PID services in the SSH*

About persistent identifiers

By inspecting a document held by a library, one can determine if it has been treated in a tidy way. This means that one can effectively establish the authenticity and version of a resource.

When it comes to electronic documents, the version of a document is much harder to establish. Moreover, its location, a URL, is not bound to be fixed over a longer period of time. Because of this, there is a call for an authority that provides for identification of electronic documents. This is exactly what services involving persistent identifiers do: a PID provides a sustainable reference to an electronic document.

Services using PIDs enable scientists to refer to or to retrieve documents, regardless of their location. Because of the importance of sustained reference to electronic documents, DASISH recommends the use of PIDs and PID services. When studying these, DASISH has identified three systems in use in the SSH.

DataCite

DataCite is a not-for-profit organization with members from several data centers around the world. DataCite provides a registry with persistent identifiers based on Digital Object Identifiers (DOI) for datasets. The underlying technology is the Handle System (HS) as is also used by EPIC. DataCite also manages basic metadata related to the datasets.

In the SSH, the CESSDA archives from GESIS, UKDA and SND are DataCite clients. In the future, SHARE-ERIC will join. By implementing DataCite services, client organisations accept both a responsibility to the integrity of the DOIs as well as to the objects these identifiers refer to.

A different ambition: EPIC

The European Persistent Identifier Consortium (EPIC) also facilitates PID services. It aims at long term registration, storing and resolving of identifiers. While DataCite services apply to citable data collections, EPIC intends to deal with all types of digital objects or fragments thereof.

An organisation offering EPIC services declares itself willing to provide appropriate and sustainable services to communities. Such services should extend and build on the current agreed EPIC service profile. The development of EPIC services is driven by GWDG, SURFsara, CSC, and DKRZ.

URN:NBN based identifiers

National libraries developed PID services with identifiers following the IETF Uniform Resource Names standard. Since these identifiers comprise the independently defined National Bibliography Numbers, they cannot refer to objects beyond the national scope.

To overcome the problem, the PersID project was launched. Its goal is to create services that resolve different types of identifiers. The first services were created by the German National Library. Within PersID, a legal body ensuring maintenance and operation of services will be set up in the future.

Which system is suitable for you?

In the absence of broader, community defined, set of requirements for PID services within the SSH, DASISH made a survey to expose these. Since scientists working in different disciplines may have different requirements, various types of communities were questioned and gave input. The result is the following list of requirements that were compared with the three PID services.

	DataCite	EPIC	URN:NBN
Basic requirements			
The PID services provided should be interoperable.	Yes	Yes	Yes
The PID service provider should also offer training.	No	Maybe	Yes
Common requirements			
The PID resolution service needs to be available under the responsibility of a reliable and long-term funded organisation. This service should be provided on a European or national scale.	Yes	Yes	Yes
PID services need to be part of a European or national network.	Yes	Yes	Yes
A PID service provider should have a clear policy describing the responsibilities of the stakeholders.	Yes	Unknown	Yes
A PID service should require and offer a minimal set of metadata.	Yes	No	No
The PID services should be reliable, fast, and scalable.	Yes	Yes	Yes
PID services should be available for centers and users in the European Union.	Yes	Yes	No
The business model for the provider should be sustainable for all stakeholders. Also, the model should be controlled by the scientific communities.	Yes	Yes	Yes
Extended requirements			
PID syntax should allow for version and fragment denotation.	Yes	Yes	No
The PID services should be able to create a view on efforts to link literature, data and authors.	Yes	Unknown	No
PID services should support different metadata formats.	Yes	Yes	Yes
PID service providers should make instances of services available for testing purposes.	Yes	Yes	Yes
Objects referred to by persistent identifiers can also be authors and organisations.	No	No	No

This comparison can serve as an aid in deciding which type of PID service will suit you best.