

BUILDING A DIGITAL LIBRARY: BEST PRACTICES IN BCU CLUJ

Róbert BODNÁR

e-mail: bodnar_robert@bcucluj.ro

Ágnes BREM

e-mail: brem.agnes@bcucluj.ro

Olimpia CURTA

Central University Library Cluj-Napoca

e-mail: pia@bcucluj.ro

Abstract: *In 2008 BCU Cluj started digitizing its special collections and periodical materials with basic tools and methodology. Since the early period the digitization laboratory grew and gained valuable experiences. It developed a workflow that enables us to manage and track every step of the digitization. DSpace software was implemented to help manage library collections and to enable us a possible connection to the Europeana digital library. The challenges and the solutions of the implementation of this open source software are described here. The paper also discusses the possibilities that the institutions with digital collection can have in providing access to the electronic materials as a new library service.*

Key words: *digital library, best practices, special collections, DSpace, Europeana, BCU Cluj*

1. Introduction

The BCU Cluj Library recognized the importance of protecting but also making available its collections, so it founded its digitization laboratory with the objectives to digitize its valuable and fragile collections. Since the early period the laboratory gained experience in the process of digitization and created one of the biggest online digital library national wide. To get to this point the library had to constantly reevaluate its work processes and improve its tools. This paper will discuss some of these.

2. Workflow

Creating a specified workflow is crucial for the digitization process. The workflow helps monitoring the processes, assures quality products and promotes collaboration between the departments of the library.

The process starts with the *planning and selection*. The lists of documents that will be digitized are selected by the colleagues who are familiar with the collections, and the lists contain all types of documents (old books, manuscripts, periodicals, maps, photos, etc.), documents that the library owns. The selection criteria are: to be a document relevant to Transylvania and Cluj-Napoca, documents with important research values, required for university curricula, or requested by a library user. In the selection process, the condition of the document is also considered: fragile or often required documents will be treated with priority. In this phase, the size of the periodicals will be verified and estimations are made on

the number of pages. This helps planning the time of the scanning process, the selection of personnel who will scan it and the tool that will be used.

Since 2010, an important step was added to the selection: to verify in the National Library's digital document inventory if someone has digitized it. The list is updated (<http://www.bibnat.ro/Inventarierea-documentelor-digitizate-s160-ro.htm>) by the National Library, and BCU Cluj is supporting the National Library's effort to keep this list accurate by sending updates periodically.

The success of planning and selection, and later the scanning process is highly depending on the collaboration between the departments. It is crucial that every department to include in its workflow the selection process, and these departments must be considered as team members in the digitization process and they must get credit for the invested effort.

When the documents are selected (based on topic, size, condition, etc.) the *scanning and processing* phase starts. The laboratory has an internal guide with the technical details that ensure the high quality of the master images. The national library also published a "Guide to digitization" that is a useful resource in setting standards and quality control. The process is showed in Figure 1.

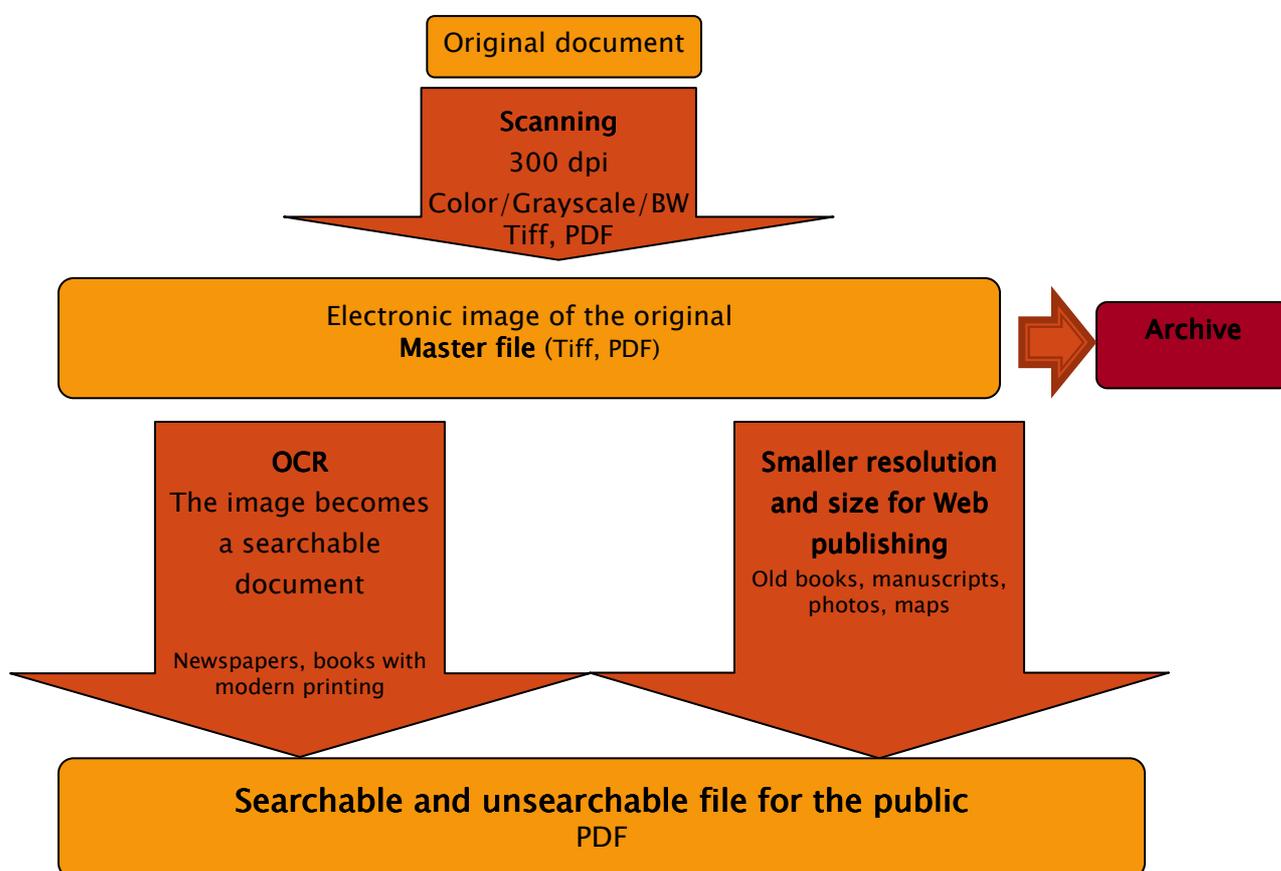


Figure 1. *Scanning and processing*

Through the process, every page of the item will have one master image in TIFF format, created in the best resolution (300 dpi), named using a standard method described in the digitization guide. These master images will be archived, thus later on, other image formats and resolution images can be derived from them.

In the third phase the metadata is created for these files. Only the derived files that are published will have metadata, the master files will not. The metadata is based on the document's bibliographical description in Aleph, the library's ILS. This requires that each document must have a bibliographical description before the scanning process.

There are two ways of creating the metadata. If the document is composed from one or two files, the metadata can be created directly in the database that will store it and publish it (DSpace). In case of periodicals, which can have hundreds of files, the metadata is created in MS Excel, then using a small VisualBasic script, the XML files are created and bundled together with the license file and the actual digital document; later these are batch imported in the database (DSpace) together with the image files.

The Figure 2 shows the metadata fields we use to describe the electronic file. We selected these fields with the idea in mind, that at some point we would like to link our digital library to Europeana. Thus we also create metadata that are required by Europeana. The green fields are filled with data from Aleph, the orange fields are created by the person who creates the metadata, and the light blue fields contains data derived from the other fields, but will be filled in by the Europeana office. The gray ones are optional ones.

Table 1. *Metadata fields*

Dublin Core fields	EUROPEANA fields	Custom created fields
dc.title	europaena.provider	AlephSisNr
dc.title.alternative	europaena.type	ScanatDe
dc.contributor.author	europaena.userTag	
dc.contributor	europaena.unstored	
dc.subject	europaena.object	
dc.description	europaena.language	
dc.publisher	europaena.uri	
dc.date.created	europaena.year	
dc.date.issued	europaena.hasObject	
dc.type	europaena.country	
dc.format		
dc.identifier		
dc.source		
dc.language		
dc.relation		
dc.coverage.spacial		
dc.coverage.temporal		
dc.rights		
dcterms.medium		
dcterms.provenance		
dcterms.isPartOf		

3. A Tool for Digital Library: DSpace

With the growing number of documents and interest for the digital documents from the users, the original HTML website (<http://documente.bcucluj.ro/>) could not satisfy anymore the expectations and needs of the users. We evaluated several systems based on functionalities, scalability, technology and cost, and decided to implement for testing the DSpace digital repository. This is an open source system that has great advantages: no license fee, it is robust, thus scalable, knows all the international standards for cataloging and communicating and it is highly customizable. The downside of the system is that it needs an IT professional, that can install and manage the application and it is difficult to connect to the library system already running (Aleph). In these conditions we gave a try to DSpace, and our experience is that it has a simple user interface, that can be set to work in several languages. It has a built-in search engine that can also search full text. The administrative side is simple, and has a built-in workflow, that enables easy document tracking and management. The batch import feature is also important. It makes possible the easy and fast upload of great number of documents. The customization of the interface requires knowledge in HTML/CSS, but it can be set to respond to the users needs. Unfortunately, the connection to the already running ILS,

Aleph, is difficult, and probably this is the reason why, the library will keep looking for the optimal tool to run its digital library on.

The DSpace digital repository system made possible to have easy access to our collections, to search them in full text, to have an accurate inventory, to have a uniform metadata description and to have fast and easy batch import for upload.

Technical aspects of the Dspace system:

- The system

Our DSpace repository is powered by its own server, this being an Intel Xeon dual core central processing unit, model 5110, each core working at a ~1.6 GHz speed, helped by 2 GB system memory. The operating system is the Ubuntu Linux, version 8.04 LTS. The main reasons for choosing this system is its cost (it is free), its power and stability. The LTS version also means that it has long time support.

- Dspace installation

The first step when installing the DSpace software is to make sure that all the prerequisite software is installed. First of all we need a database that will store our data. DSpace supports by default two database software: PostgreSQL, open source database, and the Oracle Database which also has a free version called Express Edition, with some limitations compared to the main edition. We are using the free PostgreSQL on our server. We also need to have the Java JDK installed on our server, as well as Apache Maven and Apache Tomcat. All of them are free to download from the Internet. If we are using this combination we will need the following packages installed: libpg-java (Postgres JDBC driver), ant - optional. All of these can be very easily installed on an Ubuntu system using the command line tools or simply using the Synaptic Package Manager, both available in the operating system repositories.

Having the system prepared, we can download the DSpace software from its homepage: www.dspace.org. Our repository is currently using the 1.6.0 version, but the 1.7.1 newer version is also available. After downloading the DSpace source and setting the right ownerships for the folders, we need to pack the installer using a Maven command, which will download the necessary dependencies. This might take a couple of minutes and it cannot be done without a working Internet connection. When the package is ready and we did not encounter any errors during the build, we can finish installing the software by deploying the files in the designated folder and installing the database. A couple of additional modifications need to be done to the Tomcat server to point it to our DSpace installation.

- The software configuration

The DSpace platform has a poor administration page, thus the configuration needs to be edited manually in the configuration file named: `dspace.cfg`. Although this file has more than two thousand lines of code (including the comments), and the editing needs to be done carefully, with the help of the comments even non-experts can figure it out.

There are hundreds of configuration steps; I will mention a couple of them.

First you need to set the base URL of the server, including the port number. Then you need to configure the access information for the database server, as well as the interface used by the software. Here you can choose between the JSPUI, which is based on Java Server Pages, and XMLUI, which is based on XML pages. We use the JSPUI, but it seems that the newer releases tend to use the XMLUI version more. We translated the interface in Romanian, thus we had to configure the localization as well. The administration pages remained in English

because many terms still lack a good translation into Romanian. This was a problem which took us a while to figure out, that the software needs to be rebuilt completely to add another language. The translation was done by translating the Messages.properties file.

- The interface

As we mentioned earlier, we are using the JSP user interface for our repository. We customized it only with our logo, a couple of additional links, and buttons for easy connection with the most important social sites. We also added a short code for tracking visitors, and collect user statistics. The site is multilingual; with a single click the user can choose the preferred language of the interface: Romanian, Hungarian, English and German (the Romanian version was translated by us). The system was initially in English by default, but currently is in Romanian (<http://dspace.bcucuj.ro>).



The screenshot shows the homepage of the Digital Library of BCU Cluj. At the top left is the logo of Lucian Blaga University of Cluj-Napoca. The main header features the title 'Biblioteca Digitală' and the subtitle 'Biblioteca Centrală Universitară "Lucian Blaga" Cluj-Napoca'. A navigation bar includes 'Pagina principală' and language links for 'Română', 'Magyar', 'English', and 'Deutsch'. The main content area is titled 'Caută în DSpace' and contains a search box with a 'Start' button. Below the search box is a section for 'Categoriile în DSpace' with a list of categories and their item counts: 1. MANUSCRISE - Kéziratok - Manuscripts - Handschriften [36], 2. CĂRTI VECHI ȘI CĂRTI RARE - Régi könyvek és ritka könyvek - Old books and Rare books - Alte Bücher und Seltene Bücher [168], 3. MATERIALE ICONOGRAFICE - Vizualis anyagok - Iconographic material - Ikonographische Material [26], 4. PERIODICE - Időszaki kiadványok - Periodicals - Zeitschriften [10980], 5. CĂRTI ȘI EXTRASE - Könyvek, kivonatok - Books and reprints - Bücher und Auszug [8], 6. BIBLIOGRAFIILE PENTRU CURSURI [21]. On the right side, there is a sidebar with the title 'Biblioteca Digitală a BCU Cluj' and a list of relevant publications for Cluj-Napoca Transilvania, including highly requested publications and scanned documents available upon request. A note at the bottom of the sidebar states that search is performed on metadata and the full text of scanned documents.

Figure 2. Digital Library of BCU Cluj

- The items upload

After the repository is installed and configured, we can start uploading the documents. First we have to create the communities and collections, which will give the digital library's logical structure. For upload we have two options:

1. The software's web interface allows us to upload files using a webform to enter our data. Unfortunately this option allows us to upload only one file at a time, so if we have hundreds of items to add it is not the best way to do it. But for books or other materials which consist of only one or two items, this option is the best;
2. Another way to upload items is to use the command line for batch import. DSpace includes a couple of commands that help us upload multiple files at once. The downside of this upload is that we need to prepare the bundles that will be uploaded ahead. The bundle in our case consists of the item file itself, a license file, and four xml files that contain our metadata.

4. Digital Library as a Library Service

The increasing number of requests from the users for digitization shows a real market and need for the digital libraries. The Romanian government has also created a national strategy for digitizing the national written heritage of Romania. In this context the libraries are invited

to start digitizing their own materials, and give access to them through Internet. Around the world are great examples for successful digital library initiatives, where the library gives access to the electronic materials as a service.

In Romania there are such initiatives, but their existence and growth is limited by the lack of a clear and explicit legal context and a national level coordination and collaboration.

BCU Cluj started its digital library with documents from the special collections and periodicals. At the beginning, the library decided to give only limited access to these materials in order to protect them from batch downloads. With time the policy has been changed, and following some international examples the library gives free access to its materials. Also, it will try to build up a new business model that will be able to support financially the digitization process. It is not yet decided what this model will be, but there are several models that are looked into. Whichever business model will be applied, the success of the service will depend on preparing our online catalogs, request forms and payment methods for such service. All the documents need to be easy searchable in the online catalog. There should be a link to the electronic version of that particular document, or if it is not digitized yet, an easy way to request its digitization. The payment method also needs to be easy and in a format that the users are customized with. The flawless connection of the digital library to the already in-use library software are crucial to have a successful and easily useable service.

5. Conclusions

The process of creating a digital library is a sophisticated and highly technical process. The technical side of the process defines the quality and usability of digital materials. Thus planning and written guidelines are required to succeed in building a repository. The tools used to present, store and archive the materials need to be capable to migrate the digital contents to new formats, in case the used technology becomes obsolete. This will ensure the protection of the digital collection and the invested work.

The legal aspect of creating digital collections should also be clear. The collection developments are highly dependent on copyright and its legal issues. The libraries will face difficulties providing quality service, because a considerable part of the digitization requests from the users are for copyrighted materials. There are international examples for dealing with this issue, hopefully the implementation of such solution will soon be proposed by the Romanian lawmakers as well.

BIBLIOGRAPHY

1. Curta, O. *Metode de informare documentară tradiționale și moderne cu aplicație la Bibliografia istorică a României*. Cluj-Napoca: Argonaut, 2008. 286 p.
2. *Ghid de digitizare*. Available at: <http://www.bibnat.ro/Documente-elaborate-in-cadrul-Comisiei-s159-ro.htm>.
3. *Inventarierea documentelor digitizate în sistemul național de biblioteci*. Available at: <http://www.bibnat.ro/Inventarierea-documentelor-digitizate-s160-ro.htm>.
4. *Politică Publică pentru digitizarea resurselor culturale și realizarea Bibliotecii Digitale a României*. Available at: <http://www.bibnat.ro/Comisia-pentru-digitizarea-patrimoniului-cultural-scris-s136-ro.htm>.
5. *Recomandări Tehnice pentru programele de creare de conținut digital din domeniul cultural*. Available at: <http://www.bibnat.ro/Documente-elaborate-in-cadrul-Comisiei-s159-ro.htm>.