

The 19th International Conference on Open Repositories, June 3-6th 2024, Göteborg, Sweden (OR2024 Lightning Talk (or 24x7) Track Submission)

From Invenio1 to DSpace7: Sustainability of an Institutional Repository in Practice – text of the conference proposal

Hana Vyčítalová (ORCID iD 0000-0002-8323-5790), National Library of Technology, hana.vycitalova@techlib.cz;

Petra Černošlávková (ORCID iD 0000-0001-8683-6156), National Library of Technology, petra.cernoslavkova@techlib.cz;

David Gerner (ORCID iD 0009-0007-6992-5609), National Library of Technology, david.gerner@techlib.cz



Abstract

The National Library of Technology (NTK) has operated its institutional repository for publications and other outputs for over ten years on the open-source platform Invenio 1. With a lack of technical staff over the years, the maintenance of the repository was getting more and more complicated. By 2022, the situation was no longer sustainable, and we had to face difficult choices. To ensure the repository's sustainability, it was necessary to change the software and process the data migration. We chose DSpace 7 as it meets our requirements well. Nevertheless, there are only a few implementations of DSpace 7 so far, none of them in the Czech Republic (only version 6). In all circumstances, it meant many challenges and mistakes during the DSpace 7 implementation. Our experience with the sustainability of a repository on a short budget is very fresh. We want to share it with the community, so our mistakes don't have to be repeated. The contribution briefly describes the NTK repository context and shares the challenges and mistakes we made while implementing the chosen solution.

Keywords

Institutional repository, open-source software, sustainability, data migration

Audience

The main target group are librarians, repository managers and curators. Some insights about software migration may also be useful for repository administrators.

Proposal

The Institutional Repository of the National Library of Technology (NTK) was launched in 2013 to collect, archive, and make available works produced in the NTK for staff and the public. This allowed the NTK to present its professional and cultural activities, such as

annual reports, articles, papers, photographs, and videos from professional conferences and cultural events, project outputs, etc., through another channel.

The NTK generally supports open-source technologies, and the NTK had already successfully run the National Repository of Grey Literature on the open-source Invenio 1 at that time. Therefore, the choice of software for the institutional repository was simple. One of the main advantages was shared staff for both repositories. The repository management had been handled by two staff members, one in fulfillment, metadata, file management, and user support, and the other in technical administration and software maintenance. Over time, the capacity to be devoted to the institutional repository diminished as the workforce was needed elsewhere. However, the worst for repository sustainability had yet to happen when a colleague who had consistently maintained the Invenio software for several years left in 2018. The repository was left without any technical administrator, and nobody in the NTK had any experience with Invenio.

In addition, Invenio 1 maintenance support ended the same year, and an upgrade was needed. Invenio version 2 was never complete in a way that could satisfy NTK's needs and came to the end of life together with version 1. Both versions were replaced by version 3, which is suitable only for large-scale repositories with a sufficient developer team. (CERN, 2020) The institutional repository still needed to be maintained at version 1 for several reasons, mainly due general difficulties in hiring senior developers for a governmental institution that can offer only lower salaries. Nevertheless, the occurrence of bugs and outages was increasing, and the newer security requirements became more and more challenging to fulfill.

By 2022, the situation was no longer sustainable, and switching quickly to another solution was necessary. Invenio 3 wasn't an option as it was too robust and complicated for our small repository (circa 1.700 records) (CERN, 2019). Regarding sustainability and cost minimization, we focused again on open-source software – this time on DSpace and Islandora. A brief analysis of both showed that DSpace better meets our requirements, has a solid international development community, and is used by many Czech universities and colleges. However, the actual implementation of the software, repository setup, and data migration also proved to be very challenging; instead of the planned three months, the whole process took over a year, primarily due to no previous experience with DSpace or repository migration. The timing of the transition to the new software was also not the happiest. The latest version of DSpace 7 has just been released, and so far, only a minimum of repositories have applied it, not even one in the Czech Republic (Masaryk University, 2024). The situation was further complicated by the many bugs in the new version 7, which were gradually addressed in minor versions of DSpace 7.

The NTK's use case shows how tricky the sustainability of even a smaller institutional repository can be. The challenge is obvious, at least for the Czech Republic. So far in the Czech Republic, the repositories have usually been minor projects, mostly on universities for theses and dissertations, using DSpace 6. Even though the Czech universities are testing DSpace 7, none is ready to or even plan to upgrade to version 7 in the foreseeable future. This contribution aims to share our fresh experience with the sustainability of a repository with a short budget solved by migration to another open-source software. The contribution briefly describes the NTK repository context and shares the challenges and mistakes we made while implementing the chosen solution.

References

CERN, 2019. Invenio Framework: Open Source framework for large-scale digital repositories. [online]. CERN. Invenio. [Accessed 4 January 2024]. Available from: <https://inveniosoftware.org/products/framework/> [Accessed 4 January 2024].

CERN, 2020. Maintenance Policy – Invenio 3.4.0 documentation. [online]. CERN. Invenio Framework 3. [Accessed 4 January 2024]. Available from: <https://invenio.readthedocs.io/en/latest/releases/maintenance-policy.html>

MASARYK UNIVERSITY, 2024. DSPACE V ČR. MASARYK UNIVERSITY. DSPACE CZ [online]. [Accessed 4 January 2024]. Available from: <https://www.dspace.cz/dspace-v-cr>