

Persistent Identifier (PID) a journey of making data machine-actionable

Authors:

Preston Rodrigues^{1*}, David Linke² and Thomas Bönisch²

*Lead presenter

¹preston.rodrigues@hirs.de, High-Performance Computing Center Stuttgart (HLRS), University of Stuttgart, Germany

²Leibniz-Institut für Katalyse e.V. (LIKAT), Rostock, Germany

Abstract:

We are in an age of digital convergence. This has resulted in a growing demand for data and services that can be accessed from any device at any time. However, finding and linking the data reliably has been a continuous challenge. FAIR principles states findability as one of key pillar in the journey of making data machine-actionable.

This journey is not new. A good example is a published books with its International Standard Book Number (ISBN). However, in this age of rapid digitization it is also imperative to uniquely identify digital object. In order to distinctly identify a digital object, it needs to have an universally unique identifier. Some of the well know techniques used in industry to uniquely identify digital object is Digital Object Identifiers (DOI)¹ and Handle System².

The Handle System includes an open set of protocols, an identifier space and way to modify the associated digital data. The protocol enables a distributed computer system to store identifiers of digital resources and resolve those identifiers into the information necessary to locate and access the resources. The associated information can be changed as needed to reflect the current state of the identified resource without changing the identifier, thus allowing the ID of the item to persist over changes to its location or content.

In NFDI4Cat project, we plan to leverage the flexibility of the Handle system to not only provide PID to digital object but enable linking different digital objects that use the PID. We proposes a unique solution combining group and permission based access control and domain specific digital object subscription. This combination ensures strict data privacy, a must for both academia and industry. Making sure only those PID that are made explicitly public viewable can be accessed. Furthermore, to realizing our journey towards machine-actionable data we are leveraging on linkML³. LinkML allows to define a standardized schema and express data and its semantics in a machine-readable format. This ensure data consistency, integrity, and interoperability across different systems, tools, and domains.

As a proof-of-concept we have setup a standalone Handle server⁴ on public cloud infrastructure. The handle prefix is register with Corporation for National Research Initiatives (CNRI) hence, any publicly readable values stored in the handle system will be accessible to all. NFDI consortia members can request for a domain specific space (4cat, 4chem, 4earth, 4ds, 4hum). Using REST API they can mint and manage their PIDs. The REST API provides secure access to the domain specific space enabling flexible PID integration into your existing of future solutions. This Poster will present our current progress towards our journey of making data machine-actionable.

References:

- [1] DOI Homepage. URL: <https://www.doi.org/>, last access on 15.05.2024.
- [2] Handle Homepage. URL: <https://www.handle.net/>, last access on 15.05.2024.
- [3] linkML Homepage. URL: <https://linkml.io/>, last access on 15.05.2024.
- [4] NFDI4Cat API Homepage. URL: <https://api.nfdi4cat.org> ,last access on 15.05.2024.

Keywords: PID, LinkML, API, Handle.net, permission based access, subscription