Austrian Literature Online (ALO):
How Electronic Libraries, Cultural Heritage and Service Provision for Print Disabled People work together

Klaus MIESENBERGER[1], A Min TJOA[2]

[1]University of Linz, *i³s³*, Altenbergerstraße 69, A-4040 Linz, Austria
[2]Vienna University of Technology, Favoritenstr. 9-11/188, A-1040 Wien, Austria

Abstract: Digitisation is one key objective at the beginning of the information society. More and more efforts are spent to make documents usable on computers and via the Internet. This also addresses a major need of print disabled people. Access to documents in digital format is the prerequisite to be able to apply assisitve technology to get access to documents at the Human Computer Interface (HCI).

This paper presents the initiative "Austrian Literature Online" and how this project addresses the needs of print disabled people. A virtual library system was developed for presentation and preservation purposes of old literature. This system and the documents handled are designed according to the needs of print disabled people. Research was done on a more efficient and cost effective workflow management in digitisation. Till now copyright free literature has been made available in accessible formats. The system is also developed further to provide blind and visually handicapped students a convenient access to study literature.

Key Words: ALO, Electronic libraries, Cultural heritage, Print disabled, Digitisation

1. Introduction

There are three estimations summarising the need for efforts in the field of transcription and library services for print disabled people:
· only 1-2 % of all publications are made available to print disabled persons
· most documents which are transcribed in accessible formats are used only 1-3 times
· an important number of publications and documents are produced two or more times at different locations (by private persons or professional organisations)

These estimations show that a) more documents should be made available, b) information on available documents should be distributed in a better way and c) better co-ordination of the production is necessary. Direct access to electronic documents (electronic publishing), converting digital sources into suitable formats and computer supported transcription of documents (computer supported publishing)[3] give the target group access to what everybody takes for granted. These developments need accompanying efforts for a more efficient co-operation between print disabled users, publishers, authors and service providers possible.

Services for print disabled people therefore have been and get more closely related to digitisation and metadata enrichment for enhanced usability:

"Ante Gutenberg" blind people did not have problems with printed documents. Not a lot were used. Speech was the major tool for communication and dissemination of knowledge. The Gutenberg revolution brought forward print disabilities. The more books were part of everyday life, reading skills, access to books and education based on documents became a prerequisite for taking part in almost any part of society. Therefore not only more possibilities to choose, to store and to use knowledge were brought forward; people became forced to use these possibilities, which established print disabilities as a social problem. "Post Gutenberg", since the beginning of the 80s, print disabled people get access to documents more independently and efficiently by using computers. Nevertheless only 1-3 % of all documents are accessible which still puts print disabled people at a disadvantage.[5]

Computers and especially the Internet led to a similar demand for access to documents in digital format to increase usability and efficiency. Not only newly published documents but also older documents should be available in digital formats. Activities of making cultural heritage available in virtual environments is a question of prestige for libraries, archives, museums and thereby the cultural identity. The changing culture of using documents, the increasing information overflow and competition ask for a new presentation of important cultural information to be still recognised. It is obvious that older literature and other cultural heritage playing an important role of reflection and identification for mankind have to be part of this new information society. Library services and their preservation function therefore more and more involve digitisation, digital services and digital preservation. A growing number of initiatives around the world push forward digitisation of documents like books, newspapers and magazines for cultural heritage, virtual libraries, efficient access but also for an efficient document management in complex organisations:

· Bibliotheca universalis — http://www.ddb.de/gabriel/bibliotheca-universalis/bibuniv.htm
· Gallica — http://gallica.bnf.fr/
· Setis — http://setis.library.usyd.edu.au/ozlit/
· American Memory — http://memory.loc.gov/ammem/
· MOA — Making of America — http://www.umdl. umich.edu/moa/
· Dioscorides — http://www.ucm.es/BUCM/diosc/
· Victorian Women Writers Project — http://www.indiana.edu/~letrs/vwwp/index.html.

Although primarily oriented towards cultural heritage and preservation, digitisation could be seen as the technical, economical, cultural and social convergence in the information society. "Design for all" of documents could help to address all these targets.

2. ALO — Austrian Literature Online

German literature is very rarely represented in these activities. These countries defining their identity and understanding out of a long history and the related cultural heritage are challenged by digitisation. This has been the starting point for the ALO project[1][3][6]. The striking goal of ALO is the digitisation of the 1000 most important books of Austrian literature ranging form 1800 to 1930. This shows that up to now only literature without copyright protection is taken into account in the ALO virtual

library.

The consortium (University Library of Graz, Institute for German Literature at the University of Innsbruck and the Institute Information Systems for Blind People at the University of Linz, i^3s^3) did research on how a virtual library of Austrian literature could be set up. Most of the functionalities needed for the ALO project could be found in other projects but none of them provides all aspects needed (e.g., OCR of old fonts called "Fraktur", high quality full text and bibliographic search functionalities, "Book on demand" and especially accessibility for print disabled people. ALO did intensive studies for an efficient scanning, OCR, mark up, accessible user interface design, print on demand and document management. A prototype with the most important functionalities of a virtual library was set up. Based on a list of the most important books of Austrian literature collected by the University of Innsbruck, the University Library in Graz scanned selected first editions. In addition the literature magazine *Iris* (1848 to 1849) was digitised and integrated into the virtual library system[1)3)]. i^3s^3, which runs a service, research and education centre for IT for blind and visually handicapped students all over Austria, set up the prototype and did OCR with a focus on accessibility and usability. The system is also used as a trial portal for services related to the preparation of study materials for print disabled people. ALO is open source, so it could be used and should be developed further by any other organisations interested in such applications.

3. The ALO system

The ALO software gives access to digital surrogates of documents in three different formats following the most important international standards for preservation, metadata enrichment and long term storage [1, p. 44ff]:

3.1. Formats provided by ALO
· *Facsimile:*

Documents are scanned with a resolution of 400 to 600 dpi, black and white in TIFF format. (Only if necessary documents are stored in colour). Users get access to
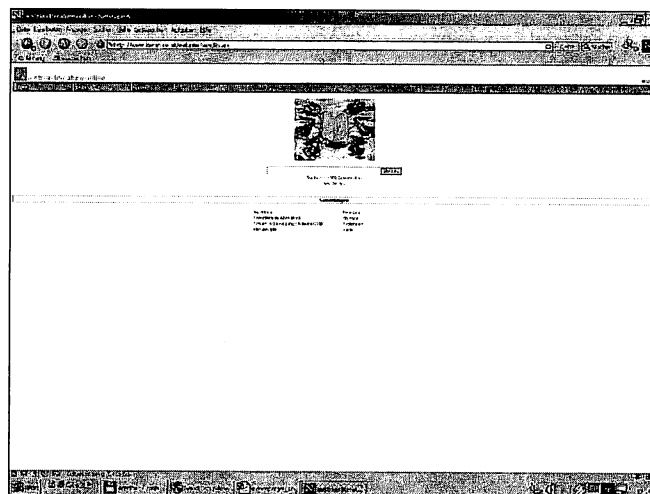

Fig. 1   http://www.literature.at

the image of the original including the typography, layout, images and graphical components via the Internet. The masters are converted into GIF format when presented following the MOA II guidelines. Before scanning the whole book, images are tested and adjusted according to the needs of OCR systems (brightness, cleaning of age-related soiling,) to enable a fast and easy production of full text.

· *Uncorrected Full Text:*

Using the prepared high quality facsimiles the production of an uncorrected full text can be done easily for a "quick and dirty" access for print disabled people and other purposes. Dependent on the quality of the book and especially the fonts used the text will have recognition errors. In almost any case this text serves the needs of full text search for professional and fast usage.

· *Corrected Full Text (on demand):*

The correction of full text and mark up of books is done according to guidelines and recommendations of the Making of America II project (MOA II) of the Library of Congress, Preservation Reformatting Division. In this correction process the metadata for structuring a document is done using the MOA II DTD. [7] The guidelines of the Text Encoding Initiative (TEI, TEIxLite-DTD) are used to include descriptive, administrative and structural metadata for usage and management[8].

Following established standards should help to ensure inter-library exchange and long-term preservation. Besides increasing the usability in general ALO offers a high level of usability for literature studies, research (e.g., dictionaries of old language) and language education. Last but not least ALO offers access to a wide range of important literature, which has not been accessible for print disabled people till today[3].
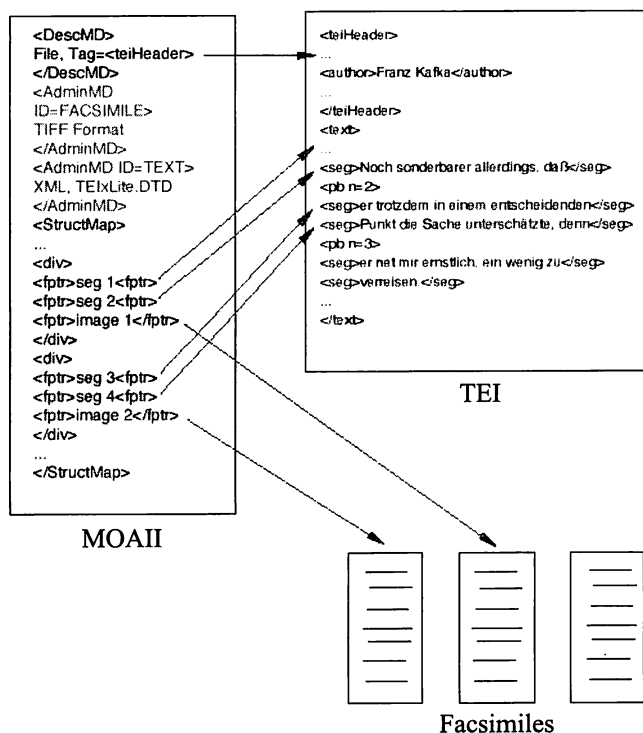

Fig. 2   ALO Formats

3.2. Functionalities of the virtual library

The ALO prototype offers basic functionalities to handle digital documents: add/delete book, access to facsimiles, access to full text, access to meta data, links between full text and facsimiles, access to whole documents, ... [1, p.50ff; 3]

The *database* is implemented in Java which gives access to tools to handle XML files and interactions with several www servers. A Java tool handles requests using Remote Method Invocation (RMI) and sends back the query results in XML files. First the MOAII file is loaded and coded in XML providing further links to facsimiles and the TEI encoded text. MOAII, TEI and facsimiles files are stored in a file system. The database handles a list of names and links to the physical elements of documents in a file system which slows the system down but is recommended for preservation and management.

The *user interface* provides three tools to search for books: *full text search, metadata search and list of books*. For all presentation functionalities servlets are employed using XQL-queries and providing XML files and XSL stylesheets. Pages can be accessed with a certain page number or forward and back buttons also enabling browsing a document. Users can switch between text and facsimiles which is of particular importance for old fonts, which are often hard to read. Facsimiles are dynamically transcribed in GIF format for display purposes.

For reading the whole document or defined parts of a document (e.g., 1; 5; or/and 7-16) a PDF or text file can be downloaded. The servlet composes the download and sends it to the browser for reading using a Plug in or saving.

Book-on-Demand allows producing and getting access to cheap reprints in a format and with a binding serving the needs of reading literature. No longer do a high number of books have to be produced and stored. Books are produced when an order comes in and are at clients' side in a few days. Therefore the economical risks for publishers and authors are minimised. With a link to one provider of Book on Demand those books can be ordered.

Due to the "design for all" principle only a few special things could be mentioned about accessibility and usability for print disabled people. People should get access to correct and well-structured text of documents. This cannot be afforded for all documents but the costs to produce full text versions "on demand" are reduced due to the availability of high quality facsimiles. "Braille Print on Demand" should be integrated into the workflow and established with Braille print service providers. As another basic support, visually handicapped people can use different sizes of page

Table 1   Costs of Digitisation

Task	Cost/unit	remarks
Scanning	0,3 €/page	flat bed scanner or camera
OCR	0,3 €/page	no proofreading
Proof reading	0,5 €/page	one time
e-book	40 €/book	
Book on Demand	138 €/book 0,3 €/page	fee for 3 years print master production

images when browsing a book at the WWW interface.

The digitised books are integrated into the Austrian cataloguing system ALEPH and — if accessible — in the special catalogue of literature in accessible formats TESTLAB. TESTLAB is a unique collection of library and service provider catalogues in the German speaking area, which is further elaborated to a document management system[5)6)].

4. Cost analysis

ALO did cost analysis studies for digitisation and WWW presentation of documents to provide profound information for decisions on further digitisation activities. A sample of 200 old books was analysed according to criteria like year/location of publication, sort of literature, number of pages, title pages, empty pages, size of page/print area, characters per line, words per line and characters per cm^2, fonts used, age related soiling, and specialities. The data collected were used to define the costs for digitisation outlined in Table 1. The costs for the digitisation of one book with about 190 pages on average are about 480 €. Without book on demand the costs are about 250 € and without proof reading about 150 € per book. Of course these figures tend to go down considerably over time.

5. Future work

During the ALO project the Meta[e] project was outlined, submitted and accepted by IST in the 5th Framework Programme of the European Commission[7)]. Meta[e] develops tools to support or automate the process of metadata generation in the digitisation process. Again, a major focus is given on the "design for all" principle to increase the usability of digitisation efforts. The Meta[e] project also works on OCR for "Fraktur". Finally a search engine for SGML/XML documents will be developed.

ALO will use the results of this project. Besides that, the prototype software is developed further at the moment in the ALO II project. Till now ALO had not set up a working environment integrating all steps of digitisation, preservation and presentation of documents. ALOII therefore will integrate its components and workflow into one user interface. Besides software, training and counselling for digitisation will be made available.

Concerning the usage of the system for a virtual library for print disabled users, the interests and the co-operation of stakeholders of copyright protected documents (publishers and authors) have to be considered. Management tools therefore have to
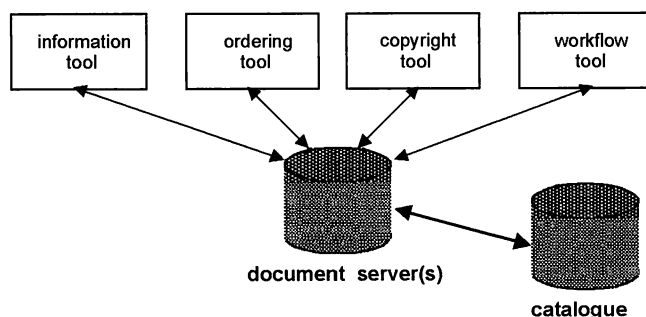

Fig. 3   Extensions of the ALO System

32

provide e-commerce functionalities like ordering, payment, secure access and delivery[4].

Another EU-IST funded project, BOOKS2YOU, allows the expansion and use of ALO software to support or automate inter-library loan, which is very cost intensive and slow. Documents should be digitised instead of sending them around and putting them in danger of being damaged. This could also help to give print disabled a better access to libraries. Other applications for smaller archives and collections are planned. Further on it is planned to expand the system to handle digital surrogates of other objects in museums and archives. This again is a challenge to make these surrogates accessible to print disabled people.

Last but not least it should be noticed that increasing the efficiency and usability of the tools should help by having people with disabilities be employed to fulfil this work in a cost effective way. The ALO consortium started two projects of this kind.

The final ALO system should enable supporting the service provision for print disabled students in all steps and tasks. Therefore the work now starts to concentrate on the components as described in the following:

Information tool: Since the resources for transcription of literature are limited, it is necessary to share the information of which titles are already available or in process. Standard features already implemented are having a catalogue as such a tool. The interfaces to the catalogue have to be made more attractive, should offer the possibility of advertising new documents, should help to get sponsors and support.

Ordering tool: Entering requests for documents, which are already available as well as requests for transcription, conversion, print, etc. should be facilitated as comfortably as possible by using the data, which are in the special catalogue, or which are imported from standard catalogues.

Copyright tool: The transcription of study literature into electronic format raises several copyright issues. As pointed out in another paper[2], the co-operation with publishers and authors depends on possibilities of insuring their rights and interest. Several possibilities for a secure document distribution system, as described in[2] and as already on their way to be implemented[9], have to be evaluated and integrated into the system. The copyright has to hinder every misuse of data that are at the service provider's side (internally) as well and especially at the client's side. All actors in the process have to have the possibility of controlling the process according to their rights and interests.

Workflow management tool: The management of the whole process requires several formal and informal agreements, contracts, management and control documents (e.g., pricing, billing) and actions by the different people involved. Clients, publishers, libraries, services providers should agree on a standardised procedure, which should be supported by a workflow management tool. Possibilities and rights to control certain parts of the process by clients, publishers, service providers, sponsors and supporters are needed.

This rough structure should be elaborated in the next months by using UML — a challenge for us and our blind colleagues in the software engineering process.

6. Acknowledgement

and the "Meta"" project funded by the European Commission, IST.

References

1) Egger, A.: Electronic Publishing als Chance für die Erhaltung Österreichischen Kulturerbes — am Beispiel der Projektes ALO, Diploma Thesis, IICM, Technical University of Graz, Austria, 2000

2) Burger, M., Miesenberger K.: Secure Distribution System for Publications in Electronic Form, in: Klaus, J., Auff, E., Kremser, W., Zagler, W., L. (Eds): Interdisciplinary Aspects on Computers Helping People with Special Needs; 5th International Conference, ICCHP '96, Oldenbourg, Wien, München 1996

3) Habitzel, Hofinger, N., K., Miesenberger, K., Muhlberger, S., Sperl, B.: Vorbereitungsprojekt und Machbarkeitsstudie für eine virtuelle Bibliothek der österreichischen Literatur: Austrian Literature Online — Österreichische Literatur Online (ALO) — Endbericht und Machbarkeitsstudie, bm:bwk und Innsbruck-Graz-Linz 2000

4) Miesenberger, K., Köttstorfer, M., Stöger, B.: From Library Catalogue to Document Management for Print Disabled Users, in: Buhler, Ch., Knops, H. (ed.): Assistive Technology on the Threshold of the New Millenium, S. 600ff, IOS Press, Amsterdam 1999

5) Miesenberger, K., Burger, F., Wagner, G.: Integration of Gutenberg on Line Books into the Austrian TESTLAB CATALOGUE, Colloque organisé par INSERM: Telematics in the Education of the Visually Handicapped, Paris 1998; catalogue TESTLAB: http://testlab.aib.uni-linz.ac.at

6) http://meta-e.uibk.ac.at/

7) Making of America, 2000; project homepage: http://moa.umdl.umich.edu

8) Text Encoding Initiative, 2000; project homepage: http://www.uic.edu/orgs/tei/

9) WESLEY, T.; MILLNS, I.: Secure Document Delivery for Blind and Partially Sighted People: the SEDODEL Project, in: Telematics in the education of the visually handicapped, Paris 1998; http://www.braillenet.jussieu.fr/ntevh/secure.htm#full