



Trustworthy persistent identifier systems of the future

Pavel Golodoniuc (1), Jens Klump (2), and Nicholas Car (3)

(1) CSIRO, Mineral Resources, Perth, Australia (pavel.golodoniuc@csiro.au), (2) CSIRO, Mineral Resources, Perth, Australia (jens.klump@csiro.au), (3) Geoscience Australia, Canberra, Australia (nicholas.car@ga.gov.au)

Over the last two decades, persistent identifier (PID) systems have seen some significant changes in their governance policies, system capabilities, and technology. The development of most systems was driven by two main application areas, namely archives and libraries. Guidelines and criteria for trustworthy PID systems have been clearly devised (Bütikofer, 2009) and many PID system implementations for the identification of static digital objects have been built (e.g., PURL). However systems delivering persistent identifiers for dynamic datasets are not yet mature.

There has been a rapid proliferation of different PID systems caused by the specific technical or organisational requirements of various communities that could not be met by existing systems such as DOI, ISBN, and EAN. Many of these different systems were limited by their inability to provide native means of persistent identifier resolution. This has prompted a decoupling of PID-associated data from the resolution service and this is where the Handle system has played a significant role. The Handle allowed to build a distributed system of independently managed resolver services.

A trustworthy PID system must be designed to outlive the objects it provides persistent identifiers for, which may cease to exist or otherwise be deprecated, and the technology used to implement it, which will certainly need to change with time. We propose that such a system should rest on four pillars of agreements – (i) definitions, (ii) policies, (iii) services, and (iv) data services, to ensure longevity. While we believe all four pillars are equally important, we intentionally leave regulating aspects of issuing of identifiers and their registration out of the scope of this paper and focus on the agreements that have to be established between PID resolver services and the data sources indicated by the persistent identifiers.

We propose an approach to development of PID systems that combines the use of (a) the Handle system as a highly distributed system of independent nodes that provides registration and first-degree resolution facilities for persistent identifiers, and (b) the PID Service tool to enable fine-grained resolution of object representations in dynamic datasets using parameterized requests. The PID Service, deployed in close proximity to data services and managed by individual organisations, gives great flexibility and control over multiple representations and versions of information objects in data stores while allowing basic resolution via the Handle system.

Through the assessment proposals and implementation example we give, we highlight a critical aspect of PID system design and implementation that we believe is often neglected – the protocols and procedures required for PID system decommissioning. These protocols and procedures are needed in order for PID systems' core data to be able to be transferred to successor systems when current systems need replacing, as we indicate they inevitably will. Not knowing what successor systems may be, we strongly believe in using open standard formats as this gives future system implementers the best possible chance of being able to work with the data export. Smooth system handover will ensure that identifiers minted today will actually persist into the future.