Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

ETH Library
Research Support

DOI3: A new software environment for the registration of DOIs

Final Project Assessment

Project Lead	Samantha Foulger (ETH Library) / Davor Kupresak (ETH IT Services)		
Created	21.07.2022		
Status	[] in progress	[] in review	[✓] released

Table of Contents

1. Purpose of this Document

The final project assessment forms the basis for the decision to close the project. It informs the client about the target-actual difference with regard to the objective and financial project and procedure goals, as well as the deadlines. The project experiences are documented in summary form. The content and dates of the project performance review are determined.

2. Management Summary

The DOI3 project was carried out between June 2022 and August 2023. The aim was to relaunch the DOI system of ETH Zurich, which offers an alternative to the usual DOI registration. The system is available to existing and potential DOI-Desk customers, the link to the code is in this final report should anyone be interested in it. Despite project delays, the project objectives were fully achieved: The system was reprogrammed, the data successfully migrated and transferred into operation. The future of the new system is secured in both personnel and finances ways, since the DOI product existed before the project. The project goals achieved are listed under point 4.
Point 5 describes the advantages of the new system for DOI customers, the Swiss Scientific Community and the DOI Desk. These are primarily the continuation of an existing service, i.e. no change for existing customers, and a simplified registration via the OAI-PMH interface for future customers. The milestones, costs and effort are listed under point 6. The delay of about 5 months is the result of technical difficulties before and during the migration and staff absences. Nevertheless, not as many personnel ressources as initially calculated were necessary. Project experiences include more generous project planning with more time and a larger project team. After the end of the project, it is necessary to retain the data from the old system, as it contains certain data from the DOI-Desk's own old handle system that was not migrated.

3. Initial Situation

The DOI registration service via ETH Zurich's DOI application is widely known on a national level. Due to the fact that the ETH Library was a founding institution of DataCite, ETH Zurich's DOI-Desk has always had its own system for harvesting records for DOI registration. However, the former DOI system, in use since 2012 was reaching its end of life and needed to be reprogrammed.
In comparison to the DOI registration directly via a registration agency like DataCite or Crossref, the data needs to be available in the metadataformat Dublin Core and as xml on an OAI-PMH interface, all of which are widely used, international standards. From this interface, the DOI system harvestest the data, converts it into the DataCite medtadataformat (DataCite MetadataSchema) and Json, then sends the data to DataCite for registration.
The service of harvesting DOIs and their metadata (DOI Desk) versus sending the data to a registration agency (DataCite) is popular among our clients and simplifies the monitoring of errors. The DOI-Desk has communicated in the past that the current system needs to be renewed but can be used by all customers as before with no extra costs. This ensures that customers of ETH Zurich's DOI Desk do not need to switch to another way of registering DOIs, which saves them both time and money.

4. Assessment of Objectives achieved

4.1. Overview of Objectives

Nr	Category	Description	Measurement	Priority	Degree of achievement
1	Replacement of current DOI system	Make sure that Swiss universities, research institutions, and libraries can continue to participate in DOI registration via DataCite once ETH Zurich's DOI registration application has reached its end-of-life.	Continued use of DOI system after end-of-life of current system	1	100%
2	Continued offer of ETH DOI registration via OAI-PMH protocol	Develop and provide to Swiss universities, research institutions, and libraries a state-of-the-art software tool that enables users to register DOIs via an established technical workflow based on the OAI-PMH protocol.	OAI-PMH protocol can be used with the new system	1	100%
3a	Use new interface	Use the DataCite REST API for the new system instead of DataCite MDS API, to ensure the long term use of the API.	Export to DataCite via DataCite REST API	2	100%
3b	Use new DataCite MetadataSchema	Use the latest release of the DataCite Metadata Schema (4.4) for the new system (mandatory fields only).	RessourceTypeGeneral list can be used in full.	3	100%

Nr	Category	Description	Measurement	Priority	Degree of achievement
4	Migration from old to new DOI system.	Migrate existing customers of ETH Zurich's DOI Desk from the current to the new software application for DOI registration.	All of the DOI pools can register DOIs in the new system.	1	100%

All objectives could be achieved.

5. Benefits for Customers, the Swiss Scientific Community and the DOI Desk

5.1. Benefits for the Customers of the DOI-Desk

For current customers of the DOI Desk, the reprogramming of the DOI management system primarily means the continued use of an existing service, described in chapter 3. For the future customers, it offers and alternative and often easier way to register DOIs.

The user interface and dashboard are now simpler, more modern and easy to understand. The latest DataCite MetadataSchema (version 4.4.) is used for mapping. The DOI management system is more performant and now uses the DataCite REST API, the only interface recommended by DataCite for the future. As DataCite still does not support parallel import, a maximum request rate has been built into the new DOI system so that the DataCite system is not overloaded. Records that have be sent to DataCite but have to wait due to the request rate are in the status "Export". This shows the registration status of the data.
Error monitoring has been implemented based on the old system, but has been partially simplified. A list of the most common errors is provided in the user manual.
The user manual was supplemented and adapted for the customers of the DOI Desk.

5.2. Benefits for the Swiss Scientific Community and the PID Community

The continuation of the DOI management systems ensures an an alternative, easy way of registering DOIs currently not offered by DataCite (or any other registration agency) for the future. Current and prospective customers can make use of this system and ETH Zurich's expertise. The DOI management system is a long term solution that will continue after the project's end in the product "DOI-Desk" which has been a service by ETH Zurich for over 10 years. It is a stable service and there is no dependency on further funding, the costs for technical maintenance is included in the existing business model of the product. The DOI management system's future role and availability are guaranteed.

Should another instutition be interested in setting up a similar system, the code is open and available at https://gitlab.ethz.ch/doi.
The information about the newly set up DOI system to customers and the community was done by distributing the information via the mailinglist swisslib, widely read by information professionals in Switzerland.

5.3. Benefits for the DOI-Desk

For the DOI Desk, the same as for the customers, the new DOI system means the continuation of the previous service. While the registration via DataCite is technically out of our hands, the registration via our own application is traceable in almost every step, which makes the monitoring of errors easier to understand. At the same time, the new system brings a significant improvement, as it is not only much more per-formant, but the dashboard easier to use and jobs that are being executed are visible.
Internally, ETH is (re)defining and/or consolidating roles with a new SLA between IT Services and the ETH Library.

6. Target-Actual Comparison

6.1. Milestones

Milestones	Date planned	Date executed
Project Kick-Off	*June 1, 2021*	*June 1, 2021*
Requirements specifications ready	*August 31, 2021*	*August 31, 2021*
Software development, installation and testing completed	*January 1, 2022*	*March, 15, 2022*
Launch of new system completed	*February 27, 2022*	*June 16, 2022*
Final evaluation	*March 21, 2022*	*August 31, 2022*

Reasons for the delay of completed launch and the final evaluation:
On the technical side, there were delays in implementing the web design according to the CD and accessibility guidelines.
It was decied not to do the migration all at once as originally planned, but iteratively in order to be able to intervene more quickly in case of errors. This took some additional time. However, it was a sensible measure, as a major problem arose at the beginning of the migration that could only be solved together with DataCite and which had not shown up in the exports of records to DataCite's test system. Also, with very large repositories on the application side, it was not possible to retrieve all data at once. This only became apparent during the testing.
Due to a long-term staff shortage at the DOI desk, less capacity was available for the project than planned during the migration period and onwards, plus other areas of work were taken over by the library project lead.

6.2. Costs (development by external programmer and webdesigner)

Year	Planned (in CHF)	Actual (in CHF)	Discrepancy (in CHF)
2021	*120'000*	*74'336*	*+45'6640*
2022	0	*41'595*	-41'595
Total	*120'000*	*115930*	

6.3. Effort (for personel ETH IT Services and ETH Library)

Personell IT Services

Year	Planned (per man-day)	Actual (per man-day)	Discrepancy (per man-day)
2021	9.3	2.2	+7.1
2022	5.7	2.2	+3.5
Total	15	4.4	+10.6

Personell Library

Year	Planned (per man-day)	Actual (per man-day)	Discrepancy (per man-day)
2021	32	5.4	+26.6
2022	7.5	18.8	-11.3
Total	39.5	24.2	+15.3

7. Project experience

Nr	Area	Experience	Relevance	Recommendation
1	Schedule	The schedule was very tight and did not consider possible difficulties or temporary loss of personell enough.	Relevant for future project planning.	Plan with more time.

Nr	Area	Experience	Relevance	Recommen-dation
2	Size of Pro-jectteam	The small size of the project team seemed accurate for a small project (in terms of time). However, this was a big disadvantage in the event of loss of team members.	Relevant for future project planning.	Plan with more people (backup).
3	Goals	The goals were very hands on and clear.		
4	Financial plan-ning	The financial planing in itself was straightfor-ward, but because of the delay and the turn of the year, neither the planned figures from 2021 nor 2022 were accurate.	Important for budegeting and/or funding.	Consider that even minor delays might have a big impact on the planned fi-nances if the year changes.
5	Long-term backup of data from old sys-tem	A good solution with different steps over several years could be found.		

8. Pending Issues and Measures

8.1. Open issues from the project

There are no open issues.

8.2. Further measures after the end of the project

Nr	Measure	Responsible	Date
1	DOI2 application remains in operation for 2-3 months after DOI3 goes live (without cron-jobs.)	IT Services	September 2022
2	Shutting down the DOI2 application i.e. the web instance. The database will remain in operation for about 2 years.	IT Services	Autumn 2024

Nr	Measure	Responsible	Date
3	Before the database is deleted, the individual tables will be exported and made available. IT services consult with the ETH Library.	IT Services	Autumn 2024

The above steps are necessary, as certain data from the DOI-Desk's own previous handle server were not migrated into the new system. The ETH Zurich-owned handle server was replaced by the registration via DataCite in 2017. Therefore, even if this data is available in the Handle System, it is currently not directly accessible for ETH Zurich via services like DataCite Fabrica. It is only very few data from the very beginning, when DataCite had no own infrastructure, but to ensure that no data at all is lost and that it remains directly accessible to ETH Zurich, these three steps were considered indispensable.