

**Reprint**  
Jahrgang 2005, Ausgabe 8

8/2005

# information

## WISSENSCHAFT & PRAXIS



**Barrierefreiheit im Internet**

**Zielgruppen**

**Rechtliche Grundlagen**

**Barrierefreies Webdesign**

**Qualitätsmanagement für Websites**

**Sprachliche Hürden**

**Barrierefreie PDF-Dokumente**

**BITV-Test zur Barrierefreiheit**

**BIENE-Award**

**Praxisbeispiel zur Einführung  
von Barrierefreiheit**

**Realisierung barrierefreier  
Web-Angebote**

**Jahresregister 2005**

# The 18oT Project in Cologne or How do I Process 180,000 Books in Four Months?

Successful cooperation between the hbz, USB Cologne and ZB MED

Astrid Großgarten, Bonn

Imagine yourself as a social sciences student searching for literature on thesis on "The Role of Women in Rural Japan on the Threshold of the 21st Century" or as a dermatologist researching "Methods of Treating Rosacea". In addition to standard works, catalogue research will also bring up hits for monographs. Or will those searching also discover items in collections of essays such as "Japanese woman working" or "Natural Healing Processes for Skin Diseases"? The question can only be answered by taking a look at the table of contents. If the researchers are actually in the library, then this does not pose a problem. However, it is a completely different matter if the research is being carried out online. What is the solution? Drive to the library and look it up there? Order the book on the off-chance using an interlibrary loan? Extending the catalogue information by including a summary of the content would therefore be a very desirable option. These types of requirements will soon become reality – most rapidly for medical professionals, economists and social scientists in scientific libraries in Cologne.

Under the heading *Catalogue Enrichment*, there is a project currently running in Cologne in which the tables of contents more than 180,000 books are being scanned, processed as full text using text recognition and finally fed into the various catalogue systems. The project started on September 1 2005 and processing of the 180,000 book titles initially estimated was completed on schedule before Christmas. At the moment the venture is still running under the working title of 18oT Project (for 180,000 books). The project will be given a final title in a later phase because, as Kathrin Gitmans from the project coordination department of the hbz points out, the project will be continued initially with the two libraries in Cologne and then with three additional partner libraries. This would mean that the 200,000 point would already have been reached by the beginning of 2006 and it would then be necessary to find a new title anyway.

How did such a project, unique in size and scale in the arena of German libraries, come

about? As Tatjana Mrowka, Marketing Manager for the project sponsor hbz (Hochschulbibliothekszenrum [Library Service Centre] of North Rhine Westphalia) explains, the topic of Catalogue Enrichment has been under discussion for quite some time. The subject has been highlighted time and again at various specialist conferences – especially from the point of view of how it might be possible to generate added value for the catalogue (OPAC) by providing additional information about the contents thereby designing a more targeted and successful literature search. Of course, in the omnipresence of Amazon and Google and faced with fastidious Internet users, libraries also consider themselves under severe pressure not to surrender the innovations field entirely to commercial providers.

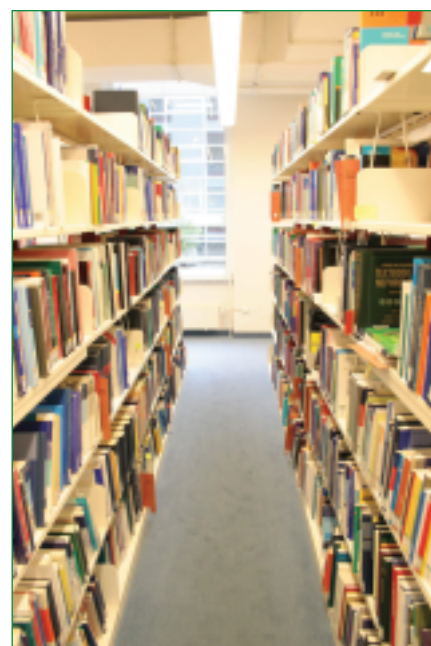
At the conference of the hbz Library Network at the beginning of 2005, the discussion surrounding Catalogue Enrichment was taken up once again and the hbz decided to tackle a project on a larger scale. The project is supported by North Rhine Westphalia's Ministry of Innovation, Science, Research and Technology.

Given its objectives, the hbz is in a perfect position to take the lead management role in a pilot project of this type since; as an establishment providing services and development, it picks up on innovative trends. With a hugely varied range of products, the hbz is a nationally active partner to libraries which supply a total of 2.5 million customers with literature and information.

Some key figures from hbz's product range:

Union Catalogue	13.750.000 title entries 30.200.000 holdings 246 participating libraries
The Digital Library/DigiBib	330 incorporated databases 170 scientific and public libraries
Online interlibrary loan	knapp 300 Bibliotheken

The hbz as the project coordinator agreed the benchmark data in cooperation with



Library "ZB Med" in Cologne

the USB Cologne (University and City Library of Cologne) and the ZB MED (German National Library of Medicine). 60,000 monographs from acquisitions made in the last five years were processed from the field covered by ZB MED and 120,000 titles from the USB Cologne, from purchases made in the last 15 years, were processed in the specialist fields of economics and social sciences. The different time approach for this retrospective catalogue enrichment is owing to the fact that the half-life for medical knowledge is substantially shorter than it is in the field of economics and social sciences.

Both libraries taking part have prominent positions. The ZB MED is the second largest specialist medical library in the world, the largest in fact if based on user numbers. The USB Cologne also caters for supranational and non-university customers with several special subject collections of the DFG (German Research Foundation), extensive special holdings and the most significant old stock in North Rhine Westphalia. The titles were chosen from the holding of economics and social sciences which accounts for approximately a third of the total hol-

Reference year for information is 2004

	USB Cologne	ZB MED
<b>Holdings of books and volumes of journals</b>	3.6 millionen	1.3 millionen
<b>Current journal titles</b>	approx. 10.000	approx. 8.000
<b>Users per year</b>	over 45.000	14.815
<b>Lendings on site</b>	1.170.000	147.000
<b>Interlibrary loans made</b>	approx. 60.000	
<b>Document delivery and interlibrary loans</b>		610.000

ding. To a large extent both collections also cover relevant titles from the whole European languages area in addition to German and Anglo-American research literature.

The scientific relevance of the specific literature from the pilot libraries selected and the fact that the data of 246 libraries is combined in the hzb's Union Catalogue clearly shows how expedient it is to place enriched title data in such a huge and heavily frequented data pool as the hzb media server. 13 million holdings with approximately 30 million copy data are currently available to those participating in the network.

If one extrapolates the 180,000 books selected for the project, to calculate the pages to be scanned; then the total number of pages is 720,000, where medicine accounts for approximately 240,000 and economics and social sciences account for approximately 480,000 pages. If this is apportioned over 4 months it corresponds to approximately 2,000 books or 7,000 pages per day that had to be processed.

It was clear from the beginning to the hzb and the participating libraries, that the libraries' own staff would not be able cope

with such a volume, in addition to the normal everyday routine work of a library. The search was then on for a service provider who could meet the specifications in terms of time, organisation and cost. These specifications stated that no additional hardware and software should be purchased for the project, the existing infrastructure (Medea3 environment of the hzb Library Network and MyBib servers of the USB Cologne) had to be harnessed, no recapturing of media data was to take place and of course there should be no interference in the normal operation of the library.

The ImageWare Components Company from Bonn, well-known manufacturer of Bookeye® book scanners and MyBib delivery systems, won the bid as the organisation having the most convincing solution.

Organisation of the project was agreed jointly:

- The participating libraries would provide network connections and work rooms for the service providers and would grant access for the service provider's staff to readings rooms and stacks
- The service provider would provide the technical equipment
- The service provider would provide a server for job processing, which would in turn be connected to the hzb server by way of an interface
- The hzb would configure the server jointly with the service provider
- In both libraries there would be permanent contacts for technical problems, quality assurance and professional questions
- The quality standards would be agreed on and compliance with them checked.

For ImageWare Components, the assignment to the project signified a huge challenge since it was entering partially virgin territory from a technical and entrepreneurial aspect due to the complexity of the requirements. Managing director Rolf Rasche considered there were far more opportunities than risks since he was "not often offered the chance to subject MyBib to such a weighty test on its own doorstep. Furthermore, it will allow us to prove the stability and quality of our systems in so doing."

In order to avoid any unpleasant surprises at the beginning of the project, it was preceded by a pilot phase in July and August

2005. First, 500 volumes for each participating library were processed. The results were checked by the contacts of the libraries and the hzb with respect to scan quality and accuracy of text recognition. The result was better than satisfactory as was the optimised MyBib-eDoc server, which seamlessly handles the administration, control and tracking of orders and recording of operating data. The hzb was confident about the test phase, "Quality assurance is of prime importance for the hzb. If these test results had not been satisfactory, we would have discontinued the project," said Hans Ollig, manager of the hzb.

Following the successful test phase, the project went into production operation on September 1, 2005 and has run to the complete satisfaction of all those stakeholders.

Since any project member can test the working results on the MyBib system at any time by accessing the Web, the project group is always kept up to date and it has been possible to limit harmonization meetings to once a month. In this case, the agenda covers assessment of the quality of the work and discussion of special cases, for which binding specifications and solutions for production operation have to be found quickly. Some examples of these are handwritten additions in the tables of contents, multi-language tables and those with formulae or Arabic and Chinese characters which are not recognised or are incorrectly recognised by the text recognition program.

So how is the work actually organised in the participating libraries? For the first project phase eight scanning stations were set up in the USB Cologne and four in the ZB MED. The equipment provided per workplace, in addition to furniture and book trolleys, is a Bookeye®-GS400 with ergonomic scan pad and barcode gun. Annegret Johann of ImageWare, responsible for production, explains the process, "The employees bring the books to the scanning station. There each book is registered first of all using the barcode gun. In the following work steps, the pages of the table of contents are scanned and cleaned of irrelevant information. Text recognition is then carried out. Every employee is urged to check the correctness of the result". If errors have been overlooked, they become apparent later in the multi-stage quality assurance process. Then the jobs in the system are marked with a query and appear on subsequent job lists for reprocessing.

The employees who carry out the scanning work only see the user interfaces of the BCS-2® scanning software used. The extremely complex MyBib system in the background, which controls the entire workflow and is actually responsible for enabling seamless job tracking from the local library systems to the hzb server, is only accessi-



Bookeye® Scanstation





Workstations, ZB Med Cologne

ble to relevant people commissioned with the project. Using MyBib, Annegret Johann generates what are known as book collecting lists, which are distributed to the scanner operators. These lists are job notes and specify the books to be processed for the employees. Every book is accounted for on the list by way of shelf mark, media number and title. The media numbers occupy a key role in the complex set of data categories. They are the unique identification for a book and are attached to every medium in the form of a barcode label. The barcode identifies the book vis-à-vis MyBib and in turn, this creates a link to the network ID of the hbz.

The hbz transferred the data obtained to its media server at the turn of the year. Conversion of the full text generated by text recognition will in this case allow indexing of the data via the search engine used in the hbz which is based on FAST technology. The enriched data will be usable for all network participants and will avoid duplicated work.

Every library customer will profit from the fact that very soon the entries for more than 200,000 monographs in the catalogues will be expanded to include the table of contents. It will then be possible for users to examine them as a digital image. The result will be a significant improvement when carrying out literature research, which will make local and interlibrary loans much easier. Perhaps as a side effect, document supply orders will be made for articles which would otherwise continue to remain on the shelves unstudied in anthologies.

For Dr. Christiane Süverkrüp and Claudia Dembek, who are looking after the project for their libraries in terms of content, the advantage is primarily an improvement in the working economy of library customers because the supplemented catalogue data will reduce the number of incorrect loans and increase the frequency with which tit-

les are used. "As a result of Catalogue Enrichment and availability of the data online, the existing holdings will also be made more public and the user service will also be extended especially for the supraregional customer groups," according to Claudia Dembek. In Dr. Süverkrüp's experience, the ability to research media online in particular has very clear effects on the frequency with which they are used. And what more can a library wish for with regard to its considerable investment in building up and maintaining holdings than that its media are used in a targeted manner, intensively and frequently? Perhaps in the future, users will leave Google and at the next opportunity choose to research in the catalogues of the ZB MED and the USB Cologne instead.

As the project ran without a hitch and completely to schedule, all those taking part began as early as November to look into the future and check the possibility of a continuation, in which catalogue enrichment could be operated not only retrospectively but also for new acquisitions. In order to cover 80%, if possible, of new acquisitions in the library network, suitable partner libraries were sought to enable seamless continuation of the project in a second phase.

The second project phase has been underway since the beginning of the year: in addition to the ZB Med and the USB Cologne, the project group now includes the ULB Düsseldorf, the UB Paderborn and the ULB Bonn, and continue to be co-ordinated by the hbz. Scanning stations based on the organisational model trialled in Cologne will also be installed in the premises of the new participants and connections will be made to the central MyBib-eDoc server which monitors the whole business process.

According to Hans Ollig, manager of the hbz, "A project of this magnitude is trend-setting and unique in Germany to date."

## AUTHOR

### Astrid Großgarten

Patrimonium Transcriptum  
Verlags GmbH  
Oppenhoffstraße 16  
D-53111 Bonn  
E-Mail: grossgarten@ptverlag.de  
Telefon: 00 49-2 28-3772 685



**Hochschulbibliothekszenentrum des  
Landes Nordrhein-Westfalen**  
Jülicher Straße 6  
D - 50674 Köln  
Telefon: 0049-221-400 75 - 0  
Telefax: 0049-221-400 75 - 180  
E-Mail: info-hbz@hbz-nrw.de  
<http://www.hbz-nrw.de>



**ImageWare  
Components GmbH**  
Am Hofgarten 20  
D - 53113 Bonn  
Telefon: 0049- 228-96985 - 0  
Telefax: 0049- 228-96985 - 84  
E-Mail: info@imageware.de  
<http://www.imageware.de>