

Mass Digitization of a Monograph-Collection.

F. O. Rump

Library of the University of Veterinary Medicine Hannover Foundation, Hannover, Germany

Friedhelm.Rump@tiho-hannover.de

AIM

The Library of the University of Veterinary Medicine Hannover Foundation was granted funds by the [German Research Foundation \(DFG\)](#) to participate in a nationwide mass digitization project for two years. The aim is to provide an exhaustive repository of monographs in high quality, i. e. resolution and easily navigable layout. In our specific part of the project the digitization of 3790 veterinary monographs published between 1597 and 1890 is being carried out.

METHODS

The requirements stated by the DFG had made it necessary to find a high quality bookscanner, capable of a resolution of true, not interpolated 600 dpi. Product information brochures did not answer all our questions. It was therefore a favourable coincidence that the [CeBIT](#), the world's largest computer fair, is an annual event in Hannover, and we could arrange to get demonstrations of various scanners we had considered for buying. The most convincing impression was made by the [Bookeye3-A2-Scanner](#) made by [Imageware](#):



Figure 1

with the following specifications:

- A2 oversize
- 400 x 600 dpi optical resolution
- rapid scanning
- integrated book cradle
- integrated glass plate
- LED lamps
- laser-assisted profile detection
- 1 GBit network interface

Also, after connecting the PC to the network interface, software for workflow support, post-scan-refinement, archiving and presentation were needed. Upon consultation with the [Center for Retrospective Digitization, Göttingen \(GDZ\)](#) we decided to utilize their solution for workflow control, the open source platform “[Goobi](#)”. This software allows for a controlled production process and subsequent indexing of metadata. It has a production and a presentation level which again are subdivided as shown in figure 2.

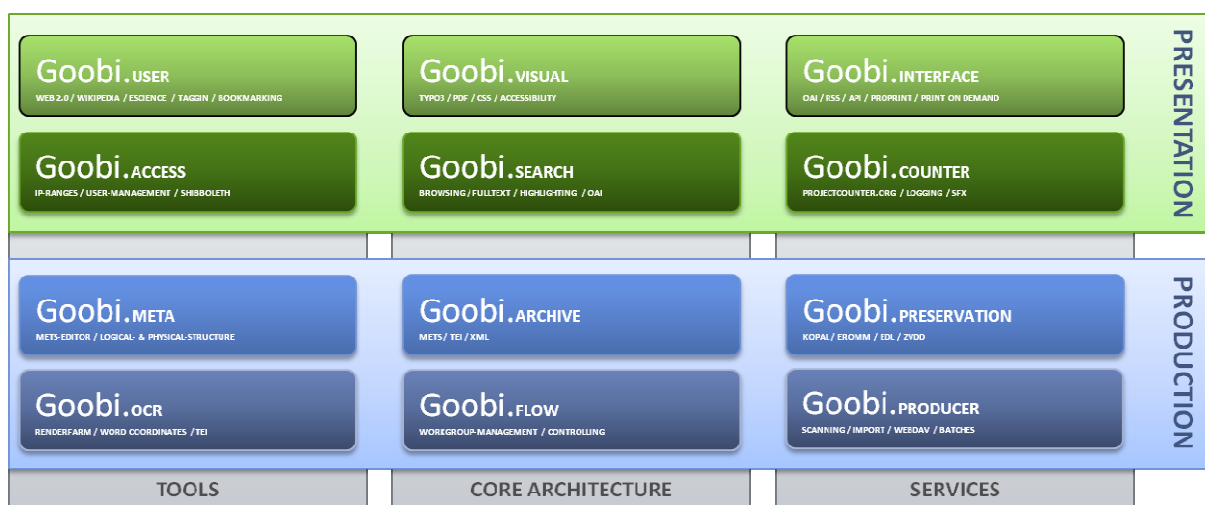


Figure 2

The digitization is done with 600dpi for bitonal originals and 300dpi for grayscale and coloured originals. Digital masterscans are saved in Tiff-format. Safety-copies of the master-files are saved in two different places before post-process-enhancement.

Following this quality is controlled. All scans are screened for readability, correct order of pages and completeness. Then the bitonal scans are post-processed. This comprises a uniform presentation with respect to brightness, contrast and text field. Also de-spreckling is performed, i. e. cleaning of the scans of spots not originally present in the text. The latter is done in batch mode.

The post-processed scans are uploaded onto a server of the [GBV \(Union Catalogue of 7 German States\)](#) which again hosts the OPAC of the library of the Veterinary University Hannover.

Direct access to the digitized documents is facilitated on the Internet via the Goobi presentation module “Goobi.visual” which is based upon the content management system [TYPO3](#). The search engine employed is [lucene](#).

RESULTS

The digitization started on September 5th, 2009 and is developed by a librarian and an informatics specialist. Student workers and specially hired part-time employees do the scanning. To date (April 26th, 2010) 147 books have been scanned. The digitised monographs are accessible on the internet in our Digital Library of Veterinary Medicine at <http://bl460-134.gbv.de/goobi/sammlung/browsen/browsen-titelliste/?DC=tiho.dfg.projekt> database called [Central Index of Digitized Imprints \(ZDDD\)](#) and can be viewed with a specially designed viewer, the [DFG-Viewer](#). The collection of documents can be searched and browsed like in any online catalogue with all the typical functionalities. Special mention should be made of [search functionality](#) which will go into the deeper structure of documents and find chapters and even single plates related to the search term. Another feature is the inclusion in the [library's new acquisitions list](#). Also the upload of a document results in its immediate inclusion in the [GBV](#).

DISCUSSION

This does not comply with the expected progress. The reasons are several. Firstly the DFG did not grant the full manpower applied for, secondly the scan-speed is not the true measure of completion of scans, as they have to be post-processed in various ways:

- Cleaning of the images (de-spreckling)
- Smoothing of the images
- Rescanning, if necessary
- Structur-data filing
- Meta-data cataloguing

CONCLUSIONS

Our digitizations are valuable contributions to the existing collection in other fields. The number of digitized books is quite behind the projected figures. The project therefore should be extended by two years and a firm experienced in batch processing of the raw scans will have to be employed with the second phase of it.

REFERENCES

German Research Foundation: <http://www.dfg.de/en/index.jsp> (last accessed: May 14th, 2010)

CeBIT: http://www.cebit.de/homepage_e (last accessed: May 14th, 2010)

Bookeye Sanner: <http://www.imageware.de/en/systems/book-scanner/bookeye3-a2-color> (last accessed: May 14th, 2010)

Imageware: <http://www.imageware.de/en/> (last accessed: May 14th, 2010)

GDZ: <http://gdz.sub.uni-goettingen.de/index.php?id=2&L=1> (last accessed: May 14th, 2010)

Goobi: <http://www.goobi.org/> (last accessed: May 14th, 2010)

GBV: <http://www.gbv.de/vgm/> (last accessed: May 14th, 2010)

Lucene: <http://lucene.apache.org/java/docs> (last accessed: May 14th, 2010)

TYPO3 <http://typo3.com/> (last accessed: May 14th, 2010)

ZVDD: <http://www.zvdd.de/sammlungen.html#Hannover> (last accessed: May 14th, 2010)

DFG-Viewer: <http://dfg-viewer.de/en/regarding-the-project/> (last accessed: May 14th, 2010)

Goobi search functionality: <http://bl460-134.gbv.de/goobi/en/sammlung/simple-search/> (last accessed: May 14th, 2010)

New Acquisitions List of the Library of the University of Veterinary Medicine Hannover Foundation, Hannover, Germany: <http://biblserv.fh-hannover.de/rss-ext/neuerw-tiho.xml> (last accessed: May 14th, 2010)