

The Digital Beethoven House: Collection Building, Presentation and Maintenance

Marion Borowski, Lars Bröcker

NetMedia
Fraunhofer Institute for Media Communication
Schloss Birlinghoven
D-53754 Sankt Augustin
marion.borowski@imk.fraunhofer.de
lars.broecker@imk.fraunhofer.de

Abstract: An increasing number of cultural institutions accept the challenge to build digital libraries. They start the process of digitising and describing their treasures in order to present them to cultural audiences as digital media. So did the Beethoven House in Bonn who owns the largest and most valuable collection of the composer Ludwig van Beethoven. In cooperation with the Fraunhofer Institute for Media Communication they realised the project *The Digital Beethoven House* (2001-2004). Original objects, manuscripts and music of Beethoven play the leading role in the digital collection, the online presence and the virtual environment. In the following paper we share our experiences in building, presenting and maintaining an awesome collection with the digital library community. Digital libraries are not an end in itself but the basis for semantic interrelations with similar repositories. We present the potential of further developments to provide additional services to the scientific community as well as the general public.

1 Introduction

The Digital Beethoven House is a cooperative effort of the Beethoven House Society in Bonn, Germany and the Fraunhofer Institute for Media Communication. The aims of this project were threefold: firstly, preservation and presentation of the cultural heritage of Beethoven. Secondly, enabling the Society to sustain the site on its own, and thirdly laying a basis for future research efforts in the fields of computer science as well as in musicology. These goals have been achieved by continually building the main three parts of the project: online presence, a collection database, and a virtual installation. The internet pages (see www.beethoven-haus-bonn.de) provide users with general information of the Beethoven House and give them access to the largest collection of manuscripts of the famous composer Ludwig van Beethoven. 26,000 sheets of music scores, letters, first prints, pictures and other original objects have been digitised in high quality. They and other data sources from the museum as well as from third parties have been imported into a digital library to build the collection repository.

The combination of a web environment on one hand with a virtual environment on the other hand is a speciality of *The Digital Beethoven House*.

Users might as well access digital treasures via the internet as visit the real Beethoven House in Bonn to experience a virtual environment with Beethoven's music. A virtual world allows users a synesthetic experience of classical music in a 3 D environment. Music of Beethoven's opera Fidelio is visualised and invites users to interact with the scene via four pillars providing interaction tools. The visualisation of Beethoven's Fidelio is the first virtual opera of the classical repertoire.

The goal of this paper is to share some of our main experiences with the scientific community to help others implementing similar projects and to further enhance semantic connections between them. We set our focus on describing the digital library and the online presence, we do not explain the virtual environment in further details. This paper is structured as follows: Section 2 gives an overview over the process of building and presenting the digital library. Section 3 describes the potential of further developments. The paper ends with the conclusion in section 4.

2 Preservation and Presentation of the Collection

The main goal of the project was the preservation of the cultural heritage Beethoven's in the form of the documents collected by the Beethoven House Society. In order to allow scientists to examine these documents for their research while simultaneously keeping these in parts irreplaceable documents safe (some of them are worth more than a million euro), a digital library solution for these documents was proposed. In a first step the requirements for the digital library were collected, concerning the digitisation of the documents, the different sources of metadata and the planned services of the library. The following sections give an overview of the steps necessary for the implementation of the digital library.

2.1 Digitisation and Postprocessing of the Images

The first major work package concerned the digitisation of the collection: the documents needed to be digitised in full colour in very high resolution, so that they can be used in printing as well as on the internet [Bo01][Gr02]. Four additional formats have been derived from the original high resolution images. These formats are: thumbnails for fast preview, medium-sized images for presentation alongside metadata on one page, larger-sized images, that are presented as sole content on a page and a compressed, original-sized format fit for print. Additionally, provisions had to be taken in order to secure the copyright of the images.

Visible watermarks are used for the middle and larger sized images. The first three image formats are available via the internet. Invisible watermarks are used for the largest format, which can be either viewed on public workstations at the Beethoven House in Bonn or purchased for publications. The invisible watermark allows monitoring of the web for illegal occurrences of these images. More details on watermarks can be found in [Bi00].

2.2 Combination of Metadata from Different Sources

The digital library of the project collects metadata from several different sources. The digital library includes image and music files from different providers, as well as transcripts of letters of Beethoven and his contemporaries, taken with permission from a letter edition published by the Henle-Verlag (see www.henle.de), each with their corresponding metadata. To import these metadata, an XML language has been developed. This language is also used for update purposes.

The Beethoven House Society files data about its collection in a legacy system at their offices. The digital library does not contain this full set of metadata; only a relevant subset is imported, modeled after the Dublin Core Metadata Standard [Du] with domain-specific additions. Updates to the digital library are performed at fixed intervals. This way, the valuable data of the Society stays encapsulated and the editors can work in the environment they are used to.

The letter transcripts are an especially valuable addition to the digital library, since they are enriched with highly detailed markup (in SGML). This markup cross references people, places or works and other letters mentioned in the letters to corresponding background information or the pages of the letter or work in question. The digital library contains only transcripts of originals present in the collection. The integration of this rich source of information on Beethoven and his contemporaries has been the subject of a diploma thesis written at the IMK [Gr03].

2.3 Implementation and Presentation of the Digital Library

The preceding sections already mentioned some of the aspects that had to be kept in mind in designing the architecture of the digital library. Access from the intranet as well as the internet with differing levels of service and updates of the database in regular intervals being the main issues. A market survey on digital library systems was conducted for the project. This proved to be a complex task, especially for the specialised requirements of a museum. The findings of this survey can be found in detail in [B01].

The project partners first intended to use the product IBM Content Manager thanks to an interesting offer of the University of Bonn. Due to severe problems with the stability of the software on the intended server architecture (SUN Solaris) the partners reluctantly decided not to take on that offer.

Since the remaining commercial systems were too expensive, the digital library was being built using open source software running on a dedicated Linux server. The metadata is stored in a relational database (MAX DB, the former, now open-sourced, SAP DB); the image data resides in a RAID connected to the server. Backup of the images is accomplished via tape. Java servlets are used in order to access objects of the digital library. They are served by Apache Tomcat from the Apache Foundation, Tomcat being the reference implementation of the Java servlet specification by Sun Microsystems.

Java servlets are used to deliver data to the web CMS for presentation purposes. The use of a web CMS ensures that the site stays alive and topical even after the technical partner has left the project. Changes in the data as well as in the design of the site can be implemented by people inside of the Beethoven House, especially for smaller institutions an important expense factor.

Digital exchange with other institutions in the field of musicology is seen to gain importance in the near future, so the digital library will allow export of metadata using the Open Archive Initiative (OAI) protocol. In order to gain experience in this new form of data exchange, the Society has chosen to become a data provider in this protocol, i.e. the digital library defines a subset of the metadata in the collection and allows transfer of this subset to a so-called service provider.

3 Potential for further Developments

The digital library of *The Digital Beethoven House* is the home for the largest digital collection of original documents concerning the work and life of Ludwig van Beethoven. The architecture of the library has been especially designed to provide opportunities for additional services to the scientific community as well as the general public, that go beyond the scope of this project.

There are already ideas for such additions to the system, some of these touching research topics like the semantic web. Some of these ideas include:

- creation of topic maps (see www.topicmaps.org) for subject areas of the digital library, thus showing users the interrelations of documents inside of the collection.
- connection with digital libraries of other institutions, preferably ones with topic maps of their own, to further the understanding of interrelations of Beethoven's work with e.g. that of Mozart's.
- creation of a repository/topic map containing information on contemporaries of Beethoven, important events in his time and information on Beethoven not covered in the collection.
- creation of a Beethoven course on his music and life, which teaches the Beethoven friend, the pupil as well as the scientist. The aim is to provide a comprehensive learning platform which is motivating, intuitive, fun and simple to operate so that the user may immerse quickly into the content.

There are far more ideas, but these are the most promising ones in the eyes of the project partners. The first three items are especially promising, regarding the opportunities for musicology as well as the realisation of the vision of the semantic web.

4 Conclusion

Projects of similar magnitude as *The Digital Beethoven House* will - sooner or later - come across many of the same questions and difficulties that our project team has had to face. With the speed, in which new technologies evolve, architectures for digital libraries need to be very flexible and open to allow inclusion of new services and modules.

Sustainability is another important question in this kind of project: is the partner able to maintain the digital library on his own, after project ends? There are some very sad examples of digital libraries that broke down after the technical partner pulled out of the project.

These two concepts, an open architecture and the sustainability of the results, guided the team of *The Digital Beethoven House*. Our team is interested in sharing experiences with other scientists involved in building actual digital libraries for and with scientists from other domains. Feel free to contact us with questions or suggestions.

References

- [B01] M. Bogen, M. Borowski, S. Heisterkamp, D. Strecker: Requirements and Architecture for Administration and Maintenance of Digital Libraries. In: *Museums and the Web 2001*, Bearman, David [Ed], 2001
- [Bi00] T. Bissel, M. Bogen, V. Hadamschek, C. Riemann: Protecting a Museums' StockthroughWatermarks. In: *Museums and the Web 2000* / Bearman, David [Ed], 2000
- [Bo98] Bogen, Manfred; Borowski, Marion: *The Digital Beethoven House*. In: *WebNet 1998 - World Conference on the WWW, Internet and Intranet* / AACE [Eds], 1998
- [Bo01] M. Bogen, C. Bonkowski, M. Borowski, J. Löffler: Digitizing a Cultural Heritage - The Key Issue for Preservation and Electronic Publishing. In: *WebNet 2000* / Gordon, Davies [Eds], 2001
- [Br01] L. Bröcker, M. Bogen, A.B. Cremers: Bridging the Semantic Gap in Content-Based Image Retrieval Systems. In: *Fifth International Conference on Information Visualisation* / Davies, Gordon [Eds], IEEE Computer Society, 2001
- [Du] Dublin Core Metadata Initiative: Dublin Core Metadata Element Set, Version 1.1: Reference, Description. Downloadable at <http://dublincore.org/documents/dces/>
- [Gr02] Grigat, Friederike: Vom Dokument zur Datei; Farbdigitalisierung von Handschriften, Erstausgaben und Bildern als Grundlage eines digitalen Beethoven-Archivs. *Forum Musikbibliothek*, 23. Jahrgang 2002/1, ISSN 0173-5187
- [Gr03] David Greiff Mehrwert-Integration in Digitalen Bibliotheken durch semantische Auszeichnung. Diplomarbeit at the Rheinische Friedrich-Wilhelms-Universität Bonn, Germany, 2003.
- [In02] Internet Software Consortium: Internet Domain Survey, July 2002. <http://www.isc.org/ds/WWW-200207/index.html>
- [Ko03] Inke Kolb, Dieter Strecker: Topicality versus Prettiness: How to enable Easy ContentManagement during Web Development. In: *Museums and the Web 2003* / Bearman, David [Ed], 2003
- [Ly00] P. Lyman, H.R. Varian, et al: How much Info? University of California at Berkeley, School of Information Management and Systems, <http://www.sims.berkeley.edu/research/projects/how-much-info/internet.html> 2000