicroCat IM-CT with optional pressure (submersible) CTD sensor series","modelIdentifier":{"modelIdentifierValue":"http://vocab.nerc.ac.uk/collection/L22/current/TOOL0022 /","modelIdentifierType":"URL"}}]
8	21.T11148/55f8ebc805e65b5b71dd	2021-04-14 14:02:37Z	A high accuracy conductivity and temperature recorder with an optional pressure sensor designed for deployment on moorings. The IM model has an inductive modem for real-time data transmission plus internal flash memory data storage.
9	21.T11148/f76ad9d0324302fc47dd	2021-04-14 14:02:37Z	["http://vocab.nerc.ac.uk/collection/L05/current/134/","http://vocab.nerc.ac.uk/collection/L05/current/350/"]
10	21.T11148/72928b84e060d491ee41	2021-04-14 14:05:56Z	[{"measuredVariable":{"variableMeasured":"http://vocab.nerc.ac.uk/collection/P01/current/CNDCPR01 /"}},{"measuredVariable":{"variableMeasured":"http://vocab.nerc.ac.uk/collection/P01/current/PSALPR01 /"}},{"measuredVariable":{"variableMeasured":"http://vocab.nerc.ac.uk/collection/P01/current/TEMPPR01 /"}},{"measuredVariable":{"variableMeasured":"http://vocab.nerc.ac.uk/collection/P01/current/PREXMCAT/"}}]
11	21.T11148/22c62082a4d2d9ae2602	2021-04-14 14:05:56Z	[{"date":{"dateValue":"1999-11-01","dateType":"Commissioned"}}]
12	21.T11148/eb3c713572f681e6c4c3	2021-04-14 14:05:56Z	[{"alternateIdentifier":{"alternateIdentifierValue":"2490","alternateIdentifierType":"serialNumber"}}]
13	21.T11148/178fb558abc755ca7046	2021-04-14 14:05:56Z	[{"relatedIdentifier":{"relatedIdentifierValue":"https://www.bodc.ac.uk/data/documents/nodb/pdf /37imbrochurejul08.pdf","relatedIdentifierType": "URL","relationType":"IsDescribedBy "}}]
100	HS_ADMIN	2021-04-14 14:02:37Z	handle=21.T11998/USER22; index=300; [create hdl,delete hdl,read val,modify val,del val,add val,modify admin,del admin,add admin]

Fig. 3. PID for instruments: https://hdl.handle.net/21.T11998/0000-001A-3905-1?noredirect.

Handle.Net®

Handle Values for: 21.11113/0000-000B-CA4C-D

Index	Type	Timestamp	Data
1	CREATOR	2017-12-07 20:59:10Z	PID Service pid-webapp-4.22.0.201711102014
2	ADM_MD	2017-12-07 20:59:19Z	https://repository.de.dariah.eu/1.0/dhcrud/21.11113/0000-000B-CA4C-D/adm
3	FILESIZE	2017-12-07 20:59:19Z	4802
4	RESPONSIBLE	2017-12-07 20:59:19Z	BeataMache@dariah.eu
5	CHECKSUM	2017-12-07 20:59:19Z	md5:d53305cfd84972afec2393bc9328c8b5
6	BAG	2017-12-07 20:59:19Z	https://cdstar.de.dariah.eu/public/EAEA0-E069-7925-F28A-0
7	PUBDATE	2017-12-07 20:59:19Z	2017-12-07 21:59:18 +0100
8	PROV_MD	2017-12-07 20:59:19Z	https://repository.de.dariah.eu/1.0/dhcrud/21.11113/0000-000B-CA4C-D/prov
9	URL	2017-12-07 20:59:19Z	https://repository.de.dariah.eu/1.0/dhcrud/21.11113/0000-000B-CA4C-D
10	DATA	2017-12-07 20:59:19Z	https://repository.de.dariah.eu/1.0/dhcrud/21.11113/0000-000B-CA4C-D/data
11	LANDING	2017-12-07 20:59:19Z	https://repository.de.dariah.eu/1.0/dhcrud/21.11113/0000-000B-CA4C-D/landing
12	SOURCE	2017-12-07 20:59:19Z	https://cdstar.de.dariah.eu/dariah/EAEA0-FA65-59AB-6BD6-0
13	INDEX	2017-12-07 20:59:19Z	https://repository.de.dariah.eu/1.0/dhcrud/21.11113/0000-000B-CA4C-D/index
14	METADATA	2017-12-07 20:59:19Z	https://repository.de.dariah.eu/1.0/dhcrud/21.11113/0000-000B-CA4C-D/metadata
15	TECH_MD	2017-12-07 20:59:19Z	https://repository.de.dariah.eu/1.0/dhcrud/21.11113/0000-000B-CA4C-D/tech
16	DOI	2017-12-07 20:59:19Z	http://dx.doi.org/10.20375/0000-000B-CA4C-D
17	INST	2017-12-07 20:59:19Z	2000
18	PUBLISHED	2017-12-07 20:59:40Z	true
100	HS_ADMIN	2017-12-07 20:59:10Z	handle=21.11113/USER02; index=1; [create hdl,delete hdl,read val,modify val,del val,add val,modify admin,del admin,add admin]

Handle Proxy Server Documentation
Handle.net Web Site

Fig. 4. DARIAH dataset: https://hdl.handle.net/21.11113/0000-000B-CA4C-D?noredirect.

balance between human and machine readability. Metadata at the PID level can provide benefits beyond basic linkage and resolution. The Handle system provides a simple and efficient way to create these records. Even though currently there is no large scale adoption and implementation of these PID records, experimentation and usage within different domains can be used for wider FAIR adoption. Combining these two categories of metadata allows for the achievement of a balance between the requirements of various use cases and the promotion of cross-domain interoperability. Domain experts, users, data stewards, data practitioners, and others with diverse perspectives must be consulted before decisions are made and actions are taken regarding these metadata. Adoption of PID level records helps FAIR implementation move toward a more standardized machine actionability and data-intensive operation arena as we continue to develop our specification for FAIR Digital Objects.

Acknowledgements

Many thanks to Dr. Ulrich Schwardmann and the FAIR Digital Objects Forum's Technical Specification and Implementation group for their work with FDO specification and the use of PID records. Also, Wouter Addink, Sam Leeflang, and Soulaine Theocharides from the Naturalis Biodiversity Center for working on the DiSSCo PID implementation.

Funding

The author reports no funding.

Conflict of interest

The author reports no conflict of interest.

References

[1] G. Coen, *Introduction to Semantic Artefacts. FAIRsFAIR*, 2019. doi:10.5281/zenodo.3549375.
[2] H. Cousijn, R. Braukmann, M. Fenner, C. Ferguson, R. van Horik, R. Lammey, A. Meadows and S. Lambert, Connected research: The potential of the PID graph, *Patterns* **2**(1) (2021), 100180. doi:10.1016/j.patter.2020.100180.
[3] S.J. Cox, A.N. Gonzalez-Beltran, B. Magagna and M.C. Marinescu, Ten simple rules for making a vocabulary FAIR, *PLoS computational biology* **17**(6) (2021), e1009041. doi:10.1371/journal.pcbi.1009041.
[4] Crossref API record, available from https://api.crossref.org/works/10.1038/sdata.2016.18 (Accessed: 03.11.2022).
[5] DataCite Metadata Schema, available from https://schema.datacite.org/meta/kernel-4.4/ (Accessed: 03.11.2022).
[6] DataCite REST API Guide, available from https://support.datacite.org/docs/api (Accessed: 03.11.2022).
[7] K. De Smedt, D. Koureas and P. Wittenburg, FAIR digital objects for science: From data pieces to actionable knowledge units, *Publications* **8** (2020), 21. doi:10.3390/publications8020021.
[8] Digital Research Infrastructure for the Arts and Humanities (DARIAH), available from https://www.dariah.eu/ (Accessed: 03.11.2022).
[9] Distributed System of Scientific Collections (DiSSCo), available from https://www.dissco.eu/ (Accessed: 03.11.2022).
[10] DOI metadata Kernel, available from, https://www.doi.org/doi_handbook/4_Data_Model.html#4.3.1 (Accessed: 03.11.2022).
[11] ePIC Data Type Registry, available from http://dtr-test.pidconsortium.eu/#objects/?query=hardisty (Accessed: 03.11.2022).
[12] FAIR Digital Object Forum, available from https://fairdo.org/ (Accessed: 03.11.2022).
[13] C. Goble, S. Cohen-Boulakia, S. Soiland-Reyes, D. Garijo, Y. Gil, M.R. Crusoe, K. Peters and D. Schober, FAIR computational workflows, *Data Intelligence* **2**(1–2) (2020), 108–121. doi:10.1162/dint_a_00033.
[14] Handle record of a sensor, available from https://hdl.handle.net/21.T11148/709a23220f2c3d64d1e1 (Accessed: 03.11.2022).
[15] S. Islam, A. Hardisty, W. Addink, C. Weiland and F. Glöckler, Incorporating RDA outputs in the design of a European research infrastructure for natural science collections, *Data Science Journal* **19**(1) (2020), 50. doi:10.5334/dsj-2020-050.
[16] N. Juty, S.M. Wimalaratne, S. Soiland-Reyes, J. Kunze, C.A. Goble and T. Clark, Unique, persistent, resolvable: Identifiers as the foundation of FAIR, *Data Intelligence* **2**(1–2) (2020), 30–39. doi:10.1162/dint_a_00025.
[17] R. Krahl, L. Darroch, R. Huber, A. Devaraju, J. Klump, T. Habermann and M. Stocker, The research data alliance persistent identification of instruments working group members, in: *Metadata Schema for the Persistent Identification of Instruments. Research Data Alliance*, 2021. doi:10.15497/RDA00070.

[18] RDA PID KI, RDA Recommendation on PID Kernel Information. Research Data Alliance, 2019. doi:10.15497/RDA00031.
[19] E. Schultes, M. Roos, L.O.B. da Silva Santos, G. Guizzardi, J. Bouwman, T. Hankemeier, A. Baak and B. Mons, FAIR digital twins for data-intensive research, *Frontiers in Big Data* **5** (2022).
[20] U. Schwardmann, Digital objects – FAIR digital objects: Which services are required?, *Data Science Journal* **19**(1) (2020), 15. doi:10.5334/dsj-2020-015.
[21] U. Schwardmann and T. Kálmán, Two examples on how FDO types can support machine and human readability, *Research Ideas and Outcomes* **8** (2022), e96014. doi:10.3897/rio.8.e96014.
[22] Technical Documentation of the IGSN, available from https://igsn.github.io/metadata/ (Accessed: 03.11.2022).
[23] M. Wilkinson, M. Dumontier, I. Aalbersberg et al., The FAIR guiding principles for scientific data management and stewardship, *Scientific data* **3**(1) (2016), 1–9. doi:10.1038/sdata.2016.18.