1. FAIR PIDS: THE ROLE OF ORCID IN STRENGTHENING THE FAIR PRINCIPLES

Paloma Marín-Arraiza[60]
Ana Heredia[61]

1.1 INTRODUCTION

The idea behind a Persistent Identifier System (PID system) is to offer a lasting reference of an entity (physical, digital, or abstract), for instance, a digital document, website, person, or institution. Some well-known PID systems are Archival Resource Key (ARK), Digital Object Identifier (DOI), Handle system, Persistent Uniform Resource Locator (PURL), Uniform Resource Name (URN) and Open Researcher and Contributor ID (ORCID iD), the latest exclusively for people.

A PID has a series of associated machine-readable metadata; therefore, they identify the object but not its location, as it happens with a URL (Dappert *et al.*, 2017). A PID can be implemented following the HTTP protocol, which makes it actionable and allows directing the reader to the page Where the resource can be found (*landing page*) (López-Pellicer *et al.*, 2016; Van de Sompel *et al.*, 2014).

However, it is important to point out that the persistence is related to the service offered by the system and not to the identifier itself. This means that an entity commits itself to keep the identifier resolvable. The identifier leads the users to the services that guarantee reference (Kunze, 2013). For example, the ARKs can be maintained and resolved through the EZID service (University of California); the DOIs are generated by the International DOI Foundation and its correspondent registration agency, such as Crossref and DataCite and data centers; the Handles are managed by the Corporation for National Research Initiatives (CNRI); and the PURL system was developed by the Online Computer Library Center (OCLC).

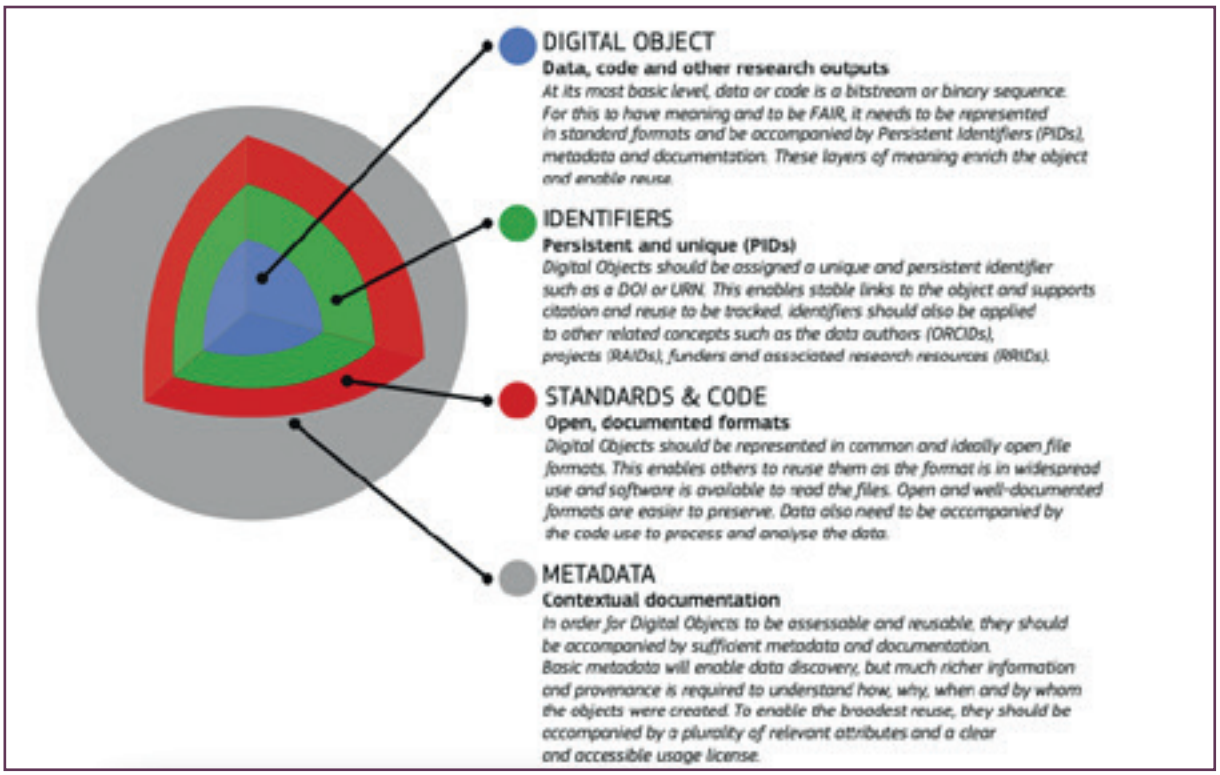
The use of PIDs in archives and research information systems is currently generalized, and PIDs are considered an essential part of the preservation process. For that reason, several research institutions have created data centers to register PIDs, with the aim to preserving their contents and turn them internationally findable and editable. The data center responsible for issuing a PID – for example a research library – must also perform the digital curation to guarantee the maintenance of the resource metadata (Johnston *et al.*, 2018).

60 Author details: PhD in Information Science (São Paulo State University), Master in Information and Scientific Communication (University of Granada), Bachelor in Physics (University of Granada), ORCID, p.arraiza@orcid.org, https://orcid.org/0000-0001-7460-7794.

61 Author details: PhD in Sciences (Université Libre de Bruxelles), Master in Cognitive Sciences and Neurosciences (Université Paul Sabatier), Bachelor in Biology (University of Santa Úrsula), Independent information consultant, heredia.a@gmail.com, https://orcid.org/0000-0001-7862-8955.

In fact, current guidelines indicate the use of PIDs, as it is the case of the first FAIR principle: "(Meta)data are assigned as globally unique and persistent identifiers". The report "Turning FAIR into a reality" proposes a FAIR Data Object model (European Commission, 2018), whose layers consist of metadata, standards, identifiers, and data.

Figure 1 – FAIR Data Object Model



Source: European Commission (2018, p.35).

To understand FAIR Data Object, the authors claim that:

> Data needs to be accompanied by Persistent Identifiers (PIDs) and metadata rich enough to enable them to be reliably found, used and cited. In addition, the data should be represented in common – and ideally open – file formats, and be richly documented using metadata standards and vocabularies adopted by the given research communities to enable interoperability and reuse. Sharing code is also fundamental and should include not just the source itself but also appropriate documentation including machine-actionable statements about dependencies and licensing (European Commission, 2018, p.35).

In addition to the identification, PIDs are used to add resources. The research results as a PID are easier to track, which makes the research monitoring activities easier. However, as mentioned earlier, the persistence is not an intrinsic characteristic of a PID, but it is related to the underlying service.

In this matter, it is possible to talk about "reliable identifiers" that are – in addition to persistent – unique, descriptive, interoperable and governed. The ODIN consortium (ORCID and DataCite Interoperability Network) proposed the following characteristics for the reliable identifiers:

a. unique on a global scale, allowing large numbers of unique identifiers;

b. resolve as HTTP URI's with support for content negotiation, and these HTTP URI's should be persistent;

c. come with metadata that describes their most relevant properties, including a minimum set of common metadata elements. A search of metadata elements across all trusted identifiers of that service should be possible;

d. are interoperable with other identifiers through metadata elements that describe their relationship;

e. Are issued and managed by an organization that focuses on that goal as its primary mission. The organization has a sustainable business model and a critical mass of member organizations that have agreed to common procedures and policies, has a trusted governance structure, and is committed to using open technologies (Ariani *et al.*, 2015, p. 19).

In addition, PIDs work as mechanisms of credit and assignment, when citing the results of the research (McMurry *et al.*, 2017). According to Wilkinson *et al.* (2016), the scientific infrastructures – for example, repositories, supercomputers or physical equipment – can also receive a PID.

1.2 FAIR PIDs AND LEVELS OF MATURITY

PIDs can be internal – when used inside an organization; for example, an employee or student identifier —, owner —When used in a single system; for example, the Scopus Author ID — or open—when they present a complete interoperability with other systems and identifiers; for example, a ORCID iD, a DOI or a Uniform Research Identifier (URI). They allow the establishment of reliable connections among resources.

Demeranville (2018) also defines FAIR PIDs, adding more desirable characteristics to PID systems:

> **FAIR PIDs:** These PIDs are not just resolvable, but can also be used to discover open, interoperable, well-defined metadata containing provenance information predictably. They are openly governed for of the community. Example: DOIs are stored either as URLs "https://doi.org/10.1/123", or simply "10.1/123". We present these to the user as links in the Registry and you can also follow those links to discover metadata describing the linked item. DOIs are governed by the International DOI Foundation and the attached metadata is available under a CC0 license, meaning that it is open to everyone. The metadata contains information about the publisher, the publication, other authors, funding, and affiliation(s), all of which help establish the provenance of the item. Other FAIR PIDs include arXiv identifiers, PubMed and PubMed Central identifiers and most ISBN identifiers. (Demeranville, 2018).

In this matter, it is important to note the maturity of the infrastructure behind these PIDs. We can consider the maturity of an infrastructure when it is in common use in the research Community and across disciples of knowledge. According to the research developed by Ferguson *et al.* (2018) in the framework of FREYA project, Only the "researcher", "publication" and "data" entities currently have mature PID systems.

The following table (Table 1) shows those PIDs whose infrastructure has a high level of maturity.

Table 1 – Entities, types of PIDs and their maturity

Research Entity	Types of PIDs used	Maturity of PIDs infrastructure
Publication	DOI, Accession number, Handle, URN, Scopus EID, Web of Science UID, PMID, PMC, arXiv Identifier, BibCode, ISSN, ISBN, PURL	Mature
Researcher (or scholar)	ORCID iDs, ISNI (also DAIs, VIAFs, arxivIDs, Open IDs, Researcher IDs, Scopus IDs)	Mature
Data	DOI, Accession number, Handle, PURL, URN, ARK	Mature

Source: Adapted from Ferguson *et al.* (2018, p. 9-10).

ORCID iDs are part of this mature research infrastructure and also contribute to FAIRfication of research data, as described below.

1.3 FINDABLE E INTEROPERABLE: THE ROLE OF THE ORCID

FAIR principles guide the entire data releasing process to make them findable - F, accessible – A, interoperable – I and reusable – R.

The role of ORCID in the context of FAIR principles is understood, since ORCID iD acts as an international standard in the persistent identification of authors. Table 2 presents this contribution for each FAIR principle.

Table 2 – The role of ORCID in the FAIR principles

Principle	Description[62]	ORCID Contribution
F1	(Meta)data receives a globally unique and persistent identifier	Providing ORCID iD as PID for "author"/" creator" and "collaborator" .
F2	Data are described in rich metadata[63]	Detail of the provenance information.
F3	Metadata clearly and explicitly includes the data identifier they describe	Inclusion of PIDs in all entries in the ORCID record.
F4	(Meta)data are recorded or indexed in a researchable resource	Availability of ORCID Public API[64] for queries. Annual publication of the ORCID public data archive [65] .
I1	(Meta)data use a formal, accessible, shared and widely applicable language for knowledge representation.	Recognition of ORCID iD as a norm ISO 27729:2012[66]. Use of international standards for the record building (e.g., CASRAI[67]).
I3	(Meta)dados include references qualified to other (meta)data.	Presentation using HTTPS PIDs[68] to discover metadata that describe the linked item

Source: Designed by the authors.

Therefore, the use of ORCID iDs (authenticated[69] if possible), contributes to the FAIRfication process. Furthermore, together with the work to improve the quality and completeness of metadata contained in ORCID records, this will make easier for ORCID to become a reliable source of FAIR data.

1.4 FINAL CONSIDERATIONS

62 Obtained and translated from https://www.go-fair.org/fair-principles/

63 It refers to the fact of having relevant atributes and provenance information a

64 ORCID Public API: https://members.orcid.org/api/about-public-api

65 ORCID Public Data File 2020: https://doi.org/10.23640/07243.13066970.v1

66 ISSO 27729:2012. Information and documentation – International standard name identifier (ISNI) https://www.iso.org/standard/44292.html

67 CASRAI: https://casrai.org/

68 Identifiers contained in ORCID records: https://pub.orcid.org/v3.0/identifiers

69 Collection process of an ORCID iD authenticated: https://members.orcid.org/api/tutorial/get-orcid-id

This text intends to present some points about PIDs and their importance in the context of FAIR data, as well as the role of ORCID in the process of data FAIRfication.

ORCID, as a non-profit organization and open infrastructure provider, continues to develop and align its work with improving metadata quality and supporting research communities. FAIR principles also underpin some of this work.

REFERENCES

ARIANI, A.; BARTON, A. J.; BRASE, J.; BROWN, J.; DEMERANVILLE, T.; HERTERICH, P.; MCAVOY, L.; PAGLIONE, L.; RUIZ, S.; THORISSON, G.; VISION, T.; ZIEDORN, F.. **D4.2**: Workflow for interoperability. [*S. l.*]: Figshare, 2015. Available from: https://figshare.com/articles/D4_2_Workflow_for_interoperability/1373669/1. Access on: 26 June 2020.

DAPPERT, A.; FARQUHAR, A.; KOTARSKI, R.; HEWLETT, K.. Connecting the Persistent Identifier Ecosystem: Building the Technical and Human Infrastructure for Open Research. **Data Science Journal**, v. 16, n. 28, p. 1-16, 15 June 2017. DOI 10.5334/dsj-2017-028. Available from: http://datascience.codata.org/articles/10.5334/dsj-2017-028/. Access on: 26 June 2020.

DEMERANVILLE, T. **Blog**: Building a Robust Infrastructure, One PID at a Time. [*S. l.*], 7 Aug. 2018. Available from: https://orcid.figshare.com/articles/Blog_Building_a_Robust_Infrastructure_One_PID_at_a_Time/7008101/1. Access on: 1 Nov. 2020.

EUROPEAN COMMISSION. Directorate General for Research and Innovation. **Turning FAIR into reality**: final report and action plan from the European Commission expert group on FAIR data. Luxembourg: Publications Office, 2018. Available from: https://data.europa.eu/doi/10.2777/1524. Access on: 1 Nov. 2020.

FERGUSON, C.; MCENTRYE, J.; BUNAKOV, V.; LAMBERT, S.; SANDT, S. V. D.; KOTARSKI, R.; STEWART, S.; MACEWAN, A.; FENNER, M.; CRUSE, P.; HORIK, R. V.; DOHNA, T.; KOOP-JACOBSEN, K.; SCHINDLER, U.; MCCAFFERTY, S. **D3.1 Survey Of Current Pid Services Landscape**. Version 1. 17 July 2018. DOI: 10.5281/ZENODO.1324296. Available from: https://zenodo.org/records/1324296. Access on: 1 Nov. 2020.

JOHNSTON, L. R.; CARLSON, J.; HUDSON-VITALE, C.; IMKER, H.; KOZLOWSKI, W.; OLENDORF, R.; STEWART, C.. How Important is Data Curation? Gaps and Opportunities for Academic Libraries. **Journal of Librarianship and Scholarly Communication**, Ames, Iowa, v. 6, n. 1, p. 2198, Apr. 2018. DOI: https://doi.org/10.7710/2162-3309.2198. Available from: https://jlsc-pub.org/article/10.7710/2162-3309.2198/. Access on: 9 Apr. 2019.

KUNZE, J. **The ARK Identifier Scheme**. [*S. l.*]: California Digital Library, 2013. Available from: https://tools.ietf.org/html/draft-kunze-ark-18. Access on: 9 Mar. 2019.

LÓPEZ-PELLICER, F. J.; BARRERA, J.; GONZÁLEZ GARCÍA, J.; ZARAZAGA-SORIA, F. J.; LÓPEZ ROMENO, E.; ABAD POWER, P.; RODRIGUEZ PASCUAL, A. F. El desafío de los identificadores persistentes y accionables. **Mapping**,

n. 180, p. 32-41, 2016. Available from: http://www.jiide.org/Jiide-theme/resources/docs/pdf/articulos/09_art_IAAA_IdentificadoresPersistentesAccionables.pdf. Access on: 10 Mar. 2019.

MCMURRY, J. A. *et al*. Identifiers for the 21st century: How to design, provision, and reuse persistent identifiers to maximize utility and impact of life science data. **PLOS Biology**, *San Francisco*, v. 15, n. 6, p. e2001414, 29 June 2017. DOI: https://doi.org/10.1371/journal.pbio.2001414. Available from: https://dx.plos.org/10.1371/journal.pbio.2001414. Access on: 1 Nov. 2020.

VAN DE SOMPEL, H.; SANDERSON, R.; SHANKAR, H.; KLEIN, M. Persistent Identifiers for Scholarly Assets and the Web: The Need for an Unambiguous Mapping. **International Journal of Digital Curation**, Edinburgh, v. 9, n. 1, p. 331–342, jun. 2014. DOI 10.2218/ijdc.v9i1.320. Available from: http://www.ijdc.net/article/view/9.1.331. Access on: 9 Mar. 2019.

WILKINSON, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. **Scientific Data**, [*S. l.*], v. 3, p. 160018, Mar. 2016. DOI: https://doi.org/10.1038/sdata.2016.18. Available from: https://www.nature.com/articles/sdata201618. Access on: 9 Mar. 2019.

How to cite this chapter: MARÍN-ARRAIZA, Paloma; HEREDIA, Ana. FAIR PIDS: the role of ORCID in strengthening the air principles. *In*: SALES, Luana Faria; VEIGA, Viviane Santos de Oliveira; VIDOTTI, Silvana Aparecida Borsetti Gregório; HENNING, Patrícia; SAYÃO, Luís Fernando (org.). **FAIR Principles Applied To Research Data Management**: Brazilian Experiences. Brasília, DF: Editora Ibict, 2024. cap. 1, p. 23-29. DOI: 10.22477/9788570131959.cap1.