

## THESEUS.FI – OPEN ACCESS PUBLISHING IN THE FINNISH UNIVERSITIES OF APPLIED SCIENCE

Herkko Hietanen, Anna-Kaisa Sjölund

In 2008, Finland's universities of applied science created a digital repository for theses and research publications at <http://www.theseus.fi>. The essential goal of this project was to build an accessible, user-friendly, flexible, transformable digital publishing platform and repository for academic thesis and research papers.

The Theseus project was founded together by Finland's Ministry of Education and the country's 26 universities of applied sciences. The joint venture consisted of two parts: An open access web journal and a repository for publications. The universities of applied sciences in Finland agreed on an open access statement to guide their publishing policies. One of the guiding objectives of the Theseus project is to promote open access publishing among the collaborating universities.

The guiding principles of open access and open source solutions are visible throughout the publishing system. Authentication of the theseus.fi service uses Shibboleth software, students can choose a Creative Commons license for their publications, and the platform is built on open source DSpace software.

### Theseus.fi

The Theseus project combines the digital repositories of the 26 participating universities of applied sciences. The repository is expected to digitally store and make available over 20.000 new academic theses annually. The digital repository solves the problems of access, storage, and preservation in a user-friendly fashion.

The repository consists of several technology layers that are provided by different technology partners. The key technology that the repository uses is the DSpace platform. DSpace is open source software that enables open sharing of content. The platform is localized by the National Library of Finland who is also responsible for the technical integration of theseus.fi system.

The Theseus repository relies on author publishing. Students upload their theses to the repository themselves. In order to ensure that the system is reliable, the publishing procedure includes electronic authentication, which will be carried out by the Haka Federation, the identity federation of Finland's higher

education and research institutions. The Haka Federation, which is operated by the IT Center for Science, uses Security Assertion Markup Language 2.0 technology and open source Shibboleth software. Haka users are able to access federation services using a single user account and password. In this case, users, students, and teachers, are able to access the services using their home organization's username and password.

### Legal issues

Copyright issues are important for any text repository. Theseus has tackled these issues by incorporating a reliable identity management system into the publishing procedure and by enabling authors to choose a Creative Commons license for their published works. Providing information for authors and their instructors about the distribution and licensing policies was one of the challenges of the Theseus project.

The creation and storing of copyright metadata is an integral aspect of the Theseus publishing procedure. The publishing system automatically creates copyright information with responding metadata for every published work. The reason for choosing Creative Commons (CC) licenses was their interoperability and popularity in the academic world. The CC licenses provide a free, easy way to express the rights that are granted for use of the work. The attachment of legal metadata serves the purposes of wide dissemination of works by creating clear legal rules for sharing those works while at the same time respecting the authors' copyrights.

The Ministry of Education has published a policy for academic thesis restricting the use of trade secrets in theses and forcing public access to theses. Until now, many of the participating universities have kept their theses available only within their own libraries. Several of the universities of applied sciences have close R&D co-operation with companies. The trade secret and open access demands of the theses policy has turned out to be complicated.

Open access offers visibility for universities, sponsoring companies, and students. However, students often write their theses about practical problems they have helped to solve with their sponsor. This means that

sponsors and students have to consider whether they have utilized trade secrets in their work. Public availability of theses in an easily accessible database means that a sponsor's competitors can freely access a project and its results.

Nevertheless, the interests of openness and students working with trade secrets do not have to conflict. Student instructors will need to advise students and sponsors early on to formulate their research questions so that trade secrets are not jeopardized.

### Problems with implementation

The implementation of Theseus has faced several challenges. While DSpace is used in several universities, there were no reference implementations for the Creative Commons licensing module. The project management had to decide the licensing features and procedures that the publishing process was to include. The DSpace platform and the use of PDF files added technical restrictions on how license selection could be implemented.

As the Creative Commons licensing model is rather novel, some of the technology providers had a hard time understanding the rationale for using such licensing. This was reflected in an unwillingness to find creative solutions for technical problems, despite their client's clear policy that the Creative Commons licensing was to be a feature of the publishing system.

The biggest problem left unresolved was the attachment of licenses to the published documents. When authors enter the publishing workflow at the Theseus system, they typically have their work prepared and transformed into a PDF document. However, the publishing system creates a permanent URN-address for the file, metadata, and licensing information that would be beneficial to have stored in the PDF document. For example, people often link directly to the main document and if the link is not attached to the PDF document, the valuable metadata information that the system generates does not travel with the work.

One way of adding the metadata would be to have an automatic operation in which the system generates a metadata page that is attached to every published document. However, the idea of automatically modifying an authors' works, which might be technically protected, raised concerns, and the project decided not to add metadata pages to documents.

### Streaming

Text is not the only form in which theses are produced by participating universities. Students also demonstrate their skills and ideas using videos, computer programs, and music. Most theses are still in the traditional written format but many of them include sound samples and video. Another important user group for Theseus is students and researchers who dependent on sign language. Streaming has been understood as a means of presenting theses and other publications in sign language. For these reasons, Theseus wanted to provide means by which audiovisual content could be stored and shared by streaming.

Streaming applications turned out to be too complicated for universities. For some of the universities, there were too many steps before the video is on the Internet. During the pilot period, only one university took advantage of the streaming module. However, Theseus does have these streaming capabilities.

### Conclusions

The twenty-six universities of applied sciences in Finland have taken a huge step towards open access to academic research. The universities of applied sciences have been open-minded in choosing to catch, together, the open access ball, which has been rolling since 1966. The active development of the Theseus platform will end in December 2009 and the software will be available for use in 2010.

One of the takeaways is that building an academic publishing platform is not merely a technological exercise. Students, teachers, and sponsors must be familiar with the rules of open access publishing well in advance. It is also necessary to have a clear mission statement for the project that all participating schools and publishing authors support. For Theseus, this boils down to the following statement: "After 1 January 2010, the Universities of Applied Sciences will require all teachers and researchers who work at the universities to save a copy of their research essays that are published in scientific publications, or a university publication series, in the open electronic library, Theseus." (<http://theseus.fi/en/julkilausuma.html>) With this statement and the Theseus publishing system, the Finnish universities of applied sciences are playing their part in the global open access movement.



**Herkko Hietanen**, PhD., Turre Legal Oy.



**Anna-Kaisa Sjölund** Library Coordinator, Open Access Project, Seinäjoki UAS