INCUNABULA ON THE INTERNET – A DIGITISATION PROJECT OF THE BAYERISCHE STAATSBIBLIOTHEK

Dr. Marianne Dörr
Bayerische Staatsbibliothek
Ludwigstr. 16
D-80539 München
Tel.: 089/28638-2600, Fax 089/28638-2672
E-mail: doerr@vd17.bsb.badw-muenchen.de

Background and scope of the project

Incunabula, the oldest documents of printing history, are valuable, rare, often unique and mostly only accessible to the general public in occasional exhibitions. The Bayerische Staatsbibliothek owns the second largest collection of incunabula in the world after the British Library. Thanks to funding from the project Sammlung Deutscher Drucke sponsored by the Volkswagen-Stiftung, the BSB was able to add a number of valuable copies to its collection in recent years. The background for the digitisation project was the intention of all libraries participating in the project Sammlung Deutscher Drucke to make some part of their holdings accessible to the public in digital form. Projects for the presentation of collections of incunabula - predominantly on microfiche - exist already. The Bayerische Staatsbibliothek did not intend to compete with such enterprises or duplicate them and therefore concentrated on digitising the "printed book illustration of the fifteenth century" (illustrated incunabula) within the DFG-programme Retrospective Digitisation of Library Holdings. The illustrations are exclusively woodcuts of varying quality and detail. About 30% of the prints have been coloured in by hand. In the age of Dürer, the technique of woodcutting became more subtle: woodcuts show a more differentiated use of line drawing and shading, the technique of printing from several plates (line and tone plate) increased the potential for design, and woodcuts thus did not require later colouring.

The choice of illustrations followed a double intention:
* extension of indexing to an area previously neglected by libraries:
the area of images and, more specifically, image subjects
* creation of a tool for art historical research
* creation of a service for a broader public: illustrations can provide immediate access, whereas texts printed in incunables, even if they are in German, are not easily comprehensible for a general public.

The aim of digitisation was on the one hand to create master files for reproduction of the prints in order to be able to fulfil regular requests for reproductions by using the digitised image, thus avoiding further strain on the original. From the very beginning, it was also intended to provide access via the internet - in reduced quality, taking into account the standard capacities for connection and transmission.

Procedure and technical aspects of digitisation

Basically, there was a choice between two options:
Digitisation from microfilm / Ektachrome or Digitisation from the original.

No films of sufficient quality existed. In addition, the problem of potentially double shift of colour from the original to the photographic image and from the photographic image to the digital image had to be considered especially for coloured originals. After some discussion, we decided to undertake direct digitisation, which was not yet practised in libraries and museums on a large scale when the project started (1997); exceptions were the project Vaticana and Lutherhalle Wittenberg, both funded by IBM, which however did not lend themselves as models due to the framework conditions, which involved very considerable strain on the original.

As camera, we chose the Progres3012 by JenOptik (formerly Kontron), which had already been tested by a contractor who provides photographic services in routine work, not merely for the creation of individual images.

The technical parameters of the camera and the digitisation are the following:

resolution capacity of the CCD array: 3,500 x 4,500 pixels = 384 dpi on A4 scale.
36Bit colour depth, which was however reduced to 24 Bit during software processing by the scanning and image processing programme used (PhotoShop).
For originals of different sizes, a range of exchangeable optical lenses were available.
We made full use of the maximum scanning resolution, but varied colour depth according to black and white and coloured originals. Black and white woodcuts were digitised with 8 Bit, i.e. 256 scales of grey; the maximum colour depth of 36/24 Bit was only used for coloured woodcuts.

IR and UV-filtered halogen lamps were used for lighting.
The required exposure time was between 30 and 45 seconds, due to internal chip technology (movement of the CCD array). This is more than during photographic work, but was in the end regarded as acceptable by the book restorers from our Institute for Restoration of Books and Manuscripts.

A further problem when digitising books is the support for books. For the Vaticana project, the originals had been dismembered – this was unthinkable for this project, if only for reasons of costs and work involved.
In co-operation with the contractor, we decided to digitise individual pages and to use a so-called book cradle by default, which allows a book to be opened at an angle of 90°. It was necessary to employ sheets of glass to keep the original open, but no massive pressure was applied to the page or the book block. The book cradle proved suitable for the protection of bound books during this project: Even books from which several hundred scans were taken showed no damage to the spine or binding.

Digitisation was carried out in the Bayerische Staatsbibliothek (in the exhibition room for treasures in the department of manuscripts, which ensured proper height of ceiling, lack of vibration, and climatisation) with special staff who had received appropriate training for this project, but with equipment provided by the contractor (camera, lighting, support for originals, i.e. book cradle, computer workstation for scanning, second computer workstation for quality control and CD-writer).

Calibration of equipment and colour management

The camera was calibrated on the Macbeth colour chart. Every morning, before starting work, the staff carried out an new calibration by software. A colour chart with integrated scale was digitised with the original. When digitisation was done (early 1997), procedures of colour management with the creation of ICC profiles were not yet available as a standard. The lighting intensity was partly reduced in the course of the work in order to protect the originals. This caused problems and necessitated occasional follow-up editing of the images: the colour values of digitised colour chart were measured with a spectrometer, the values found were compared with the values of the digitised chart and corrected by software. The in terms of colouring very variable results of photographic reproductions of the image files by a range of contractors who themselves had no access

to the original showed further how important it is to come to an agreement on colour measuring and colour space, in order to achieve reproductions from the digitised image with correct colours.

Results of digitisation

Within a period of three months, 6,424 illustrations from 76 incunabula were digitised, ca. 30% of the images were coloured and were scanned in colour. The average number of shots was ca. 150 when digitising in scales of grey (at the end, 180 shots were achieved), and ca. 80 when digitising in colour. The reason for this difference was not so much the longer time required for scanning in colour, but the time required for follow-up editing: focussing of the image etc., which made intensive use of computing power. The sizes of files created averaged ca. 45 Megabyte for colour images (14 for grey scales). These are factors which need to be considered when undertaking high-quality digitisation. Some cameras which are currently available and achieve an even higher resolution (e.g. Picturegate by Anagramm with ca. 7,500 x 7,500 pixels) produce file sizes of 200 MB and more. When planning a project, consideration needs to be given to the adequate relation between the intention (purpose of images) and parameters for digitisation and thus to questions of economy and costs.

The estimated time depends also on the handling of the material: careful and protective treatment of valuable and ancient books and images requires time which needs to be invested in such enterprises.

Indexing of the illustrations

Only the illustrations were digitised, but access to the bibliographical data of the complete edition is required. For this, previous cataloguing work could be used: The machine-readable catalogue records of the Incunabula Short Title Catalogue (ISTC) were downloaded as the basis of the bibliographical description and loaded into a database. They were translated into German and extended by variant forms of names and titles as used in German incunabula catalogues, i.e. the incunabula catalogue of the Bayerische Staatsbibliothek. In addition, an abstract of the contents of the text and/or brief information about the place of the edition in the history of printing were composed.

The incunabula selected (the main criterion for selection was richness in or significance of illustrations) give a survey of the types of books which were illustrated in the fifteenth century. The main genres represented are secular literature in vernacular languages (prose versions of courtly romances, heroic epics, travel books), religious literature (Bible, liturgical books like plenaries, collections of saints legends), herbals and medical treatises, chronicles. The only book digitised as a whole is the Augsburg edition of the Nuremberg Chronicle written by Hartmann Schedel; here, digitisation included mere text pages.

The main work was the indexing of image subjects. With a view on international standards, which naturally have gained importance in the age of the internet and the potentially world-wide reception, we decided to undertake indexing based on the IconClass classification developed for art historical projects. IC attempts to organise potentially all objects and subjects of art in a system of 9 main classes (Religion and Magic; Nature; Human Being, Man in General; Society, Civilisation, Culture; Abstract Ideas and Concepts; History; Bible; Literature; Classical Mythology and Ancient History). This can lead to very detailed, complicated classifications. For our project, preference was given to the searchable verbal description rather than the numerical notation, which is however entered as well.

Indexing aims at providing one suitable description of the depicted scene as a whole rather than analysing its constituent parts (as it is done for projects which undertake the indexing of every individual object depicted in an image, like items of clothing, shoes, for the sake of a study of material culture).

The indexing is done using four categories: Up to two classifications from IconClass can be given. There is a field for captions transcribed from the original and another freetext field in which the indexer can enter terms not covered by the classification which ought to be searchable. An example

for this are personal names: the classification describes a scene of homage, but the incunabula text explicitly refers to emperor Maximilian.

This means that searches are integrated both on a more general level for iconographic comparison as well as more specifically. In addition, keywords from the caption can be entered in modern German form and are thus suitable for retrieval.

This kind of indexing may not allow for all imaginable queries, but offers a range of access points which can be provided within a pragmatically acceptable amount of work.

Presentation and internet service

The service aimed at offering three different types of access to the user:

- navigation in the sequence of images as printed in the individual incunable (browsing between illustrations)
- search for subjects of images, e.g. portraits of rulers, topographical views, from different printed editions
- navigation from the systematic classification to the images associated with it.

Regarding the third option it has to be said that not the complete classification was integrated into the presentation. The reason for this was that too many notations would have lacked corresponding images. The notations entered during indexing were merely traced to their root (the relevant main group), which makes it possible to link the notation to a higher level (ignoring intermediate levels) like nature or magic. Thus, the user who descends from nature only encounters notations which are in use, e.g. linked to images.

Realisation was done in several steps:
The bibliographical description and the description of subjects of images were entered into a relational database, in which the necessary connections were made between the list of images, the title data, the classification and the descriptions of individual images.
Following this, the data were exported with automatic SGML tagging according to self-designed DTDs. Two separate SGML-files were created: the SGML document "Incunabula", which contains all title data of the incunabula and the descriptions connected to individual images. In addition, there is a SGML-encoded file for the classification, which contains the main groups, the notations used and links to the relevant images of the incunabula titles.
The user can choose between the two forms at the beginning.
Naturally, the data model is more complicated than for individual images which need not to be kept in a particular sequence. In a database system, navigation by browsing between illustrations could not have been offered.
For retrieval, several search interfaces were designed which demonstrate to the user the options for combination of search terms in the form.

By digitising, indexing and processing of the incunabula illustrations, a complex tool for search and presentation of the genre was created.

At the moment, additional technical features which are in principal desirable for digitisation and the access to images cannot be offered: e.g. zooming or cutting out of details. Presentation is geared towards a use via internet browsers without further software. Other forms of processing or separate editing of individual title with CD-ROM software product which allow such kinds of reproduction can be imagined.