Extended Abstract for the Eurocriss

A PID Metaresolver to enable accurate citation and attribution of digital objects by providing reliable access to metadata and associated resources

Authors: Sven Bingert (GWDG), Themis Zamani (GRNET) 0000-0003-1148-1738, Konstantinos Kagkelidis (GRNET) 0000-0002-2449-7045, Fotios Basios (GRNET) - 0009-0003-6339-3278, Serafeim Chatzopoulos (OpenAIRE AMKE) 0000-0003-1714-5225 , Ali Reza Sajedi

Introduction

Persistent Identifiers (PIDs) are elementary ingredients in nowadays and future research data management processes. Looking at various data life cycles, identifiers help to achieve certain goals in a very efficient way, supporting steps to find, relate, integrate and publish data. If you look at these activities, it becomes clear that the FAIR criteria also support internal data processing, even before publication. In fact, various FAIR criteria explicitly mention the use of Persistent Identifiers. But the landscape of Persistent Identifiers is quite complex. Historically, different systems were established and still play important roles in various disciplines (e.g. see Klump&Huber 2017[1]). Those PID systems have specific properties, and provide specific information, in respect of their resolution method. To include those in data management procedures, or more importantly in data analytics processes, an in-depth understanding and specialized software is required. It would require considerable effort to implement all the necessary interfaces to be able to resolve different PID systems.

To ease the use of Persistent Identifiers in the research process, a generic resolver that can communicate and provide a seamless interaction with different identifier systems would be of great importance. Such a generic resolver is called a Metaresolver. As mentioned in the Strategic Research and Innovation Agenda (SRIA)[2] of the European Open Science Cloud (EOSC) in the area of further developments:
"Each PID provider provides its own resolver, while a meta resolver could form a single service which can recognise different PID types and redirect to the appropriate resolver, regardless of issuer." The development of a PID Metaresolver was categorized as a priority and is part of the objectives outlined in the aforementioned report.
In the context of the EU funded project FAIRCORE4EOSC[3], that establishes new core components identified by the SRIA, we developed a PID Metaresolver[4] that we provide as

[1] Klump, J and Huber, R 2017 20 Years of Persistent Identifiers – Which Systems are Here to Stay? Data Science Journal, 16: 9, pp. 1–7, DOI: https://doi.org/10.5334/dsj-2017-009

[2] European Commission, Directorate-General for Research and Innovation, Strategic Research and Innovation Agenda (SRIA) of the European Open Science Cloud (EOSC), Publications Office of the European Union, 2022, https://data.europa.eu/doi/10.2777/935288

[3] https://faircore4eosc.eu

[4] https://pidmr.argo.grnet.gr/

open source[5], as part of the Components Beta release (November 2023)[6]. To the best of our knowledge, this is the first attempt of such a Metaresolver.

The PID Metaresolver

The Persistent Identifier Metaresolver (PIDMR) is a service that aggregates information about various types of persistent identifiers and provides a means to resolve them to the associated digital objects or resources. When a user submits a PID to the Metaresolver, it first identifies the type of Persistent Identifier; then, it contacts the relevant PID provider service to retrieve the requested information, i.e., metadata, landing page or resource location, for the given PID. Once the mode of the PID (e.g. metadata) for the requested PID has been retrieved, the metaresolver displays this information to the user, either through a web interface or via an application programming interface (API). This allows users to access and interact with the digital object associated with the PID. The PID Metaresolver plays a crucial role in enhancing the interoperability and accessibility of objects by providing a unified interface for resolving PIDs from different systems. They simplify the process of locating and accessing research data, publications, and other scholarly outputs, thereby facilitating collaboration, citation, and reuse in the academic community.

Different target groups (scientists, data scientists, data managers, data stewards, but also software developers and, of course, the PID service providers) played an important role in the design and development of the PIDMR. By defining possible uses by these target groups, it was possible to determine the core technical requirements. These requirements range from simply resolving individual PIDs to solutions that can be exploited in batch (and automated) processes. It is evident that an intuitive interface and a scalable solution that can support resolving PIDs in batch are both essential. However, the biggest technical challenge is the actual integration of the different PID systems themselves: dealing with the ambiguity of individual PIDs in the technical solution, but also in communication with the enquirer, is a decisive success factor for the future use of the Metaresolver. The large number of PID systems supported, as well as the different response options available, is also a challenge in terms of implementation and communication.

The PID Metaresolver consists of three components: the backend, an application programming interface (API) and a user interface (UI). Each of these components fulfills different requirements. The UI offers the option of resolving individual PIDs, checking and validating unknown PIDs. In addition, the UI forms the interface for service operators to register their own PID service and store it with information. The API allows the PID Metaresolver to be integrated into various workflows. The API is addressed by the UI and stores the provider data. The PID Metaresolver backend performs the actual resolution step. The necessary logic is stored centrally and used by all components of the PID Metaresolver. The PID Metaresolver backend is based on the established handle system, which is known for the distributed and scalable resolvability of PIDs.

---

[5] https://github.com/FC4E-WP5

[6] van de Sanden, M., Ahmed, R.-B., Azzouz-Thuderoz, M., Bingert, S., Chatzopoulos, S., Bennet, M., Cannizzaro, G., Gruenpeter, M., Hugo, W., Kesäniemi, J., Lienhop, H., Nielsen, L. H., Medves, M., Monteil, A., Quazi, R., Ross, S., Snyder, K., Steinhoff, W., Suominen, T., … Wilk, R. (2023). D1.3 FAIRCORE4EOSC Components Beta Release Report. Zenodo. https://doi.org/10.5281/zenodo.10518813

A PID metaresolver offers significant benefits for researchers, organizations, and the scholarly community by promoting interoperability, streamlining access to digital objects, and enhancing the discoverability and persistence of research outputs and data. Unified Access is one of the advantages of the service. A PID metaresolver provides a single access point for resolving different types of PIDs. Users don't need to be familiar with the specific PID Provider services used by each Organization or community. By providing a centralized location for resolving PIDs, metaresolvers contribute to the discoverability of research outputs and data. Users can more easily find and access relevant resources, leading to increased visibility and impact for scholarly work. The only thing they have to do is to simply submit the PID to the metaresolver and obtain the associated information or digital object. It hides the technology complexity of the different PID Providers services. At the same it promotes interoperability between different organizations, communities. Researchers can easily access and reference digital objects across boundaries, facilitating collaboration and knowledge exchange. Facilitates Citation and Attribution: Metaresolvers enable accurate citation and attribution of digital objects by providing reliable access to metadata and associated resources. Researchers can confidently reference PIDs in their publications, knowing that readers can easily verify and access the cited materials.