1st Conference on Research Data Infrastructure

Poster presentations II

https://doi.org/10.52825/CoRDI.v1i.361

Published: 07 Sept. 2023

PIDs in the Natural Sciences

Thomas Schörner[1][https://orcid.org/0000-0002-7213-0352], Anton Barty[1], Markus Demleitner[2], Harry Enke[3][https://orcid.org/0000-0002-2366-8316] , Oliver Koepler[4][https://orcid.org/0000-0003-3385-4232] , Martin Köhler[1][https://orcid.org/0000-0003-0617-3319] , Bridget Murphy[5][https://orcid.org/0000-0002-1354-2381] , Sonja Schimmler[6][https://orcid.org/0000-0002-8786-7250] and Lisa-Marie Stein[1][https://orcid.org/0000-0001-7905-0462]

[1] Deutsches Elektronen-Synchrotron DESY, Germany

[2] Universität Heidelberg, Germany

[3] Leibniz-Institut für Astrophysik Potsdam AIP, Germany

[4] TIB Leibniz-Informationszentrum Technik und Naturwissenschaften Universitätsbibliothek, Germany

[5] Christian-Albrechts-Universität zu Kiel, Germany

[6] Fraunhofer-Institut für offene Kommunikationssysteme FOKUS, Germany

Abstract. We report on ongoing discussions and plans of NFDI consortia in physics and reltated natural sciences with respect to (persistent) identifiers.

Keywords: Persistent identifiers, Physical sciences, Natural sciences, FAIR principles, PID guidelines, Knowledge distribution

1. Introduction

Identifiers, and in particular persistent identifiers (PIDs), are a ubiquitous phenomenon, a desideratum, a challenge and a requirement in all branches of science, as e.g. reflected by NFDI working group of the NFDI section “Common Infrastructures” on the topic, and a topical proposal for the first call of the Base4NFDI initiative.

In the physical sciences – and with a little broader horizon: in the natural and technical sciences – such identifiers are used for a wide variety of entities, tangible and digital: data sets, publications, software, samples of certain materials or chemical compounds, hardware devices from individual chips to entire detectors, beamline instrumentations, …

When choosing adequate identifiers to be used in the digital context, one has to consider several questions:

First and foremost, one has to clarify the purpose of the identifier, and its scope, or level of detail. If the identifier stands for a limited collection of data sets, there is almost always a lean procedure, available from the immediate scientific context, to provide order and identification. Lifting such a collection out of this context, for example through further processing steps, may make additional identifiers (and metadata) necessary. In German, there is the term “Erschließungstiefe” that summarises the required careful deliberation process on which information to carry on. This is particularly important if the identifier points to a real-world object (with its potentially infinite depth of detail).

The next question is associated with the purpose of the additional identifiers referred to above. Here, the FAIR principles can provide ample guidance. The answer should also contain information on whether there are already identifiers in use within the community / discipline in question that might easily serve the purpose.

The third question to consider is connected to the (intended or expected) persistence of both, the identifier itself and the object that is identified. For physical samples as objects, there is additionally the issue of the lifetime of the sample to be taken into account. For digital objects (e.g. datasets) their life-cycle must be taken into account and properties should match those of the digital identifier.

For the digital identifier itself it needs to be clarified if it should have the property to be persistent and resolvable. In all three cases the question of (institutional) commitment (e.g. DOI) must be raised.

The topic of (persistent) identification leads furthermore to the necessary awareness of defining the status of the data the scientist is dealing with. Because once it is classified which type of data we are addressing – ranging from primary to secondary, from raw experimental data to derived data sets – the question arises which outputs should be published and which should only be given a public status within the individual infrastructure. (Persistent) identifiers are a fundamental principle in the publication process that need to be further promoted. But also at this point it is necessary to create an awareness for their use, since type and use are also conditioned by the juxtaposition of *public* and *published*. In summary, for a community, the choice of identifiers also requires clarification of the criteria that determine the decision to publish an object.

2. Physical Sciences

In the physical sciences, a structured discussion on the matter of PIDs was started some time ago between players from various structures in the field:

- The NFDI consortia DAPHNE4NFDI and PUNCH4NFDI cover the disciplines of photon and neutron science, and of particle, astroparticle, hadron&nuclear physics and astrophysics. The issue of metadata and of identifiers plays a significant role in both their work programmes.
- Within the DIG-UM community self-organisation for the ErUM-Data action plan, eight sub-communities and their research committees are represented and work, among other things, on “Research Data” within the topic group with the same name: Astroparticle physics [1], elementary particle physics [2], accelerator physics [3], research with neutrons [4], research with synchrotron radiation [5], research with nuclear probes and ions [6], hadron and nuclear physics [7], Rat Deutscher Sternwarten [8].

There is, naturally, significant overlap between these structures.

3. Other Natural Sciences

In the context of natural sciences, NFDI4Chem and NFDI4Cat are bringing similar and additional aspects of PID applications to the table:

- Referencing data from Electronic Lab Notebooks, chemical reactions, collections of datasets derived from a unique sample;
- referencing analytical instruments and their configuration used to generate data.

4. Goals

The goals of the discussion are manifold:

- Knowledge distribution: The discussion between players from different branches of physics, and with representatives from neighbouring disciplines like chemistry, material science, etc. – is ideal for spreading knowledge on PIDs and of discussing common views and differences, thus contributing to a more informed use of identifiers.
- Use case collection: A collection of use cases is well suited to illustrate the breadth of PID applications and choices and can thus cross-fertilise expertise from various communities.
- Development of PID guidelines for the field, with recommendations for the large science system and the entire NFDI. To this end, the connection to the NFDI working group is maintained in the discussion.
- On the practical side, a list of guiding questions is being designed as a practical help for individual researchers that are confronted with the challenge of designing PIDs for their objects.

The topics mentioned above are being collected in a white paper, conceived as a living document that is constantly updated and maintained.

This contribution to the CoRDI conference will function as a progress report and summarise the status of the discussion on PIDs in the physical sciences and of the white paper, connecting this discussion to the wider NFDI scope.

Data availability statement

There are no data underlying this contribution, which is entirely based on information discussions among members of various NFDI consortia and NFDI sections.

Author contributions

TS and HE wrote the original draft, MD, OK, MK, LMS reviewed and edited it, AB, BM and SS reviewed and commented on it. All participated in the discussions that led to this submission.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the DAPHNE4NFDI, PUNCH4NFDI, NFDI4Cat and NFDI4Chem NFDI consortia.

The DAPHNE4NFDI project is funded by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) in the NFDI Funding Programme (grant no. 460248799).

The NFDI4Cat project is funded by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) in the NFDI Funding Programme (grant no. 441926934).

The NFDI4Chem project is funded by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) in the NFDI Funding Programme (grant no. 441958208).

The PUNCH4NFDI project is funded by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) in the NFDI Funding Programme (grant no. 460248186).

Acknowledgement

We kindly acknowledge the in-kind support of DESY, Universität Heidelberg, AIP, TIB, Fraun-Hofer FOKUS for the presented work.

References

1. “Komitee für Astroteilchenphysik KAT”. https://www.astroteilchenphysik.de/das-kat.html (26 April 2023)
2. “Komitee für Elementarteilchenphysik KET”. https://www.ketweb.de (26 April 2023)
3. “Komitee für Beschleunigerphysik KfB”. https://www.beschleunigerphysik.de/de/ (26 April 2023)
4. “Komitee Forschung mit Neutronen KFN. https://www.sni-portal.de/de/nutzervertretungen/komitee-forschung-mit-neutronen (26 April 2023)
5. “Komitee Forschung mit Synchrotronstrahlung KFS”. https://www.sni-portal.de/de/nutzervertretungen/komitee-forschung-mit-synchrotronstrahlung (26 April 2023)
6. “Komitee Forschung mit nuklearen Sonden und Ionenstrahlen KFSI”. https://www.sni-portal.de/de/nutzervertretungen/komitee-forschung-mit-nuklearen-sonden-und-ionenstrahlen (26 April 2023)
7. “Komitee für Hadronen- und Kernphysik KHuK”. https://khuk.uni-mainz.de (26 April 2023)
8. “Rat deutscher Sternwarten RdS”. https://www.astronomische-gesellschaft.de/de/rat-deutscher-sternwarten (26 April 2023)