

Distributed design approach in persistent identifiers systems

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

(1) CSIRO Minerals, Perth, WA, Australia, (2) Geoscience Australia, Canberra, ACT, Australia

The need to identify both digital and physical objects is ubiquitous in our society. Past and present persistent identifier (PID) systems, of which there is a great variety in terms of technical and social implementations, have evolved with the advent of the Internet, which has allowed for globally unique and globally resolvable identifiers. PID systems have catered for identifier uniqueness, integrity, persistence, and trustworthiness, regardless of the identifier's application domain, the scope of which has expanded significantly in the past two decades.

Since many PID systems have been largely conceived and developed by small communities, or even a single organisation, they have faced challenges in gaining widespread adoption and, most importantly, the ability to survive change of technology. This has left a legacy of identifiers that still exist and are being used but which have lost their resolution service. We believe that one of the causes of once successful PID systems fading is their reliance on a centralised technical infrastructure or a governing authority.

Golodoniuc et al. (2016) proposed an approach to the development of PID systems that combines the use of (a) the Handle system, as a distributed system for the registration and first-degree resolution of persistent identifiers, and (b) the PID Service (Golodoniuc et al., 2015), to enable fine-grained resolution to different information object representations. The proposed approach solved the problem of guaranteed first-degree resolution of identifiers, but left fine-grained resolution and information delivery under the control of a single authoritative source, posing risk to the long-term availability of information resources. Herein, we develop these approaches further and explore the potential of large-scale decentralisation at all levels: (i) persistent identifiers and information resources registration; (ii) identifier resolution; and (iii) data delivery. To achieve large-scale decentralisation, we propose using Distributed Hash Tables (DHT), Peer Exchange networks (PEX), Magnet Links, and peer-to-peer (P2P) file sharing networks – the technologies that enable applications such as BitTorrent (Wu et al., 2010).

The proposed approach introduces reliable information replication and caching mechanisms, eliminating the need for a central PID data store, and increases overall system fault tolerance due to the lack of a single point of failure. The proposed PID system's design aims to ensure trustworthiness of the system and incorporates important aspects of governance, such as the notion of the authoritative source, data integrity, caching, and data replication control.