

The ETH Zurich DOI Desk: DOIs as standard persistent identifier and best practice in Swiss Open Access repositories

Barbara Hirschmann, ETH-Bibliothek, ETH Zurich

Over the last couple of years it has become best practice in Swiss Open Access repositories to assign Digital Object Identifiers (DOIs) as persistent identifiers to digital objects such as dissertations and theses, preprints and postprints, research data or digitized library material.

The ETH Zurich DOI Desk, the main DOI registration service for the Swiss university and research sector, has registered – in cooperation with its clients - over 700.000 DOIs since 2009 and serves 20 customers running 27 different services (such as repositories) all over Switzerland.

How did it come about that the DOI registration service has been set up with such success? Why has the DOI become a de facto standard in Swiss Open Access repositories while in other countries there is much more diversity in the persistent identifiers in use?

The DOI Desk has been set up at ETH Zurich between 2008 and 2012 as part of the national innovation and collaboration project “e-lib.ch: Swiss Electronic Library”. It is now run as a regular service by the ETH-Bibliothek, the main library of ETH Zurich, with technical support from the ETH Zurich IT Services.

The DOI Desk acts as a member of DataCite – an international not-for-profit organisation formed in 2009 that aims at promoting citability and easier access to research data on the internet. DataCite is a member of the International DOI Foundation (IDF) which is the governance and management body for the federation of registration agencies providing DOI services. As one of these official registration agencies DataCite has built up services promoting the use of DOIs for research data so that these can be found, accessed and reused more easily. Research data in the context of DataCite can be almost any research output in digital form that might be a foundation for further scientific investigation, including primary data as well as secondary data such as data papers, grey literature or theses. DataCite acts through its member organizations such as ETH Zurich, the California Digital Library, the British Library, or the German Leibniz Institute for Social Sciences (GESIS).

These institutions have in turn built up their own additional services that are geared towards the specific requirements of their clients.

Since the ETH Zurich DOI Desk has been envisaged as a service for the whole Swiss university and research sector we tried to build an infrastructure that can be used by a diverse range of publishing services and relieves our clients from complex technical prerequisites that especially smaller institutions might not be able to meet.

This is why ETH Zurich has setup its own DOI infrastructure which acts as an intermediary between its clients on the one hand and the global DOI infrastructure governed by the IDF as well as the technical services run by DataCite on the other hand.

This intermediary technical infrastructure consists of a Handle Server that registers DOIs and exports them to the global DOI System, which is in charge of DOI resolution, as well as a DOI Management web application that can harvest DOIs and their associated metadata from our clients and feed them to the Handle Server as well as to DataCite's metadata service.

The harvesting mechanism used by the DOI Desk is based on the OAI-PMH protocol for metadata harvesting. Since an OAI interface is implemented as a standard feature by most repositories, especially Open Access repositories for textual resources, most repository operators in Switzerland could very easily meet this requirement for DOI registration. However, we also have the possibility of harvesting DOIs and their associated metadata from an ATOM feed if the repository operator cannot provide an OAI server. For clients who will only register a very small amount of DOIs per year we provide a manual workflow where DOIs can be entered and updated directly via the DOI Management web application.

When setting up an agreement for DOI registration the repository operator and the DOI Desk agree on a syntax used for the DOI names of the service. The repository operator then assigns a unique DOI name to each existing and forthcoming record and publishes them as part of the metadata record via its OAI interface or via ATOM feed together with the current URL to which the DOI name must resolve.

The DOI application regularly harvests new and updated DOIs from the interfaces of its clients and stores them in our local infrastructure together with the associated URLs and metadata. This information is then exported to the Handle Server for resolving purposes and to DataCite for metadata indexing. DataCite has issued its own metadata schema (schema.datacite.org) which describes certain mandatory, recommended and optional elements that should be provided when registering a DOI. However, since this metadata standard has not been adopted widely by repositories yet, the ETH Zurich DOI Desk has implemented a mapping mechanism which allows repository operators to publish DOIs and associated metadata in Dublin Core Simple format. This metadata is stored locally in the DOI Management Web Application and consequently transformed to DataCite's metadata schema when exported to the DataCite Metadata Store.

From the DataCite Metadata Store the DOIs and metadata of our clients can be reused by third parties for building discovery or other services.

A successful DOI service however does not only depend on technical infrastructures. To guarantee the stability and citability of as well as long term access to a digital object referenced via a DataCite DOI we have to build strong relationships between publishing agents, allocation and registration agencies. Based on DataCite's Business Models

Principles the ETH Zurich DOI Desk has developed a “DOI Policy” which defines the type of institutions that are eligible for our service, and describes the requirements for the digital objects, metadata and storage system, that each DOI client has to fulfil.

If during our conversations with a potential new client we come to the conclusion that they can meet these requirements, we will set up an individual agreement with them to define the reciprocal rights and obligations of each partner in the DOI registration process.

As part of our service we also try to keep our clients up to date with the latest developments from DataCite and the International DOI Foundation, share with them interesting initiatives from the data citation community as well as best practices for data citation. We are an active member in the DataCite Policy and Best Practices Working Group and our newest service will be a regular newsletter that we will send out to clients and other interested parties on a regular basis.

In line with the vision of DataCite which aims at improving visibility, accessibility and reuse of research data, at ETH Zurich we have built a DOI service that can be used by a variety of publishing agents for a variety of research outputs. We believe that scientific content published in university repositories – no matter if it is primary data or secondary data such as theses or grey literature – are an important part of a researcher’s scientific record. By offering a service such as DOI registration repository operators can improve not only their technical infrastructure (by using a global best practice standard for persistent identification) but also better market their service to researchers – since these are very often familiar with DOIs and strongly appreciate it as a value-added service to receive one when depositing an object.