

Purchasing an Institutional Repository alongside a Current Research Information System: Necessary or unnecessary?

Author: Frank Diepmaat (Tilburg University, The Netherlands)

Background

Until 2013, Tilburg University had two separate applications for capturing metadata of research output and other scholarly activities of its researchers (CRIS¹: Metis) and for storing full text of research output (IR²: Arno). Both applications had reached end-of-life and needed to be replaced.

At Tilburg University it was felt that it would be more efficient to combine the functionalities of a CRIS and an IR in one system, rather than continuing to use two separate systems that had quite a lot of overlap in content. At the time, we mainly had these reasons for switching to a single application:

- Pure, Elsevier's CRIS system, offered facilities for storing full text of research output and a portal through which full text could be downloaded by a wide audience
- Replacing two applications with one, would reduce the need for functional and technical administration
- With one application, there is less risk of outdated or inaccurate data
- Elsevier's Pure allows individual researchers to upload their own files

De Castro already noted back in 2014 that although IRs and CRISs have somewhat different missions, there are some overlapping areas in the tasks they perform and there is a gradual convergence of these two types of systems. He talks about many institutions that have fully integrated their IR and CRIS, either by enabling systematic metadata transfer between the two applications or by having either application take over the functions of the other.³

Other Dutch universities also decided to combine functionalities of a CRIS and IR in one system and made the decision to implement Pure to achieve this goal. In his excellent lecture at the Pure conference in 2021, Guicherit of Tu Delft showed the result of a survey among the members of the Pure Dutch user group. It showed that 2/3 of the institutes no longer set up a separate IR system and that the majority now use Pure as an IR. The reasons for abandoning a separate IR were similar to those given above by Tilburg University.⁴

Plan S, OpenDOAR and its implications

In 2018, Plan S was launched by cOAlition S, a group of national funding organizations supported by the European Commission and the European Research Council. The main goal of Plan S was to have all scholarly publications funded by public or private grants from (inter)national research councils and funding agencies published openly through Open Access Journals, Open Access Platforms or Open Access Repositories without embargoes, starting in 2021.⁵

¹ CRIS= Current Research Information System

² IR = Institutional Repository

³ de Castro, P., Shearer, K., Summann, F. (2014). The gradual merging of repository and CRIS solutions to meet institutional research information requirements. *Procedia Computer Science* 33, 39-46

⁴ Guicherit, A. (2021, November 21). Can Pure also be an Institutional Repository?

<https://www.youtube.com/watch?v=rnHQZdahBVM>

⁵ cOAlition S., (2018, September 4). Why Plan S? <https://www.coalition-s.org/why-plan-s/>

Plan S includes the section "Part III: Technical Guidance and Requirements", which contains mandatory and highly recommended criteria for Open Access Repositories. One of the mandatory technical conditions for scholarly publications is the use of persistent identifiers with versioning, such as DOI (preferable), URN or Handle.⁶ Although Elsevier's Pure offers the possibility of registering an already existing DOI, it is not possible to have a DOI automatically created via the API at a DOI registration agency like DataCite or Crossref.

Plan S also states that the IR must be registered in OpenDOAR⁷, which implicitly indicates that the IR must meet the technical requirements of OpenDOAR. The preservation policy recommendations for an IR state, among other things, that:⁸

- items will be retained indefinitely
- the original bit stream is retained for all items, in addition to any upgraded formats
- Items may not normally be removed from the repository
- URLs will continue to point to 'tombstone' citations, to avoid broken links and to retain item histories.

Tilburg University unfortunately concluded that none of these important functionalities are currently covered by Pure. Although Dutch universities have repeatedly called for these improvements in Pure, so far this wish has not been heeded.⁹

Approach Tilburg University

Tilburg University ultimately decided not to wait for possible changes in Pure and to once again choose a separate application as its IR. In its strategy paper for 2027 the Executive Board established that 'In 2024, Tilburg University has a Plan S compliant and open source institutional repository outfitted for the archiving and publication of all types of written research output, connected to relevant external sources to increase the visibility and impact of Tilburg University's research.'¹⁰ Pure continues to serve as the CRIS system for our university.

Tilburg University had two guiding principles when starting market research for a new IR:

- The new IR should be based on open source rather than proprietary software. The reasons for this are twofold: reliable and well-functioning open source IR applications are already available, and this way the university can address researchers' concerns about the growing monopoly of research systems owned by only a few commercial institutions.¹¹
- The new IR should be hosted by an external partner with in-depth knowledge of the chosen application. Reasons for this are the cost and lack of resources at Tilburg University to maintain the application and develop the source code.

⁶ cOAlition S., (2018, September 4). Why Plan S? <https://www.coalition-s.org/why-plan-s/>, Requirements for Publication Venues

⁷ OpenDOAR= Directory of Open Access Repositories

⁸ OpenDOAR (2017), Open access repository policy (minimum recommended): Preservation Policy. <https://v2.sherpa.ac.uk/opensdoar/policytool>

⁹ Guicherit, A., (2021). Can Pure also be an Institutional Repository? <https://www.youtube.com/watch?v=rnHQZdahBVM>

¹⁰ Snijders, P., Meijdam, L. (2022), Uitvoeringnotitie Research Infrastructure 2027. *Uitvoeringsnotitie Strategie 2027*,6

¹¹ Schaafsma, J., van der Meer, M.(2021). 'Pure' misery: how Open Access makes us more dependent on big publishers, Universonline 2021-11. <https://universonline.nl/nieuws/2021/11/15/pure-misery-how-open-access-makes-us-more-dependent-on-big-publishers/>

Market research results

A project was launched with the goal of finding a suitable IR for Tilburg University. First, a market survey was conducted among already existing open source IR applications. Of the 17 applications found, four were selected based on the following criteria. The application must:

- be used by multiple institutions (at least more than one)
- support basic functionalities of an IR as specified in Plan S and OpenDOAR
- have regular version updates
- have an active user community
- be maintained by reliable companies that can host and further develop the IR system

Based on these criteria, the selection was narrowed down to four IR applications. At Tilburg University, we contacted national and international universities that were already using one of the applications and held interviews with them. This resulted in three IR applications on our shortlist: DSpace, EPrints and InvenioRDM. These systems were tested against the technical requirements and recommendations of Plan S and OpenDOAR. Then the systems were tested against Tilburg University's additional requirements, and DSpace came out on top. All this led to a recommendation to the steering committee of the IR project to choose DSpace as our future IR application and to launch a tender procedure for hosting, management and further development of the product by an external party.

The project's steering committee adopted this recommendation after which preparations began for a European tender with the goal of finding a hosting partner for DSpace. At the time of writing this summary, the tender period was still open until mid-April 2024.

Discussion

Since there is quite a bit of overlap between data captured in a CRIS and in an IR, it may have seemed logical in the past to choose one system for both information needs. With the advent of Pure, many universities chose to no longer set up a separate IR system and to use Pure as both a CRIS and a Repository. The question is (in retrospect) whether this was a wise choice, perhaps motivated more by considerations of cost savings than sustainability and diligence. The requirements established by Plan S and OpenDOAR make it clear that we are on thin ice when it comes to the required functionality for sustainable and complete storage of the full text of scholarly articles by our researchers.

Should we not accept that bringing together a CRIS and IR in one system is an impossible task and that we would be better off opting for two separate systems that can communicate with each other based on open standards?