Report on the Knowledge Exchange seminar on Persistent Object Identifiers

Executive summary
On 14 and 15 June 2011, more than 40 experts involved in various Persistent Object Identifier (POID) communities met for a Knowledge Exchange seminar to discuss the challenges and opportunities involved in interoperability between multiple PID-systems. Three major systems – Handle, URN:NBN and DOI – presented their current state of affairs and examples of their systems in practice. A quick survey of the audience showed a roughly equal distribution of representatives from the three communities present. Bringing together these different communities into one room for two days in itself was a first, and an important step in exploring interoperability – not just between systems and machines but also between people, organisations and policies. Views expressed on what is needed covered a wide spectrum. Some advocated providing more clarity about what each system does and how they compare, rather than how they differ. Others stressed awareness raising about the importance of the use of POIDs as such, since this is not as widely acknowledged as is sometimes assumed. Many felt that interoperability requires effort to harmonise systems’ policies. Some, on the other hand, had difficulties coming up with a compelling case for an interoperability effort at system level.

There was a clear interest in connecting the POID systems to the linked data standards. This led to the ‘Den Haag Manifesto’, which outlines a series of concrete actions to join the POID and Linked Open Data communities. In the summing up of the discussions the suggested actions were deemed “achievable, with a reasonable chance of something happening”; and it would be “powerful if in some twelve months from now we’d have machine interoperability in some way, even if not necessarily communities interoperating...”

The seminar offered ample input for the DIGIODUNA study which is being conducted by the European Commission and should offer recommendations in this field which are relevant for scientific data infrastructures.

Introduction
The seminar on Persistent Object Identifiers took place at the Data Archiving and Networking Services (DANS)\(^1\) offices in The Hague, the Netherlands, from lunchtime on June 14 until mid-afternoon on June 15. The Knowledge Exchange Working Group on Interoperability of Digital Repositories organised this seminar as a closing event to their work on POIDs. They wished to provide a platform for the communities of different POID-systems to explore and discuss the present state of affairs of these systems, their development and the challenges they each face. Their second aim was to have these experts assess the possibilities and questions involved in cooperation to ensure greater interoperability between the various systems. Three POID systems: Handle, DOI and URN:NBN, were invited to inform the audience on the current state of affairs and development plans for their systems. Following this three projects presented their experiences with the practical application of each of the systems. To put these system oriented talks into context, a few experts discussed POID issues from a more general perspective. The overall aim of the workshop was to ask all these experts to identify a roadmap for further work on PID systems and interoperability, more specifically at policy level.

In his welcoming words Peter Doorn, director of hosting organisation DANS, touched on the importance of this event, and on what it would collect from its participants. He quoted Thomas Carlyle: “Permanence, perseverance and persistence in spite of all obstacles, discouragement, and impossibilities: It is this, that in all things distinguishes the strong soul from the weak”.

---

\(^1\) www.dans.knaw.nl
From Peter’s perspective sustained access requires increased and persistent cooperation, and for this we need strong souls....

State of affairs from Handle, DOI and URN:NBN

Juha Hakala (National Library of Finland) shared his thoughts on issues to be solved in a world of Interacting Persistent Identifier Systems. “There is benefit in cooperation between different POID systems, and nothing to lose” in Juha’s view. One common issue for most systems is that they were developed quite a long time ago. This shows they are persistent, but they have not modernized their syntax. Other common issues are the ‘cool URI’s’, and the unclear relation between POIDs and traditional bibliographic identifiers. Juha advocated joining forces to find common approaches and solutions to these issues, and said that all may benefit from some level of IETF support to raise the level of standardisation of the POID systems. He will also start work on a ‘registry of services’, to enable flexible specification of services, making these more widely applicable than just URN. “The more versatile POID systems and services are, the clearer the difference with cool URI-based approaches” was Juha’s parting shot.

Laurents Sesink (DANS) talked about the outcomes of the PersiD project, specifically about the importance to not just focus on technology – as is so often the case -, but to align policy, communication and technology. The project aims to develop a trusted POID infrastructure, with an international meta-resolution service, governed by a transparent policy document. This will refer to national and local policies; this documented policy is to grow out of a code of conduct that will be developed first. Implementation of the shared policy will be built on trust rather than certification.

Clifford Lynch (CNI) welcomed the renewed focus on questions about POIDs, as there is less experience with serious interoperability than he would like to see, especially at the policy level. He referred to the Blue Ribbon Taskforce, an initiative by US funding agencies and JISC, who have addressed the understanding of sustained preservation of digital objects at a fundamental level. In their report they suggest to move away from ‘preservation into perpetuity’, and opt for preservation for a limited period of time (20-30 years) and then review whether to continue preservation or follow another strategy. This calls for transparent, orderly transitioning mechanisms and policy.

Other points high on his agenda for further work:
- IDs for other entities than objects, especially author IDs
- Preservation of results from more volatile new practices like virtual research organisations
- Criteria for trustworthiness, relating to the object and to the organisation assuming archival responsibility

http://brtf.sdsc.edu/
Objects and authors
The presentation by Magchiel Bijsterbosch (SURF) on the Dutch Digital Author Identifier raised a number of issues specific to this type of ID. How to realise disambiguation – more important to DAI than persistency – in the international context, was one important question discussed, with national privacy rules and trust issues being important constraining factors.

Maurizio Lunghi (FDR) showed how the PersID project builds a trusted architecture by working at the different levels of content, resolution, meta-resolver and extension to other (trusted) resolvers like Handle, DOI and ARK. An important common element here is that the work at the different levels is done by trusted organisations. That this is a requirement is shown by the preliminary results from a questionnaire in the APARSEN project. This network of excellence aims to synergize ongoing research, e.g. on identifiers and citability. The recent questionnaire on the use of and requirements of PID for objects, persons and organisations also shows that many respondents (mostly libraries) still use internal PID’s for objects, none for authors, and organisational IDs are not considered a key issue.

Persistence (de)constructed
Talking about persistence and identifiers, Andrew Treloar (ANDS) ‘decomposed’ the term POID and identified five levels at which persistence can pose problems: the object, the identifier, binding between the two, the service to resolve from id to object and the service to allow the updating of the binding between id and object. One way to approach the problems is to regard them in terms of architecture and brittleness, and to think about the role of PIDs as an indirection layer that reduces the level of brittleness in getting from ids to objects. The technology itself is easy, using and maintaining the technology is hard. Success in his view requires a combination of sustainability model (business, shared community activity or other legal framework), policies, procedures to implement the policies, and the will or intention to follow those procedures – the latter being a ‘conditio sine qua non’.

Highlights from usage experiences with Handle, DOI and URN:NBN
“Based on my experiences so far I am sure the Handle system will be flexible enough to deal with resolving the URN:NBN namespace system as well. It’s a very generic resolver system.”
Pieter van Beek, SARA (NL)

DataCite uses a DOI to link data sets to articles for example published in ScienceDirect; a reader can go from the article citation to the data without having to access – and pay for – the article itself.
Plans for further development include uploading metadata into the Web of Science and other indexes, cooperation with CrossRef to find article-data links and cooperation with CNRI on template DOI and parameter-based resolution.
Jan Brase, DataCite

The German National Library is using URN:NBN. Future goals are promoting awareness of POIDs and publishing new policy which will offer more transparency for the user. They plan to launch www.culturegraph.org which will focus on offering access to descriptions of entities and interlinking with linked data.
– Jurgen Kett, DNB (Germany)

Closing thoughts
- Technology = Easy
- Technology((Use|Maintain)) != Easy
- Success here requires a combination of:
  - Will
  - Procedures
  - Policies
  - Sustainability model

Group discussions
The participants split into three break out groups to investigate main challenges for POID systems, main challenges and benefits of running multiple POID systems and further steps in the field of Linked Data & Persistent Identifiers. Through a process of identifying challenges and setting the agenda the groups identified where cooperation could be beneficial in the future. These groups came up with the following recommendations:

- Go for wider acceptance and use of POID systems in the context of long-term availability
- Sketch a landscape of use cases and approaches in order to raise awareness
- Keep talking to each other, in smaller groups, to bring the discussion and mutual understanding forward
- Experiment with the use case of a repository manager developing a new system – how to build an interface dealing with the multiple systems
- Regarding Linked Data: avoid ‘us and them’ talk – look at it as different classes of technology
- In order to connect with the linked data community take up the approach outlined in the ‘Den Haag Manifesto’

Future
With this seminar Knowledge Exchange achieved its objectives to raise awareness for PID issues and to bring experts together to work on a common agenda. At the close of the seminar, chairman Bas Cordewener (Knowledge Exchange/SURF) thanked all participants for their open and constructive contributions. It is now up to the stakeholders, which were all present at the seminar, to continue cooperative work towards a trusted and transparent PID service infrastructure.

---