EDIT - Repository construction in then oldest Hungarian university

ELTE
Eötvös Loránd University is the oldest and largest, continually operating university of Hungary. At the turn of the last century, it was added to the list of the world's fifteen largest universities.
There are currently eight faculties of education with more than thirty thousand students, of which two thousand come from abroad.

University Library
The University Library is the central library of ELTE, the first university library of Hungary, which has been operating since its foundation.
It is the third largest library in the capital, eighth in the country. Until nineteen forty nine its collection based on encyclopedic scope, after that it became more specialized.
The size of the substances have been expanding for more than four hundred years, the number is more than two million library units.

ULS
The University Library of Eötvös Loránd University with its 8 faculty libraries and more than 50 institutional and departmental libraries established the University Library Service (ULS), a provider of services coordinated library network.
We have a common integrated library system (Aleph), the professional work is supervised by University Library, as a central library.

The goals of creating a repository
The objectives of creating a repository were: increasing the publicity of research results, recording and archiving the institutional scientific publications, representing the University's common knowledge assets, supporting the management processes and education, publishing PhD theses, e-books.
In addition to the scientific publications, we could have a way to publish so-called grey literature.
The repository might be ideal to store non-public materials - for instance educational issues - written for institutional usage.
Beside documenting projects, contracts, patents, EDIT can be a mass storage for internet and intranet portals.

First steps
To realize ELTE institutional repository we were supported by an EU tender.
We released a procurement tender for a repository, that was won by Monguz Ltd.
The "cloud computing" by ELTE Informatics Directorate gave us the Technical preconditions. The background of the common library support was provided by reforming the library network and establishing the ULS.
Regarding our demands, the flexible and open source DSpace software was chosen as repository framework.

That can ensure the multi-faceted tracing and register of digital data and materials as well as full text search capabilities.

The repository was named EDIT, that is an acronym of the initials of the Hungarian name of the repository.

It was a real challenge to build this repository, due to the size and organizational complexity of ELTE and the wide scope of documents to be stored.

Until two thousand and thirteen spring, Monguz Ltd. developed the repository with our increasing special needs.

In May 2013, with some colleague we took over their work, keeping the custom tailored novelties. Adopting the former development we started new ones and after a while we changed to a newer XML based user interface.

Structure

In September, two thousand and thirteen we made this new XML interface available.

That time the structure of the system mirrored the university hierarchy, so it could be managed by institutional units.

Although this structure proved to be well-managed and well-usable, after two years we decided to change it.

Instead of reflecting the organizational structure of ELTE, our collections were organized into thematic categories.

The conversion was done in December, two thousand and fifteen, and we concurrently changed version.

We are using the latest version now, and will try to follow all the updates as soon as possible.

It is always a big challenge for us, considering, that we have patched and rewritten many sources and tables in the database since we got the development.

So the simple DSpace offered update methods and patches won't work with us, we always have to pay attention on implementing all our own changes.

When changing the structure, we had to keep in mind the little size of our administrative network.

To facilitate the administrative tasks, preserving collections by faculties within some categories was justified. With transforming the structure, EDIT became more transparent and useful.

Why DSpace

It was vital for us, that the chosen system must be able to provide a wide range of file formats, so DSpace is ready to store more file formats, the most part of the textual documents can be discovered by its indexed content.

DSpace provides more search and browse possibilities that were custom tailored with changing the configurations and self-made development.

The multi-level access was very important for us, that can be refined with multi-level rights management.

DSpace is basically built on the conception of open access, though you can find documents that can be discovered only inside ELTE network, in closed network for only some groups or person.

DSpace has built-in SWORD, RDF, OAI module, and SOLR search platform.

Of course, statistics is always a very important part for us. We can increase the effectiveness according to the data coming from it. You can get usage and many other kind of systematic statistical data. DSpace is able to generate statistics according to the log files, SOLR.

Administrative network, rights
A system containing such a big data set and documents from different sources should have an appropriate network administration.
The network administration needs to have determined operating method.
Rights management by collections provides us with more opportunities, specifying the roles, the tasks will be well separated.
To thrive the collections requires different workflows, in which the roles of ELTE citizens can be other and other.
EDIT is basically built on self-upload method where the uploader must accept a placement agreement as a part of the submission process.
Three members of University Library staff control the development, the maintenance and the twenty faculty administrators, who are assigned to special collections.

Custom tailored needs
URL based file upload for the submission was individual need as well as uploading more files concurrently in zip package.
It was important for the development to create so-called private folders, where the registered users can store documents and save search terms that are remarkable.
They can even share documents by e-mail and social media.
The scope of documents to be stored in EDIT can be very broad, the uploaded materials might have unique characteristics.
Therefore we created different types of forms containing special variety of metadata to the different document types.
Each documents in EDIT are given unique Handle identifier, which can be used as a permanent access reference to the documents, even if the storage server or the original url changes.
Building interconnection between EDIT and our electronic library catalogue was essential. The metadata of our documents that are in our catalogue will appear in the repository transferred by batch upload.
We use Qualified Dublin Core scheme in EDIT, while we use Marc21 in our integrated library system following the Hungarian and local standards.
The Storage of Hungarian Scientific Works (MTMT) is a bibliographic database established by Hungarian Academy of Science (MTA). The metadata of scientific publications have been recorded in MTMT can be transferred with file attachment into EDIT by SWORD protocol.
We prefer PDF/A format for text files, that covers the long term preservation of electronic documents. Beside this we implemented DjVu complement to store OCR version of any text content, so the texts stored in the files are discoverable, similarly to PDF/A.
We registered ourselves in Google Search Console and uploaded sitemaps, that contain all our discoverable items. After indexing by Google, now we get statistical data, and have more chance to appear in result list in google.
Beside this we are registered in other reporting services like Awstat, Google Analytics.

Technical realization

Normal documents and images are displayed by browser plug-ins.

If a document has more images with the same orientation and size, we use FlipDoc, that converts images of a document into scrollable books. For example codicies in EDIT.

As you see, the document in FlipDoc is pageable and zoomable.

Manuscripts and books with more pages that have different orientation or size are displayed with Gallerific manager, or we convert them into PDF/A.

In Gallerific you can jump to the proper page with a click or find a page by it's number.

Video or audio files are played by browser plug-ins and strobe media player for flash files.

DSpace has batch import processing, then the packages may contain more items with attachments, license file, content file and metadata xml.

We can export/import metadata in some different formats.

We implemented RDFa Lite recommendations too, that adds a set of attribute-level extensions to HTML, and XML-based document types.

As both EDIT and our integrated library system, Aleph have OAI interfaces, we implemented VuFind.

The goal of VuFind is to enable our users to search and browse our resources by replacing the traditional OPAC.

Aggregators, search engines

Cooperation with aggregation-services is an important part of EDIT-like storages, OAI-PMH compatibility of EDIT makes the stored data available.

PhD thesis stored in EDIT are harvested by DART-Europe.

We registered our repository at DPR Celestial.

We joined to eduOER project with our video store.

We have been cooperating with the Hungarian National Digital Archive and Film Institute for years, which means that through their database our museum collections can get into Europeana.

OpenAIRE compatibility of EDIT allows our research results and publications supported by EU to be stored.

We registered in OpenDOAR.

EDIT can be found in Ranking Web of Repositories international gradation from January this year.

Hungarian Academy of Sciences created a common search engine, in which you can find EDIT as well.

Virág Gabriella
Repository Manager
ELTE University Library
Department of IT and Development

Cerva Csaba
System Developer and DBA
ELTE University Library
Department of IT and Development