

Producing eBooks on Demand - A European Library Network

Silvia Gstrein and Günter Mühlberger, University of Innsbruck

Introduction

Following the creation of the European Commission's digital strategy in recent years, European libraries have been meticulously digitizing parts of their cultural heritage. At the same time they have witnessed increasing demand from researchers and readers for access to digital resources, particularly historical books. This was the impetus behind the eBooks on Demand (EOD) network, which provides a trans-European digital document delivery service for readers all over the world. Currently the EOD network consists of over 30 libraries from 12 European countries. Since 2007 between four and five thousand e-books have been generated, delivered to users from over 30 countries worldwide and subsequently made available to the public through the repositories of the participating libraries. User reactions have been very positive and more libraries are interested in offering the service. This chapter examines the service in general, the libraries involved, the experience from both user and library points of view, and future developments.

Background

In 2005, the European Commission launched i2010, a policy framework for the information society and media, which aimed to promote 'the positive contribution that information and communication technologies (ICT) can make to the economy, society and personal quality of life'. Recently the i2010 strategy has been superseded by a new initiative, the Digital Agenda for Europe 2010-2020 which aims to 'maximize the benefit of the Digital Revolution for all'. In line with both of these initiatives, European libraries and other cultural institutions are gradually digitizing and making available their cultural heritage to a wider public. However, it will still take decades for the majority of books, journals and other library material to become available in digital form due to the sheer volume of source material, despite the digitization efforts of Google Books.

In an ideal world, all of the works in a collection would be considered for digitization. In practice, however, this is rarely feasible and difficult choices must be made. According to the MINERVA Good Practice Handbook (Minerva Working Party 6, 2004) the following basic criteria can be used to identify material which should be prioritized:

- material which would otherwise be unavailable or of limited availability
- very popular material which would be more widely and easily accessed digitally
- delicate material which would be preserved by making digital versions available as an alternative
- material which is suitable for the physical process of digitization
- material which is appropriate for online viewing
- material which has a theme or subject matter identified as of interest to researchers or the general public

- material which can be legally digitized, taking into account copyright and intellectual property rights issues
- material which has not already been made available digitally
- material which can feasibly be digitized within the funds available.

In reality, the approach chosen is often determined by the availability and source of funding, in combination with the institution's digitization policy. This method of deciding which material to digitize only very rarely acknowledges user interests other than the influence of popularity, as mentioned above. This raises the following questions: how can the process take account of the individual users' needs, and what happens to the material not covered by any of the criteria, for example who will digitize books in minor languages or those from smaller or more specialist collections and institutions?

All of these considerations form the starting point for the eBooks on Demand network: an electronic document delivery service initiated and part-funded by the user. The service meets the need for a model of ongoing rather than project-led digitization, and is a co-ordinated European initiative. The selection of material for digitization works bottom-up rather than top-down - the individual reader begins the process by requesting that a certain book be digitized, and then contributes financially towards the scanning process. Subsequent costs for storage, access, long term preservation and migration are not funded by the reader, and will not be addressed in detail in this chapter.

The eBooks on Demand service

It was recognized that it would be difficult to maintain a user-orientated library service such as EOD within the context of the usual methods of mass digitization, as additional resources for order management, customer communication, and payment procedures would be required. Therefore, a structure allowing for efficient processing of orders was designed for the service. This employs elements of both centralization and decentralization, where some processes such as the optical character recognition (OCR) of scanned images and online payment for completed orders are hosted centrally but other processes such as the scanning itself are managed by the individual libraries which are members of the network.

From October 2006 to June 2008, a pilot Digitization on Demand project was carried out, with funding from the European Union eTEN programme (Europa, 2006). This involved thirteen national and university libraries from eight European countries: Austria, Denmark, Estonia, Germany, Hungary, Portugal, Slovakia and Slovenia. After the pilot project was evaluated, the self-sustaining eBooks on Demand network was set up by the 13 founding libraries. Since then EOD has attracted further associated members (Mühlberger and Gstrein 2009).

Briefly, EOD functions as follows. The starting point is the online catalogue of the participating library. The EOD button is visible on the catalogue entry for all items available for digitization. Presently these are books which are in the public domain and not yet digitized

by any of the network libraries. Any user interested in obtaining a copy of a book simply clicks on the EOD button to initiate the process (see Figure 1.3.1). From 2011, a search engine covering as many as possible of the online catalogues of participating libraries will go online, which will make it easier to search from one single point. The link will be made available on the network website at <http://books2ebooks.eu> .

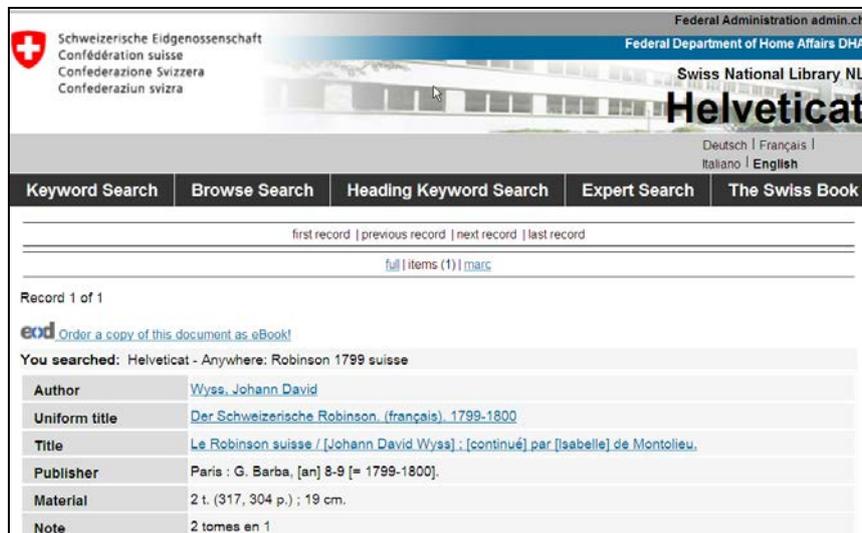


Figure 1.3.1 Example of EOD button in Swiss National Library online catalogue

Next, the user fills out the order form, where they can choose the type of delivery that they require. This may take the form of a download or can be delivered via a data carrier such as DVD (see Figure 1.3.2). After this, the user is invited to follow the process of the order on their personal tracking page.



Figure 1.3.2 Example of order form after clicking on the EOD button

The relevant library receives the order in real time, then scans the requested book and transfers the scanned images via FTP to the central server located at Innsbruck University, which is the network co-ordinator. Each library manages and processes its own orders using a central database, the Order Data Manager, accessible via web browser (see Figure 1.3.3).

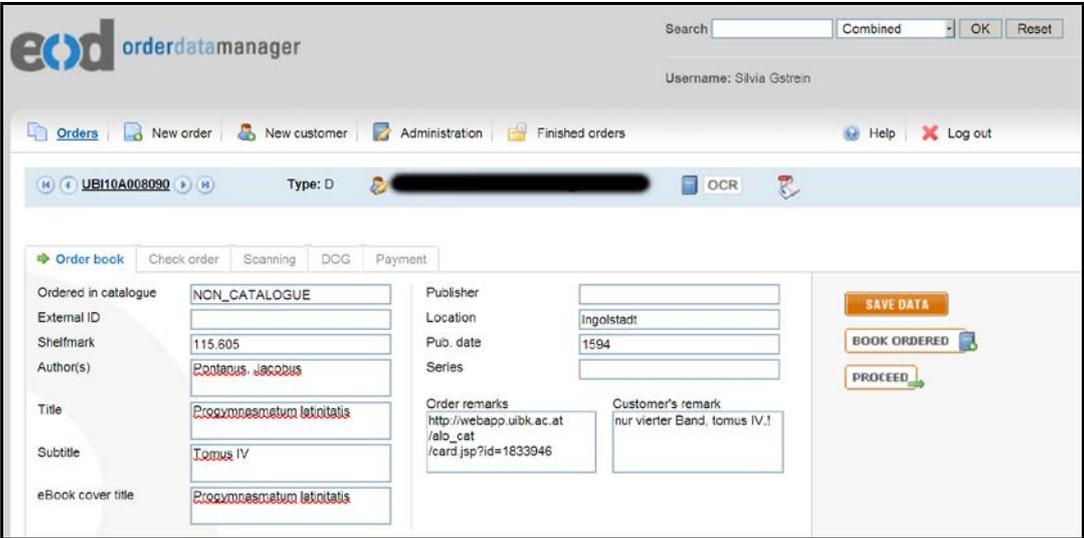


Figure 1.3.3 The Order Data Manager database

The workflow for the creation of the digitized book is also tracked via the Order Data Manager (see below). After completing the payment process, which supports online credit card payment, the user downloads the PDF from their personal tracking page. After a period of time, the library adds the digitized book to its digital library or repository and thus makes it available to the general public.

The Order Data Manager includes another core component: the Digital Object Generator, which is used to create e-books by applying OCR and generating the PDF file (see Figure 1.3.4).

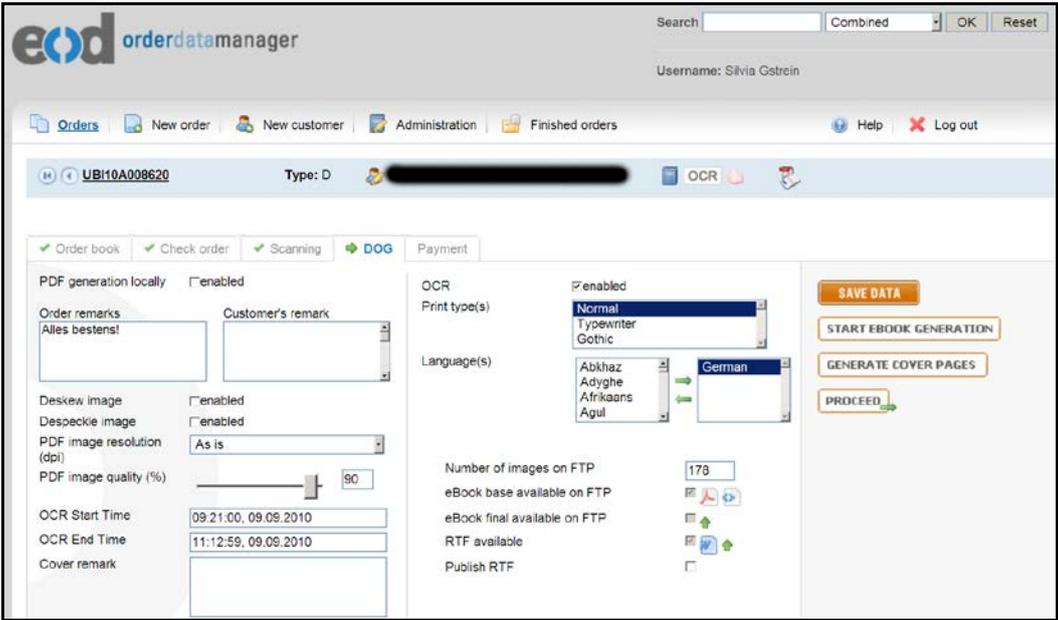


Figure 1.3.4 The Digital Object Generator

The Digital Object Generator supports OCR of all common typefaces from the 18th to the 20th centuries, including Roman and Gothic fonts as well as Cyrillic, Hebrew and Greek scripts, all of which are important when digitizing a wide range of historic texts. It is also possible to generate a cover automatically, including the relevant metadata and displaying the logo of the delivering library. The creation of output files such as PDF, RTF and OCR XML, as well as the generation of the streaming link for downloading, is rendered during the process. The PDF e-book delivered to the user consists of two layers: the first layer containing the scanned images and the second layer containing the automated recognized text.

Multilingual access is a very important aspect of a pan-European service. All generic content, such as the text on the order form, is offered in all the national languages of the participating libraries, currently eleven languages in total. All other library specific text is provided in the respective national language as well as in English, as a minimum. According to its needs and objectives, any participating institution can adapt and customize different content, such as the information linked from the order form and tracking page as well as automatically generated email text.

The EOD network - participating libraries

Currently almost 30 libraries from 12 European countries offer the service, ranging from Portugal in the far west to Estonia in the East. For up to date information about network members see www.books2ebooks.eu/partner. Several types of libraries are currently involved, including national libraries, university and state libraries as well as research and academy of science libraries.

<p>Austria</p> <ul style="list-style-type: none"> • University of Innsbruck, Library (co-ordinator) • University Libraries of Graz and Vienna, Library of the Medical University of Vienna • Vienna City Library • St. Pölten Diocese Archive 	<p>Czech Republic</p> <ul style="list-style-type: none"> • Moravian Library in Brno • Research Library in Olomouc • Library of the Academy of Sciences in Prague • National Technical Library
<p>Germany</p> <ul style="list-style-type: none"> • University Libraries of Regensburg, Greifswald, Leipzig and Humboldt-Universität zu Berlin • Bavarian State Library • Saxon State and University Library (Dresden) 	<p>Hungary</p> <ul style="list-style-type: none"> • National Széchényi Library of Hungary • Library of the Hungarian Academy of Sciences
<p>Denmark</p> <ul style="list-style-type: none"> • The Royal Library 	<p>Portugal</p> <ul style="list-style-type: none"> • National Library
<p>Estonia</p> <ul style="list-style-type: none"> • National Library • University Library of Tartu 	<p>Slovakia</p> <ul style="list-style-type: none"> • University Library in Bratislava • Slovak Academy of Sciences
<p>France</p> <ul style="list-style-type: none"> • Medical and Dental Academic Library of Paris 	<p>Slovenia</p> <ul style="list-style-type: none"> • National and University Library
<p>Sweden</p> <ul style="list-style-type: none"> • Umeå University Library 	<p>Switzerland</p> <ul style="list-style-type: none"> • The Swiss National Library

Table 1 Survey of EOD libraries (September 2010)

Efforts are also being made to include archives and other cultural institutions in the network, with the aim of providing digitization on demand far beyond the world of books in the narrow sense. For example, the St. Pölten Diocese Archive (Austria) began to implement the service in 2010.

User experience

To date, between four and five thousand books have been digitized and delivered to customers through the EOD - books that would not otherwise have been digitized. Over a million scanned pages have been produced, and nearly 2,500 readers have used the service worldwide. The three libraries which process the most digitized texts receive one request per working day on average and deliver 250 to 350 books per year each. The average delivery time is one week.

The average price of an e-book ordered via EOD is 50 EUR. This is calculated by starting with a minimum fee (on average 10 EUR) and adding a cost per page scanned (generally 0.15 to 0.30 EUR). At the moment the pricing system is fairly heterogeneous, as each library sets its own prices depending on national and local policies and conditions, however harmonization of prices would be desirable in the future.

Revenue from the customer only partly funds digitization, covering the personnel costs for the scanning process and administering the service, but not contributing to the initial purchase of equipment or the costs of data storage and preservation. Added to this, the customer-orientated service provided by EOD is more costly on a per-book basis than mass digitization, which places huge emphasis on efficiency and economies of scale. However, given that the master files remain with the library and are later made freely accessible to the public, the overall benefits greatly exceed those to the individual who initiated the process, so it is felt that the customer should not be required to pay the total real cost of digitization.

In 2008 a survey of EOD customers was carried out to collect feedback on the service, and to find out what users thought about the pricing policy. The findings were that 30% of customers felt the price was high or very high, but overall value for money was still found to be acceptable by the majority.

The overwhelming majority of users are either researchers or readers requiring e-books for professional or scientific use (over 60%). The second largest category of users (16%) is made up of book collectors and readers from special interest groups such as amateur historians, collectors, or ethnographers (see Figure 1.3.5).

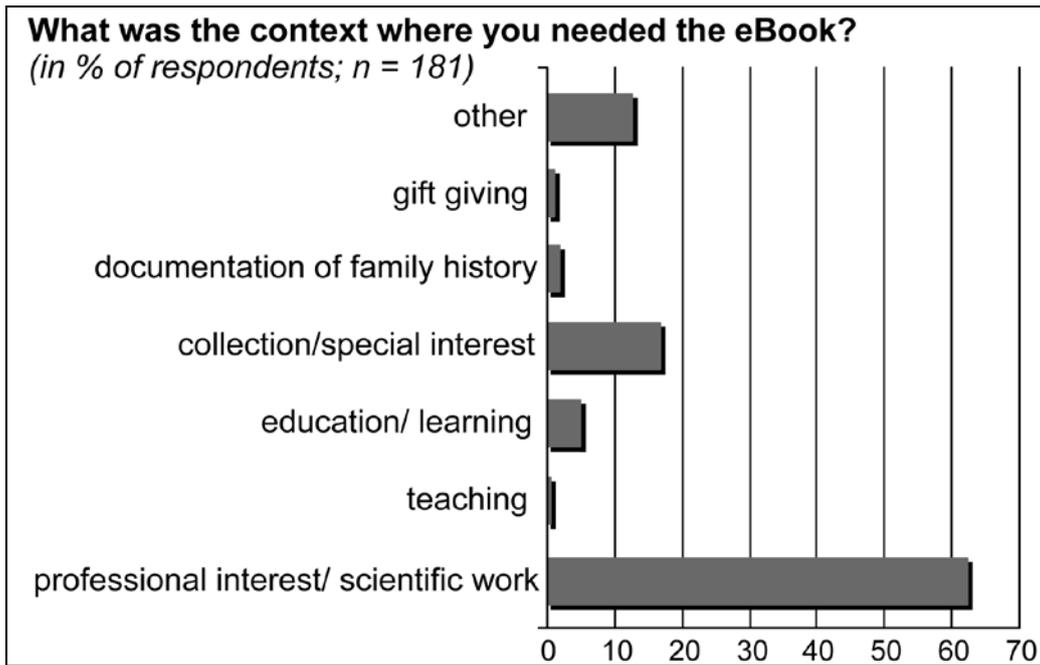


Figure 1.3.5 Users' area of interest

Asked why they had opted for the EOD service, almost half of the interviewees answered that without EOD the book would have been 'impossible or difficult to access'. This shows that EOD has achieved one of its main goals, that of being a practical alternative to accessing printed books in a library or archive collection (see Figure 1.3.6).

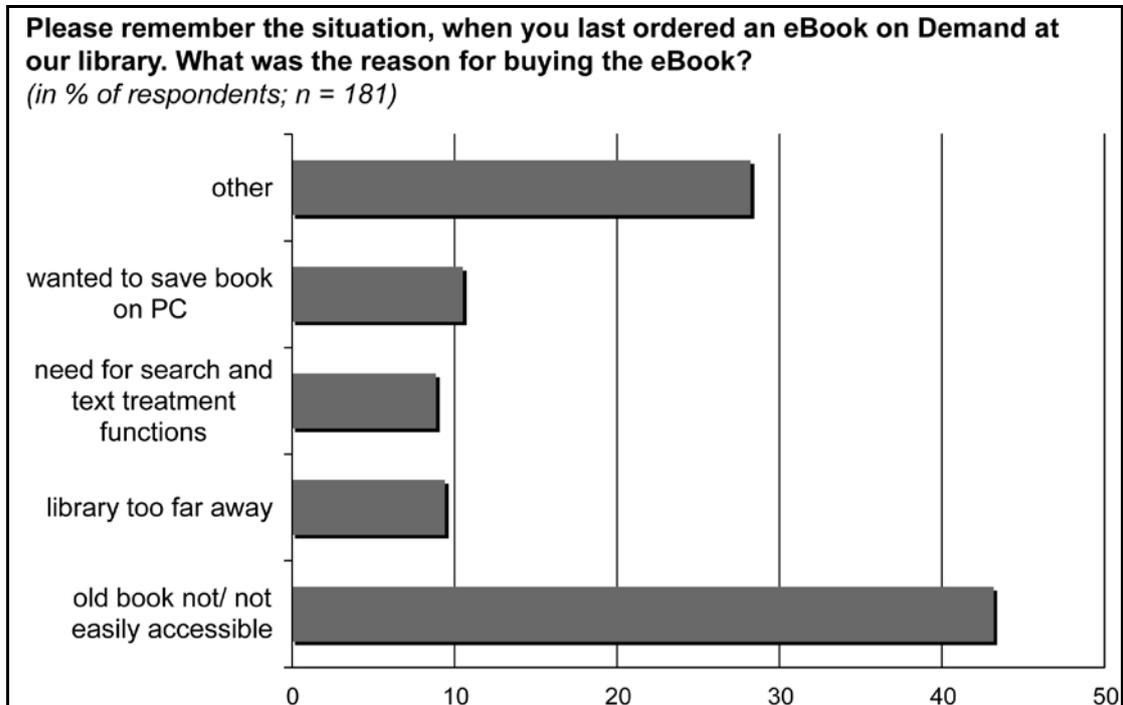


Figure 1.3.6 Reasons for ordering

In the same survey 60% of customers said that they usually print out selected pages or even the whole book. Thus, there is an obvious demand for 're-materialization' of digital material. In response to this, the EOD network has recently begun to offer 'reprints on demand': the

provision of historical books in paper format supplementing the digital file. This option can be selected during the ordering process, in addition to requesting the PDF. From the perspective of the participating library, the service is straightforward. Each library carries out the scanning of the images and provides some additional metadata. All other processes, such as image enhancement and the creation of pre-press PDFs and related files, are carried out by the central co-ordinator at the University of Innsbruck library.

Infrastructure and other organizational aspects

Staff and departments involved

Depending on the structure of the individual library, EOD may be run by a variety of different departments including those focusing on digitization, document delivery and user services, for example. In almost all cases local library staff carry out the digitization, only rarely is scanning outsourced.

Equipment

Due to the wide range of different institutions offering the service, many different types of hardware and software are used, depending on the local infrastructure already in place. Scanning is carried out largely using overhead scanners, but automatic or semi-automatic book scanners and photo cameras are also used in some libraries.

The member libraries have agreed common minimum digitization standards in order to guarantee good quality reproduction, with a focus on optimizing OCR results. The accuracy of OCR not only depends on the quality of the scan, but also on the print type and language of the original, for example texts published prior to 1850 can cause very severe difficulties for accurate conversion from image to text. The European project IMPACT tackled these difficulties through comprehensive research into OCR and language technology in relation to the processing and retrieval of historical documents, as well as by promoting the sharing of expertise to build digitization capacity across Europe.

Use of fees

Most of the revenue from the fees paid for digitizing books is put directly back into the digitization services of the member libraries and only rarely into general library funds. As mentioned above, the fee covers the scanning and order handling costs, but not the infrastructure or costs of subsequently providing the e-book to the general public.

Life after digitization

After the books have been scanned and successfully delivered to customers, the EOD network libraries make them publicly available by uploading them into their own repository after a certain period (in most cases about two months), and then providing a link to the full text from the records in their catalogue.

Currently, the option of providing a central repository for all books digitized via EOD is being discussed within the network, and this would be of particular benefit for those libraries without a repository of their own. Additionally, from 2011 onwards the metadata of books digitized via EOD will be made available through Europeana, the library of digital objects created by cultural institutions throughout Europe. This is because EOD acts as the content aggregator for its member libraries for the European Commission project EuropeanaConnect, which aims to share best practice in the area of digital libraries.

Currently there is no centralized provision for digital preservation; each library makes its own arrangements for the long term care of the scanned images and resulting files.

Holdings offered

Currently, only books in the public domain that have not previously been digitized are offered through the EOD digital document supply service. Within the new 'Reprints on Demand' service, works already digitized will also be offered in print format. In the future, the network may also be able to offer out of print books. However, in most European countries the copyright status of each individual item still needs to be checked, making the digitization of out of print books which are still in copyright time-consuming and costly.

To avoid duplication of effort, a tool to check whether a book has already been digitized by a member of the network is in development. However, this is still a work in progress, particularly as various European libraries use different identifiers in their catalogues for the same book title, making it difficult to match results accurately. As a result, identifying whether a specific book has already been digitized by another member library is still labour intensive.

Communication

Within the network, meetings are organized at least twice a year. There is a general meeting in spring where overarching objectives and the methodology for collaboration are discussed. This is followed by a session in autumn which focuses more on practical and technical issues such as scanning workflow, hardware issues and so on.

Between the main meetings, communication takes place via email and an internal wiki. If necessary, face to face meetings between the co-ordinator and partner libraries are held. On occasion distinctive national features need to be discussed, in which case it has been highly effective to bring together all of the library representatives from the same country, with or without the network co-ordinator.

Difficulties of running a pan European service

Despite the various methods of communication described above, difficulties are still encountered at times. The various local, regional and national contexts have to be taken into account, and problems tend to arise in areas where there are a variety of practices in use. In a decentralized network this is particularly the case in the area of pricing, which varies

according to both national and individual library policies. It can also be seen in the need to handle different currencies, languages and terms and conditions.

Future developments

Recently a new EOD project began, within the framework of the European Commission Culture Program 2007-2013. The project focuses on the larger scale involvement of three main target groups: participating libraries, individual users and the general public. Twenty libraries from ten European countries are concentrating on the following objectives: firstly to enlarge the EOD network by the inclusion of additional European libraries, especially those from countries not yet represented; secondly to take EOD as a best practice model for other Europe-wide networks of this kind and to train stakeholders (libraries, museums, or other cultural organizations) to run a multinational cultural service based on state-of-the-art information technologies; and finally to support intercultural dialogue among readers and users of historical books with the help of web 2.0 technology.

During recent years the internet has become noticeably more participatory in nature. Users not only create and edit information, but take part in the digital community by sharing texts, images, videos, opinions and reviews with other users. Whereas many sites such as library catalogues or repositories are only visited occasionally for the purpose of retrieving clearly defined data, individuals frequently log in to their personal accounts at web 2.0 applications such as Twitter and Facebook, and use them as that they might read a newspaper, meet friends in a café or visit a museum. It follows that cultural organizations need to not only make their cultural heritage accessible via the internet in a passive way, but also to proactively feed web 2.0 platforms with their content, since these are the sites where many users are actively engaging in social and academic discourse.

As a way of moving towards this, web 2.0 based social platforms such as Wikipedia (www.wikipedia.org), LibraryThing (www.librarything.com) and Goodreads (www.goodreads.com) will be supplied with information about selected historical books. Readers all over the world who are interested in a specific book will be able to interact easily with one another, exchange information and share reading and research experiences independently of their geographical location and professional, social or academic background. Initially EOD will explore the technical feasibility of automatically generating Wikipedia source-code for selected digitized books, and then will add metadata to articles about the relevant authors. Other social platforms will be tackled at later stages of the project.

Conclusions

European libraries host millions of print books published between 1500 and 1900. Due to their age and value they are often only accessible to users who are prepared to visit these libraries in person. With the eBooks on Demand service these hidden treasures are now becoming available to all, with just a few mouse clicks.

The experience of running the EOD network has shown that such services are a desirable means of providing researchers and other customers with the material that they need 'here and now'. However, this is just a first step. In the future, it may be possible to establish further on-demand services, such as digitization for visually impaired people and the creation of fully navigable e-books rather than simple image files, with corrected full text of up to 99% accuracy. This would pave the way for the transfer of e-books onto mobile devices, making the written word available anytime, anywhere.

Acknowledgements

This work program has been funded with support from the European Commission. This publication reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

Useful Links

Culture Programme 2007-2013

http://eacea.ec.europa.eu/culture/programme/about_culture_en.php

Digital Agenda for Europe 2010-2020

http://ec.europa.eu/information_society/digital-agenda/index_en.htm

eBooks on Demand

<http://www.books2ebooks.eu/>

eTen Digitisation on Demand project factsheet

http://ec.europa.eu/information_society/activities/eten/cf/opdb/cf/project/index.cfm?mode=detail&project_ref=ETEN-518635

Europeana

<http://europeana.eu>

EuropeanaConnect project

<http://www.europeanaconnect.eu>

i2010 - A European Information Society for growth and employment

http://ec.europa.eu/information_society/eeurope/i2010/index_en.htm

IMPACT: improving access to text

<http://www.impact-project.eu/>

References

Minerva Working Group 6: Identification of good practices and competence centres, (ed.)

(2004): *Good practices handbook, Version 1.3.*

http://www.minervaeurope.org/structure/workinggroups/goodpract/document/goodpractices1_3.pdf

Mező, Z., Svoljsak, S., Gstrein, S. (2007) EOD - European network of libraries for eBooks on demand. In: *Research and advanced technology for digital libraries*, 570–572.

Mühlberger, G. and Gstrein, S. (2009) eBooks on Demand (EOD): a European digitisation service. *IFLA Journal* **35** (1), 35–43. <http://archive.ifla.org/V/iflaj/IFLA-Journal-1-2009.pdf>

Svoljsak, S. and Gstrein, S. (2007) EOD - eBooks on demand. In: *Users and use of DL & economics of digital libraries*, 229.