

10 YEARS WITH GREY LITERATURE AT TOMAS BATA UNIVERSITY IN ZLÍN

Lukáš Budínský, Ondřej Fabián

budinsky@knihovna.utb.cz, fabian@k.utb.cz

Tomas Bata University in Zlín, Library, Czech Republic

This paper is licensed under the Creative Commons licence: CC-BY-SA-4.0 (<http://creativecommons.org/licenses/by-sa/4.0/>).

Abstract

The Library of Tomas Bata University in Zlín has been working on accessioning academic grey literature systematically for the last 10 years. During this time, one of the first information systems storing theses and dissertations changed from a closed, manually updated system into a fully automated and fully open repository with a number of integrations with both internal and external systems. New types of documents were added and several services were introduced for working with grey literature effectively. This contribution will describe its current state and outline the hurdles encountered getting there.

Klíčová slova

Grey literature, theses, dspace, fulltext, license agreement

Introduction

Tomas Bata University in Zlín (TBU) is a dynamically developing university whose six faculties provide a wide range of studies in the humanities, natural sciences and technical and artistic disciplines. It is a centre of top-class science and research in the country and, in

many aspects, on an international scale. With around 10,300 students, TBU is a medium-sized university in the Czech Republic.

The library at TBU is known for its active approach to new challenges that affect it either directly or indirectly. It uses these opportunities to develop the range of services it offers students, academic staff and the public.

One of the challenges to which it has responded, as one of the first institutions in the Czech Republic to do so, is that of making dissertations and theses available in electronic format.

The origins of electronic dissertations and theses at TBU

It was already clear back in 2004 that the situation in taking in and making dissertations and theses accessible did not respect the possibilities and needs of all sides involved. Students were forced to print out 3 copies of their electronic documents, which piled up at institutes and were subsequently handed over to the library to be processed. Each paper was put into a catalogue and placed in the archive by hand. The library at that time was housed in a building with limited space and the ever-growing number of printed papers could no longer be stored. Students only had the opportunity to search them using basic metadata, before collecting them from the archive.

All these negative factors could be eliminated by making dissertations and theses available in electronic format. The document was created electronically, meaning there was no reason not to leave it as such. The library seized the moment in updating the “Uniform formal regulation of university dissertations and theses, their storage and accessibility” guidelines, in which electronic access to dissertations and theses was mentioned for the first time, specifically in the following points of Guidelines of the Rector No. 41/2004¹:

(3) All papers shall be kept on record in the electronic catalogue of the TBU University Library and in the Digital Library of University Dissertations and Theses (hereinafter referred to as the “UDT Digital Library”), where they shall also be stored.

(4) After having been processed, papers which have been defended shall be accessible within the UDT Digital Library in accordance with Licence Agreement.

It must be recognised that pushing through the e-UDT was a step into the unknown because no UDT digital library even existed at the time the guidelines came into force. The Study Information System (IS STAG) was modified during the winter so as to be able to link one file with full text to the submission of a paper. At the same time, an electronic licence agreement form was prepared for students to fill in when handing in their work and in doing so restrict the possibility of accessing the full text of the paper. Metadata was exported from IS STAG and a MARC 21 record was machine-created for import to the library catalogue.

¹ UNIVERZITA TOMÁŠE BATI VE ZLÍNĚ. Jednotná formální úprava vysokoškolských kvalifikačních prací, jejich uložení a zpřístupnění. Univerzita Tomáše Bati ve Zlíně, 1.11.2004. Článek 3, Uložení a zpřístupnění. Guidelines of the Rector No. 41/2004. Available at: <http://www.utb.cz/file/12033/>

Let the problems begin

The switch to electronic format demanded a number of changes in processing UDT that did not pass without any teething troubles. The year 2005 was therefore considered transitional and the obligation to submit UDT in electronic format was anchored in the updated Guidelines of the Rector No. 5/2006². Most of the shortcomings were gradually, but successfully resolved. Unfortunately, some are related to the system itself and persist to this day. What exactly were the problems?

- **Formatting papers and subsequent conversion to PDF**

Given the accessibility of full texts in electronic format to the general public, students were required to stick to uniform rules relating to the formal layout of their work. Templates were created for individual faculties in which students formatted their work. Unfortunately, poor computer literacy among students, combined with insufficient support for the use of templates, led to the situation in which library employees spent hours on end helping with the formatting of papers. Even then, the storage site contains a number of incomplete (no title page) or wrongly formatted e-UDT.

Back in 2005, of course, there was no comfortable tool available to facilitate conversion of documents from MS Word format to PDF file. Students used various tools that added undesirable watermarks in the resulting texts or that required a password for printing or even reading. There were no rules for checking legibility when adding a file. A problem also arose during subsequent full-text indexing, when certain programmes managed to properly convert the text visually, but Czech accents were entirely missing when extracting the text.

Solution: both problems essentially solved themselves over time as a result of students becoming better informed. Comments were made on templates and these were simplified so much that students no longer had any problem properly formatting the resulting text. Removing the licence agreement meant that they no longer had to integrate it with such complication in the resulting text.

A tool for uniform conversion to PDF documents (and instructions) was installed and pre-set on all computers in the library's study rooms. Over time, quality conversion to PDF documents became a standard feature of the main word processors.

- **Students adding information about their work**

Although efforts are made to ensure that information about a UDT is taken from the study system to as great an extent as possible, one integral part of submitting a paper is the provision of information by the students themselves, particularly the abstract and key words. Key words are still something that many students do not understand to this day. Whole sentences appear in the records, frequent misspellings and phrases such as "there are none" and "don't know". The fact is that key words are important when searching documents and for their indexing, particularly in a library system.

² UNIVERZITA TOMÁŠE BATI VE ZLÍNĚ. Jednotná formální úprava vysokoškolských kvalifikačních prací, jejich uložení a zpřístupnění. Univerzita Tomáše Bati ve Zlíně, 17.4.2006. Guidelines of the Rector No. 5/2006. Available at: <http://www.utb.cz/file/11946/>

Solution: two alternative solutions were tested. First of all, the library tried to increase awareness among students at dissertation seminars and then in the form of hints by the box with key words. Unfortunately, neither method had any significant effect on the quality of key words provided. The other route was to correct key words before import to the library system. Nonsense was first removed by machine. Then several librarians divided up thousands of words among them and checked for misspellings. This was manageable in the first year given that they no longer had to catalogue all papers by hand. In the second year, correcting around 60,000 key words met with clear opposition and was eventually dropped.

The eventual solution came with the application of the DSpace system, which indexes documents full-text, meaning that key words became less important to the end user.

- **Importing e-UDT to the library catalogue**

Importing to the library catalogue is closely related to the previous problem. Picking complete metadata from the study system meant that the next step was to end the cataloguing of UDT by hand and import machine-generated records. Unfortunately, this plan met with strong opposition from the cataloguers, who considered the machine record to be imperfect and unsuitable for import. It was able to add papers to the catalogue for several years, but a change of concept eventually became a necessity as a result of the key words mentioned above.

Solution: there are several appropriate ways that can be chosen according to preference.

1. Continuing with the import of machine records and their subsequent modification. It is not a problem to create your own base for UDT in the catalogue, one that will not be mixed in indexes with hand-generated records of books acquired. It is now nothing special to import a collection of tens of thousands of electronic books records directly to a catalogue. It would be possible to leave out hand-generated and error key words.
2. Putting a separate system for providing access to e-UDT into operation. Informative education helps users fast become accustomed to looking for a certain type of document elsewhere than in the catalogue. The situation now has been simplified thanks to the existence of discovery systems that know how to integrate search results across sources.

- **Full texts and access to them**

Access to the full texts of e-UDT mainly underwent development in the first 5 years after being launched. Students had the opportunity to restrict access to their work back in 2006 using a licence agreement as follows: (1) prohibit its publication entirely; (2) a time embargo on publication of 1, 3 or 5 years; (3) restriction to the internal TBU network, and; (4) free publication. The opportunity to entirely prohibit publication was replaced in 2007 by an embargo for up to 30 years. These possibilities led to a frequent embargo on papers or restriction for an extremely long time.

Solution: first of all, the embargo was lowered to 10 years in 2008, before the licence agreement was abolished entirely in 2010. All papers are published in the following regime to this day:

1. all papers are mandatorily published without restriction;

2. in special cases, access to a paper may be restricted to the TBU internal network at the request of the dean of the faculty;
3. students have the chance to keep companies anonymous in the case of sensitive data.

Other documents we could class as grey literature have individual accessibility settings according to the requirements of faculties, institutions and authors.

- **The unsuitability of the study system for long-term archiving and providing access to documents**

Storing e-UDT and making these accessible in the study system was understood to be provisional from the very beginning. It was important to begin gathering data in any way, but making it permanently accessible was more of a marathon than a sprint.

Why is the study system unsuitable for permanent accessibility to e-UDT when all metadata is produced therein and full texts are stored there? The reasons are primarily as follows:

1. Technical:

The study system is one of the key systems in running the university and deals with all operative activities. The data therein must be safely stored and quickly available. In comparison with this, e-UDT are more like archive material that must be accessible over the long-term. After several years of storing full texts, the study system began placing huge demands on disk arrays, backing-up and the IT infrastructure as a whole.

2. Library-related:

Entirely different requirements are made of a system for the permanent accessibility of documents from the librarian's perspective. The system must be robust, secure and scalable. Nonetheless, it should also be open and comply with standards for cooperation with the surrounding world. The aim is to integrate data about e-UDT in a continually changing environment. The form of the library catalogue has already changed in the 10 years of its existence -> repository -> meta-searching -> discovery system.

Solution: several possible methods were considered. One that was quickly rejected was modification of the study system for the needs of long-term archiving and making documents accessible. The requirements of the two systems differ so greatly that it would have been technically demanding and in all likelihood unsustainable over the long-term.

The second option was to use the Aleph library system, for which the library has an ADAM module for providing access to electronic documents. As was shown in applicability tests, however, this module was created only as a quick stand-in for a fully-fledged system and was not developed any further. It did not, therefore, meet the functional demands of storing e-UDT.

Acquiring a fully-fledged commercial system did not even come under consideration in light of budget restrictions and so the library turned its attention to open-source projects. There

were rapid developments in this field in 2009 and the range on offer became sufficiently diverse. The DSpace system was eventually chosen, mainly for its active Czech user group³.

DSpace at TBU – from grey literature to publication activity

Open-source systems have an undoubted advantage in terms of their acquisition costs, although the costs of implementation and of their operation must also be taken into consideration. These can be covered by the user itself or in the form of outside services. In the case of the library at TBU, these costs were spread over time. Implementation was entrusted to a contractor. A tool for importing records of e-UDT from IS STAG was developed to order. Ing. Ivan Masár from the library's IT department then took charge of the subsequent operation and development of the system and his active approach saw him become an official DSpace contributor.

What did the library and users gain from the application of DSpace⁴?

- a robust and expandable system capable of long-term archiving of e-UDT;
- automatic import of papers after defence, with reports on results;
- persistent identifier of records (Handle);
- a uniform, logically-structured metadata storage site;
- the storage of full texts and other documents relating to records (appraisals, defence);
- the opportunity to store non-text appendices and display them (photographs, videos);
- full-text searching and indexing of documents;
- the opportunity to restrict access to full texts according to licencing agreements;
- standardised protocols for communication with external systems;
- access to the source code and the option to broaden the functionality of the system.

The library was able to concentrate on broadening the type of documents stored after putting a stable system into operation. Faculties were first offered the electronisation of published scripts which had not been sold at the planned level of circulation for some considerable time. Although students unanimously voted in favour of electronisation of scripts in a survey, conservative faculties resisted for several years, insisting on loss-making printing. Only the Faculty of Applied Informatics and the Faculty of Humanities took up the offer and now publish scripts and documents in electronic format, with access restricted to within the university. Other activities that promoted electronic scripts at other faculties were discontinued due to lack of interest.

The DSpace platform saw most development at TBU in the form of the institutional repository. The path that was chosen, for many reasons, was complete separation of the digital library of e-UDT and the repository. Although the repository of publication activity is a unique application of the DSpace system, it deals with the issue of grey literature only peripherally (automatic archiving of published articles, conference proceedings etc.). For

³ <http://dspace.cz/>

⁴ <http://digilib.k.utb.cz>

more detailed information on this issue, we recommend older pieces written by the author⁵ and⁶.

Opening up grey literature to the world

The application of the DSpace system allowed grey literature to be accessed within the university and at the same time opened it up to the world. The first stage of integration, however, had come earlier with the use of services of the Theses system. Given that plagiarism is part of the study agenda, the development of this interconnection continues in the STAG system. It now functions on an entirely automatic basis – full texts which are submitted are immediately recorded on Theses servers and checked, with the result of this check being displayed in the study system.

However, the connection of TBU to the network of repositories through the OAI is of greater benefit to the public, in that they offer fast and user-friendly access to grey literature. Among the many are the National Repository of Grey Literature⁷, the Bielefeld Academic Search Engine⁸ and the regional Zlín Libraries project⁹.

Conclusion

The TBU digital library for providing access to e-UDT, scripts and other documents has been under construction for 10 years now. It experienced a number of teething problems and has now reached productive age. The services which it offers continue to grow, although not at such a rapid speed as with other projects at the TBU library. Its most valuable benefit for the Library itself is not the space saved in the fund or the decrease in unnecessary work with cataloguing, but the experience of the implementation team. The team is still dedicated to developing library services and can avoid certain mistakes ... so that it has the opportunity to make new ones. The TBU digital library is available at this address: <http://digilib.k.utb.cz/>.

References

BUDÍNSKÝ, Lukáš a Ondřej FABIÁN. Institucionální repositář jako nástroj podpory VaV. In: INFORUM 2012: 18. ročník konference o profesionálních informačních zdrojích, Praha 22.-24. května 2012 [online]. Praha: Albertina icome Praha, 2012 [cit. 2015-09-30]. ISSN 1801–2213. Available at: <http://www.inforum.cz/sbornik/2012/12/>

BUDÍNSKÝ, Lukáš. Cesta je ten nejdůležitější cíl – hledejme tu správnou v podpoře vědy a výzkumu. In: INFORUM 2015: 21. ročník konference o profesionálních informačních zdrojích,

⁵ BUDÍNSKÝ, Lukáš a Ondřej FABIÁN. Institucionální repositář jako nástroj podpory VaV. In: INFORUM 2012: 18th Annual Conference on Professional Information Sources, Prague 22-24 May 2012 [online]. Praha: Albertina icome Praha, 2012 [cit. 2015-09-30]. ISSN 1801–2213. Available at: <http://www.inforum.cz/sbornik/2012/12/>

⁶BUDÍNSKÝ, Lukáš. Cesta je ten nejdůležitější cíl – hledejme tu správnou v podpoře vědy a výzkumu. In: INFORUM 2015: 21st Annual Conference on Professional Information Sources, Prague 26-27 May 2015 [online]. Praha: Albertina icome Praha, 2015 [cit. 2015-09-30]. ISSN 1801–2213. Available at: <http://www.inforum.cz/sbornik/2015/40>

⁷ <http://nusi.cz>

⁸ <http://www.base-search.net/>

⁹ <http://knihovnyzlin.cz/>

8th Conference on Grey Literature and Repositories: proceedings [online]. Prague: National Library of Technology, 2015 [cit. 2015-12-15]. Available from: <http://nrql.techlib.cz/index.php/Proceedings>. ISSN 2336-5021.

Praha 26.-27. května 2015 [online]. Praha: Albertina icome Praha, 2015 [cit. 2015-09-30]. ISSN 1801–2213. Available at: <http://www.inforum.cz/sbornik/2015/40>

UNIVERZITA TOMÁŠE BATI VE ZLÍNĚ. Jednotná formální úprava vysokoškolských kvalifikačních prací, jejich uložení a zpřístupnění. Univerzita Tomáše Bati ve Zlíně, 1.11.2004. Guidelines of the Rector No. 41/2004. Available at: <http://www.utb.cz/file/12033/>

UNIVERZITA TOMÁŠE BATI VE ZLÍNĚ. Jednotná formální úprava vysokoškolských kvalifikačních prací, jejich uložení a zpřístupnění. Univerzita Tomáše Bati ve Zlíně, 17.4.2006. Guidelines of the Rector No. 5/2006. Available at: <http://www.utb.cz/file/11946/>