

Improving the visibility of the institution, researchers and publications by introducing specific identifiers (PIDs)

Luc Hourlay

Belgian Health Care Knowledge Centre (KCE), Brussels, Belgium

Abstract

In the course of 2021, the Belgian Health Care Knowledge Centre (KCE) has decided to give further thought to improving the visibility of the KCE publications. This has led it to develop a project to set up three types of persistent identifiers (PIDs): one for the institution, another for researchers and the third for the publications themselves. The purpose of this text is to retrace the various stages in the project's implementation and to share the initial findings.

Key words: personally identifiable information; reference standards; publishing; persistent identifier.

Introduction

The Belgian Health Care Knowledge Center (KCE) is a type B parastatal funded by the Belgian federal authorities. Its mission is to provide scientific advice on subjects relating to healthcare and it is not involved in the ensuing political choices.

It works in five areas of expertise: the organization and financing of healthcare in the broad sense (HSR), the evaluation of medical technologies (HTA), the production of clinical practice guidelines (GCP), the production of methodological manuals aimed at establishing valid working methods (Methods), and the coordination of the Belgian non-commercial clinical research program (KCE Trials).

The KCE produces publications in the form of reports, summaries, supplements, COVID contributions, etc. Reports and associated documents must be legally distributed within 30 working days of approval by the Board of Directors. The library service, in collaboration with the researchers and the communications service, makes these documents available via the institution's website, the library catalogue, the institutional repository and the legal deposit of the Royal Library of Belgium.

Objectives

The objectives of this project are multiple and are intended to respond to a series of needs and/or problems

that are sometimes less obvious. We aim to facilitate the identification of the institution as a research organization, improve the identification of authors and their scientific output, facilitate the dissemination of documents produced as part of studies carried out within the KCE and improve the management of access to documents over time.

Project

The library has developed a project structured around three types of persistent identifier (PID):

- setting up a specific PID for the institution (to be implemented in 2021);
- the introduction of PIDs for authors of the institution's publications (to be introduced from 2022);
- the implementation of PIDs for the institution's publications (to be introduced during 2022).

It was decided to work in three phases to spread the workload across the many other tasks already carried out by the department.

In practice, these phases are part of a cross-functional dynamic that is not limited to the library, but also covers human resources, communication, layout, the research program and knowledge management.

Setting up

Phase 1: the persistent identifier for the institution

The persistent identifier (PID) (1) for a scientific insti-

Address for correspondence: Luc Hourlay, Belgian Health Care Knowledge Centre (KCE), Administrative Centre Botanique, Doorbuilding (10th floor), Boulevard du Jardin Botanique 55, B-1000 Brussels, Belgium. E-mail: luc.hourlay@kce.fgov.be

tution has become a key element of identification and recognition in recent years. More than 17 organizations offer this type of identifier, whether commercial, non-commercial, national, or international.

It allows the unique identification of the institution's name and, by extension, the association of its members. It provides a permanent link between researchers, projects, and publications, while adding value in terms of recognition, evaluation, and monitoring of research results.

In our context and in addition to pre-existing PIDs such as ISNI (2), ISIL (3), Crossref Funder ID (4), the first step was to choose the PID corresponding to the specificities of the institution. Among the selection criteria identified, we determined that the non-commercial aspect and the scientific and/or research field were the reference points to be considered.

The Research Organization Registry (ROR) (5), which is defined as an open directory of permanent identifiers for research organizations, was therefore a logical choice following the transition of the work carried out by the GRID (6) to the ROR in Q4 2021.

In 2023, it had over 102,000 entries for more than fifty integrations with different systems. It is also the default identifier supported by Crossref DOI, DataCite (7) DOI metadata and ORCID.

This directory is managed centrally. New registrations are created using a web form and go through a committee that checks the information provided, the scope and the metadata before validating them. This information is made available within a maximum of 6 weeks.

Phase 2: the persistent identifier for researchers/authors

Researchers have a myriad of solutions for making their Academic profile web site available (8). Due to the pre-existing use of the Open Researcher and Contributor ID (ORCID) (9) for some of our researchers, we focused on this one which meets our objectives.

ORCID is a PID that is widely known and disseminated in the world of research and scientific publications. It has many advantages (10), such as identifying and tracking an author by eliminating the risk of ambiguity with homonyms, and tracking collaborations in which the author may have participated. It also helps authors to have their work recognized (11) within a specific institution and is increasingly integrated into the

workflows of publishers and funding agencies.

Two main approaches were considered, creation by researchers or centralized creation of ORCIDs. This second approach was not considered realistic due to the library's limited human resources and the risk of duplication for researchers who already have an account.

It should be noted that it has been decided not to make it compulsory for researchers to create an ORCID ID. An internal presentation was made on (2021/04/19) to ask researchers who already have this PID to send it to the library service and to invite researchers who do not yet have an account to create one.

It included a presentation of the ORCID ID, an explanation of how to create and populate a new account, and a description of the possible benefits for the various stakeholders involved:

- for the library, the addition of this information in the library catalogue at the level of the descriptive author record. This allows customers to uniquely identify the author and his other internal or external publications;
- for human resources, which, once it has been integrated into the researcher's personal file, can use it to identify publications to be considered when promotions are made, or new posts opened;
- for the researchers themselves, by establishing a link between them, their publications and, by extension, the institution. A proposal for specific support from the library service for the creation of the account and the automated addition of these references from the information available in the catalogue was communicated internally. A specific document explaining this procedure has also been created and made available on the institution's intranet to serve as a guideline in the process.

This phase, which began in 2021, has become one of the tasks of the library to support the institution's new employees. A communication plan has been drawn up, with regular reminders to researchers to create their accounts and update their content by email and/or at weekly team meetings.

Currently, 55% of the institution's researchers have responded positively and have an active account.

Phase 3: the persistent identifier for documents

The concept of persistent identifiers in the web world is relatively old and has produced many emanations. Without going into detail, we can already think of the

Uniform Resource identifier (URI) put forward at the end of the 90s, the Persistent Uniform Resource Locator in 1995 and the Archival Resource key (ARK) in 2001. Each of these has its advantages and disadvantages (12).

20 years after the launch of the International DOI Foundation (IDF) (13), and the emergence of the Digital Object identifier (DOI), it is possible to consider it as a mature PID that is prominent in the world of research and scientific publication.

Its creation to provide a unique identifier for an object separate from its location and its characteristics of flexibility, actionability, resolvability and interoperability meets the FAIR Principles (13). It makes it possible to link an object to its metadata on a digital network. This object can take many forms such as books, periodicals, and journal articles.

It originally consisted of 4 parts: a schema (doi:), a prefix (designating the naming agency), a separator (/) and a suffix (alphanumeric value from the naming authority).

Nowadays, practice has modified its composition by replacing the scheme with the link resolver (server name). More specifically, for this project, it allows us to make the link between the object and the author, and then from the object to the institution.

The DOI services pyramid is made up of 5 roles (14). The first two correspond to the Registration Authority (RA) and the Registration Agency (RA). The KCE fulfils 2 other roles. It is the registrant, managing and maintaining the data and URLs and providing the suffixes. And as customers, through contacts with researchers, quality control of publications by our Board of Directors and management of the infrastructure needed to preserve and share documents.

The final role is that of users.

Approach

Identification of the documents concerned

An analysis of the KCE's situation has identified more than 1,438 documents produced by the 370 studies and collaborations carried out over the last 20 years. These documents are subdivided into collections (KCE Reports, KCE collaboration, etc.) and sub-collections defined based on 4 of the institution's 5 areas of expertise (HSR, HTA, GCP, Methods). Each study produces different types of documents, the scientific report (from 1 to 9 documents), the synthesis (from 1

to 3 documents), the supplement (from 1 to 4 documents) in three languages (Dutch, French, English). This gives us 1345 documents for this project, which focuses solely on the products of studies requiring a DOI.

Selection of the registration agency (RA)

In order to have DOIs, it is necessary to work with a registration agency (RA). This provides the DOI prefix, registers the DOI and provides access to the infrastructure needed to declare and manage document metadata. A total of 12 (15), of them, are located all over the world, offering different types of services on a paid or free subscription basis. The KCE selected a RA based on two key criteria for us.

The correspondence between the scope of the RA and that of the KCE. Two agencies were identified on this basis: Crossref and mEDRA (16).

The second criterion was the geographical location of the agency and its servers. The servers must be in Europe for the primary storage of metadata, in order to guarantee compliance with the GDPR rules. The mEDRA agency was chosen after contacts with us confirmed that the data supplied by the KCE was indeed stored on servers within the borders of the European Union.

After several discussions with this agency, we have selected a DOI Bracket "2" subscription which includes 170 new DOIs per year and a subscription for the 1338 catalogue DOIs which correspond to documents prior to 2022.

Implementation

Once we had received the identification codes and the prefix to be used, we carried out a more in-depth analysis of the metadata needed to create new PIDs based on the documentation (17) available on the agency's website and the "ONIX for DOI metadata schema" set up by mEDRA. The necessary information was already centralized in our institution's catalogue (18) (a Content Management System open access PMB) and managed by the library service. This catalogue also includes the institutional repository.

Test phase

mEDRA offers two types of registration interface. An XML upload accompanied by a code validation service or a Web editor.

This editor allows you to choose the type of publication to be registered (monograph, book chapter, journal, series, etc.) and a web form containing the various fields to be filled in to collect the document's descriptive metadata.

We decided to test the web editing interface to manually add the publications concerned. This simple interface consists of 7 parts: Message (identification of the institution), DOI (Suffix and URL of the document), Monograph Data (Meta-data describing the document); Additional Data (abstract, keywords, audience, etc.), Relations (Work, Product) Citations Data and Confirmation (sending the information to mEDRA).

In practice, the workflow was divided into 6 stages:

1. checking and validating the data available in the library service catalogue;
2. introduction of metadata via the Web editing interface;
3. verification of document access via the link provided;
4. modification of the document record in the library service catalogue by adding the DOI link;
5. addition of the DOI link to the web page describing the publication on the institution's website;
6. addition of the DOI to the KCE's internal publications database, which only includes the "Scientific reports" publication type in English.

Our approach was to reverse-encode from the most recent publication to the oldest to give priority to the latest publications. However, following internal discussions, we modified this option by working directly on a specific sub-collection (HTA) to facilitate its integration for updating an external database.

This test phase showed that creating these DOIs was relatively simple, but time-consuming. The time taken to create DOIs for all the publications concerned was estimated at between 18 and 24 months, considering ongoing projects and recurring tasks.

Adapting the process

Because of this relatively long lead time, it was decided to work directly on the XML upload. As all the metadata is already available in the institution's catalogue, we contacted our service provider (19) to have an export file developed using the ONIX for DOI metadata schema provided by mEDRA.

After two months of development, we had a stable export model. In line with our internal policy, this devel-

opment has been made available to the CMS user community so that other libraries can use it.

The workflow has been adapted as follows:

1. checking and validating the data available in the library service catalogue;
2. export of the XML file of references to be processed;
3. validation of the file using the verification tool provided by the service provider;
4. upload of the validated XML file;
5. verification of document access using the link resolver <https://doi.org/>;
6. modification of the document record in the library service catalogue by adding the DOI link;
7. the DOI link is added to the citation block generated on the page describing the publication on the institution's website;
8. addition of the DOI to the KCE's internal publications database, which only includes the "Scientific reports" publication type in English.

Although apparently longer, it enabled us to finalize the processing of all the publications concerned in three months.

Experience feedback

Several observations can already be made about this project, which has been integrated into the library service's missions. First, the involvement of everyone within the institution is necessary in order to be able to adapt the process for welcoming new researchers to the institution, to make them aware of the importance of PIDs, to inform them of the impact that these can have in their professional contexts and to integrate the DOI link into the various models of documents relating to KCE studies.

These new tasks can be time-consuming at many levels, such as communicating about PIDs, helping researchers to manage the ORCID ID and creating DOIs for studies producing documents. These tasks need to be monitored and integrated into day-to-day work.

Setting up DOIs for scientific publications produced by the institution's studies has an impact on the structure of these collections. For example, when a study is updated and a new report published, we must give a new unique number in the collection to generate the DOI. Previously, for this type of situation, the original number was simply retained.

There are questions about DOIs (what happens to them if the institution disappears, what is the impact of "obsolete" DOIs (no longer corresponding to an object accessible online), etc. which will need to be answered in the future.

Conclusion and prospects

This work has taken a relatively long time to complete due to the difficult health circumstances of recent years. Despite some positive feedback from researchers outside the KCE, from students and from visitors to our website, it is still too early to be able to measure the impact of the combination of these three PIDs.

It should also be considered that this is in-depth work which does not end with the closure of a project, but which leads to further reflections to refine and improve it. The possibility of integrating these references into Crossref with the help of mEDRA as a service provider, with the aim of increasing the impact of the work already carried out, or of setting up specific DOIs for the results of clinical trials funded by the KCE, remains open. These are just the first steps in an adventure that should continue over time.

Received on 23 August 2023.

Accepted on 10 November 2023

REFERENCES AND WEBSITES

1. Vierkant P, Schrader A, Pampel H. Organization IDs in Germany. Results of an assessment of the status quo in 2020. *Data Science Journal*. 2022;21.
2. ISNI International Agency (ISNI-IA). What is ISNI ? [cited 2023 2023-07-25]; Available from: <https://isni.org/page/what-is-isni/>
3. Royal Library of Belgium (KBR). ISIL. [cited 2023 2023-07-25]; Available from: <https://www.kbr.be/en/isil/>
4. Lammey R. Funder Registry. 2020 [cited 2023 2023-07-25]; Available from: <https://www.crossref.org/services/funder-registry/>
5. Research Organization Registry (ROR). About. [cited 2023 2023-07-25]; Available from: <https://ror.org/about/>
6. Global Research Identifier Database (GRID). GRID passes the torch to ROR. [cited 2023 2023-07-25]; Available from: <https://www.grid.ac/>
7. DataCite. DataCite Support. [cited 2023 2023-07-25]; Available from: <https://support.datacite.org/>
8. Zhang L, Li C. Investigating science researchers' presence on academic profile websites: a case study of a Canadian Research University. *Issues in Science and Technology Librarianship*. 2020(95).
9. Researchers CRa. About ORCID. [cited 2023 2023-07-25]; Available from: <https://orcid.org/>
10. Covay DT. ORCID @ CMU: Successes and failures. *Journal of eScience Librarianship*. 2015;4(2).
11. Grove K, Soper D. Launching an institutional ORCID initiative at Florida State University. *Issues in Science and Technology Librarianship*. 2021(96).
12. Vasile A, Guillaume S, Aouini M, Michaud A. Le Digital Object Identifier, une impérieuse nécessité? L'exemple de l'attribution de DOI à la Collection Pangloss, archive ouverte de langues en danger. *I2D - Information, données & documents*. 2020;2(2):155-75.
13. Liu J. Digital Object Identifier (DOI) under the context of research data librarianship. *Journal of eScience Librarianship*. 2021;10(2).
14. LIU J. Digital Object Identifier (DOI) and DOI services: an overview. *Libri*. 2021.
15. Foundation d. Existing Registration Agencies. [cited 2023 2023-07-25]; Available from: <https://www.doi.org/the-community/existing-registration-agencies/>
16. EDISER SRL. mEDRA. [cited 2023 2023-07-25]; Available from: <https://www.medra.org/en/>
17. SRL E. Technical Documentation. [cited 2023 2023-07-25]; Available from: https://www.medra.org/en/t_documentation.htm
18. Belgian Health Care Knowledge Center (KCE). KCE Library. [cited 2023 2023-07-25]; Available from: <https://kce.docrsources.info/>
19. Services P. PMB Services. [cited 2023 2023-07-25]; Available from: <https://www.sigb.net/>

