# Grant Agreement Number: ECP-2005-CULT-038074

#### EDL

### EDL Report on Digitisation in European National Libraries 2006-2012



Deliverable number	D-3.1.bis
Dissemination level	Public
Delivery date	29 <sup>th</sup> February 2008
Status	Final
Author(s)	Hans Petschar, Max Kaiser, Bettina Kann, Joachim Korb, Lucien Scotti, Svein Arne Solbakk, Alexander Sperl, Astrid Verheusen, David Fuegi



*e*Content*plus* 

This project is funded under the *e*Content*plus* programme<sup>1</sup>, a multiannual Community programme to make digital content in Europe more accessible, usable and exploitable.

<sup>&</sup>lt;sup>1</sup> OJ L 79, 24.3.2005, p. 1.

# TABLE OF CONTENTS

EXECUTIVE SUMMARY	5
_INTRODUCTION	10
1.1 KECENT EUROPEAN UNION LEVEL INITIATIVES INFLUENCING NATIONAL LIBRARIES AND THEIR DIGITISATION STRA	<u>regies</u> 10
2 CONTENT DIGITISATION IN EUROPEAN NATIONAL LIBRARIES	<u>18</u>
2.1 CONTEXT OF THE EDL PROJECT'S CONTENT SURVEY	18
2.2 Role of National Libraries,	18
2.3 LIBRARY STATISTICS AND DIGITAL CONTENT	19
2.4 Methodology.	20
2.5 BOOK DIGITISATION	•••
<u>OVERVIEW BY LUCIEN SCOTTI, BIBLIOTHEQUE NATIONALE DE FRANCE.</u>	23
2.0 THE STATE OF OCK OVERVIEW BY LOACHING KORD, AUSTRIAN NATIONAL LIDRARY, TEL BLUG BROUGGT	25
2 7 SUMMARY	<u>23</u> 27
	<u></u>
3 DIGITISATION STRATEGIES AND PROGRAMMES.	<u>28</u>
3.1 Summary	
3.2 Austria:	30
3.3 Belgium:	31
<u>3.4 Bulgaria:</u>	32
3.5 Croatia:	33
3.6 Cyprus:	34
<u>3.7 Czech Republic</u>	35
<u>3.8 Denmark:</u>	<u>36</u>
<u>3.9 Estonia:</u>	<u>37</u>
3.10 FINLAND:	<u>38</u>
3.11 FRANCE:	<u>39</u>
<u>3.12 GERMANY:</u> 2.12 CERMANY:	<u>40</u> 41
2.1.4 Hunderny:	41 42
3.15 ICELAND	<u>42</u> 43
3.16 IDELAND	<u></u> 45
3 17 ITALY	<u></u> 46
3.18 LATVIA:	47
3.19 Liechtenstein:	
3.20 Lithuania:	
3.21 Luxembourg:	51
3.22 Macedonia:	52
<u>3.23 Malta:</u>	53
3.24 Moldova:	54
3.25 The Netherlands:	<u>55</u>
<u>3.26 Norway:</u>	<u>56</u>
<u>3.27 POLAND:</u>	<u>57</u>
3.28 PORTUGAL:	
<u>5.29 RUSSIA:</u> 2.20 Say Marrier	<u></u>
2.20 SAN IVIARINO	01 62
3.32 SLOVAVIA:	<u>02</u> 63
3 33 \$LOVENIA.	<u></u> 64
3 34 Spain.	<u></u> 64
3.35 Sweden:	
3.36 Switzerland:	
3.37 United Kingdom:	68
A DICITICATION DECITO IN ELIDODEAN NATIONAL LIDDADIEC	60
4 DIGITISATION PROJECTS IN EUROPEAN NATIONAL LIBRARIES	<u>09</u>
4.1 Austria:	69
4.2 Belgium:	<u>71</u>
4.3 Bulgaria:	73
4.4 CROATIA:	74
4.5 CZECH REPUBLIC:	<u>75</u>

4 / DSTUNIA	78
4 8 FINI AND'	79
49 FRANCE	<u></u> 80
4 10 GERMANY'	<u></u>
A 11 GREECE:	<u>01</u> 82
4.11 OKEECE	
4.12 ITUNGARY	<u>03</u> 07
4.1.5 ICELAND.	
4.14 IRELAND:	<u>88</u>
<u>4.15 Italy:</u>	
<u>4.16 Latvia:</u>	<u></u>
4.17 Liechtenstein:	<u></u>
<u>4.18 Lithuania:</u>	94
4.19 Luxembourg:	
4.20 Macedonia:	<u>96</u>
4.21 Malta:	<u>97</u>
4.22 Moldova:	
4.23 The Netherlands:	
4.24 Norway:	
4.25 POLAND:	
4.26 PORTUGAL	104
4 27 Russia.	105
4 28 Serbia	109
4.29 SLOVAKIA.	112
1.2) SLOVAKA.	112
A 21 Spapy	115 116
4.51 SPAIN	<u>110</u> 110
4.32 Sweden.	110
4.55 SWITZERLAND.	120
4.54 UNITED KINGDOM.	121
5 CASE STUDIES	
5.1 British Library	
Digitisation Strategy	<u></u> 122
5.2 Norwegian National Library	128
THE DIGITAL NATIONAL LIBRARY	128
Legal Rights	129
Financing,	<u>130</u>
Status	<u>130</u>
Future plans	131
ESTADI SUBIC A BRODUCTION I DIE DOD THE DICITIZATION OF DOOKS	120
ESTABLISHING A PRODUCTION LINE FOR THE DIGITIZATION OF BOOKS	132
Dismounting books.	
<u>Dismounting books</u> OCR and structural analysis	<u>133</u> 134
<u>Dismounting books</u> <u>OCR and structural analysis</u> The National Library's digital long term repository	
Dismounting books OCR and structural analysis The National Library's digital long term repository Search engine	
Dismounting books OCR and structural analysis The National Library's digital long term repository Search engine Authentication and authorization – access control	
Dismounting books OCR and structural analysis The National Library's digital long term repository Search engine Authentication and authorization – access control PRODUCTION LINES	<u> </u>
Dismounting books OCR and structural analysis The National Library's digital long term repository Search engine Authentication and authorization – access control PRODUCTION LINES Description	<u> </u>
Dismounting books OCR and structural analysis The National Library's digital long term repository Search engine Authentication and authorization – access control Production lines Prioritization Ordering and extraction from the stocks	<u> </u>
Dismounting books.         OCR and structural analysis.         The National Library's digital long term repository.         Search engine.         Authentication and authorization – access control.         PRODUCTION LINES.         Prioritization.         Ordering and extraction from the stacks.         Divitization	133 134 134 135 135 135 135 135 135
Dismounting books OCR and structural analysis The National Library's digital long term repository Search engine Authentication and authorization – access control PRODUCTION LINES Prioritization Ordering and extraction from the stacks Digitization	133 134 134 135 135 135 135 135 135 135 135
Dismounting books. OCR and structural analysis. The National Library's digital long term repository. Search engine. Authentication and authorization – access control. PRODUCTION LINES. Prioritization. Ordering and extraction from the stacks. Digitization OCR/DSA.	133 134 134 135 135 135 135 135 135 135 136 136
Dismounting books. OCR and structural analysis. The National Library's digital long term repository. Search engine. Authentication and authorization – access control. PRODUCTION LINES. Prioritization. Ordering and extraction from the stacks. Digitization OCR/DSA. Digitial preservation.	133 134 134 135 135 135 135 135 135 135 136 136 136
Dismounting books.         OCR and structural analysis.         The National Library's digital long term repository.         Search engine.         Authentication and authorization – access control.         PRODUCTION LINES.         Prioritization.         Ordering and extraction from the stacks.         Digitization         OCR/DSA.         Digital preservation.         Indexing	133 134 134 135 135 135 135 135 135 135 136 136 136 136
Dismounting books.         OCR and structural analysis.         The National Library's digital long term repository.         Search engine.         Authentication and authorization – access control.         PRODUCTION LINES.         Prioritization.         Ordering and extraction from the stacks.         Digitization         OCR/DSA.         Digital preservation.         Indexing.	133 134 134 135 135 135 135 135 135 135 136 136 136 136
Dismounting books. OCR and structural analysis. The National Library's digital long term repository. Search engine. Authentication and authorization – access control. PRODUCTION LINES. Prioritization. Ordering and extraction from the stacks. Digitization OCR/DSA Digital preservation. Indexing. IMPORTANT LESSONS LEARNED THIS FAR.	133 134 134 135 135 135 135 135 135 135 136 136 136 136 137
Dismounting books.         OCR and structural analysis.         The National Library's digital long term repository.         Search engine.         Authentication and authorization – access control.         PRODUCTION LINES.         Prioritization.         Ordering and extraction from the stacks.         Digitization         OCR/DSA         Digital preservation.         Indexing.	133 134 134 135 135 135 135 135 135 135 136 136 136 136 136 137
Dismounting books.         OCR and structural analysis.         The National Library's digital long term repository.         Search engine.         Authentication and authorization – access control.         PRODUCTION LINES.         Prioritization.         Ordering and extraction from the stacks.         Digitization         OCR/DSA         Digital preservation.         Indexing.	133 134 134 135 135 135 135 135 135 135 136 136 136 136 136 137 137
Dismounting books.         OCR and structural analysis.         The National Library's digital long term repository.         Search engine.         Authentication and authorization – access control.         PRODUCTION LINES.         Prioritization.         Ordering and extraction from the stacks.         Digitization         OCR/DSA.         Digital preservation.         Indexing.    IMPORTANT LESSONS LEARNED THIS FAR.          Scope, complexity and implementation.         Actual efficiency.         Quality.	133 134 134 135 135 135 135 135 135 135 136 136 136 136 136 137 137 138 138
Dismounting books OCR and structural analysis The National Library's digital long term repository Search engine Authentication and authorization – access control Production LINES. Prioritization Ordering and extraction from the stacks Digitization OCR/DSA. Digital preservation. Indexing. IMPORTANT LESSONS LEARNED THIS FAR. Scope, complexity and implementation Actual efficiency Quality OCR/DSA.	133 134 134 135 135 135 135 135 135 135 136 136 136 136 136 137 137 137 138 138 138
Dismounting books	133 134 134 135 135 135 135 135 135 135 136 136 136 136 136 137 137 138 138 138 139
Dismounting books.         OCR and structural analysis.         The National Library's digital long term repository.         Search engine.         Authentication and authorization – access control.         PRODUCTION LINES.         Prioritization.         Ordering and extraction from the stacks.         Digitization         OCR/DSA.         Digital preservation.         Indexing.         IMPORTANT LESSONS LEARNED THIS FAR.         Scope, complexity and implementation.         Actual efficiency.         Quality.         OCR/DSA.         The digital long term repository – scaling and performance.         Statistical tools – production supervision tools.	133 134 134 135 135 135 135 135 135 135 136 136 136 136 136 136 137 137 138 138 138 138 139 139
Dismounting books. OCR and structural analysis. The National Library's digital long term repository. Search engine. Authentication and authorization – access control. PRODUCTION LINES. Prioritization. Ordering and extraction from the stacks. Digitization OCR/DSA. Digital preservation. Indexing. IMPORTANT LESSONS LEARNED THIS FAR. Scope, complexity and implementation. Actual efficiency. Quality. OCR/DSA. The digital long term repository – scaling and performance. Statistical tools – production supervision tools. Exception handling.	133 134 134 135 135 135 135 135 135 135 136 136 136 136 136 136 137 137 137 138 138 138 138 139 139 139
Dismounting books.         OCR and structural analysis.         The National Library's digital long term repository.         Search engine.         Authentication and authorization – access control.         PRODUCTION LINES.         Prioritization.         Ordering and extraction from the stacks.         Digitization         OCR/DSA.         Digital preservation.         Indexing         IMPORTANT LESSONS LEARNED THIS FAR.         Scope, complexity and implementation.         Actual efficiency.         Quality.         OCR/DSA.         The digital long term repository – scaling and performance.         Statistical tools – production supervision tools.         Exception handling.	133 134 134 135 135 135 135 135 135 135 136 136 136 136 136 136 137 137 137 138 138 138 139 140 140

5.3 Netherlands	
6 RECOMMENDATIONS	155
ANNEX	
6.1 Statistics	
7 METADATA	
8 ACCESS AND SERVICES	

# 1

# **Executive Summary**

This report gives an overview on how much digitised content there is now and will be in 2012 in European national libraries including the metadata position and solutions for converting image-mode collections [e.g. newspapers] into searchable collections. This important snapshot will be of interest to all European national libraries and to many other stakeholders in the development of the European Digital Library. It provides in convenient form a great deal of specific up to date information about the strategic plans, projects and achievements in the field of digitisation of the 37<sup>1</sup> European national libraries that responded to a survey undertaken in 2007 as part of the EDLproject. All member libraries with digitisation programs and projects responded. It also records their opinions on the obstacles remaining in this field. The report draws on the survey and on desk research including information from The European Library and <u>CENL (Conference of European National Librarians)</u> members and from the MINERVA EC network.

CENL sees a continuing need to provide statistical information on content digitisation and digital content available in European national libraries and recognises that the need may not be met by national statistical offices. Accordingly, CENL has established an annual reporting system for its members. It is hoped that parts of the survey used for this report will be completed on a regular basis by member libraries to keep the information current. This work is seen as a natural partner to <u>The European Library</u> through which CENL members provide access to their digital content. CENL members will also consider whether this report can provide a starting point for a roadmap on co-ordinated digitisation in European national libraries.

In December 2005 the CENL Executive Committee, having identified information exchange on digital content in European national libraries as a key element for future strategic planning, set up a new CENL Working Group on Content (CCWG) chaired by Hans Petschar of the Austrian National Library. The role of the Content Working Group is to act as an advisory board for CENL on digitisation issues and to contribute to planning on a strategic level for EDL. Additionally, the CCWG will provide recommendations and identify best practice on how to proceed with digitisation activities. The CWG is represented in EDL project by the Austrian National Library. The report contains a mass of information with benchmarking potential for all 47 CENL members and other organisations interested in strategic planning in this field.

The report also contributes to developments in the wider world. Just as EDLproject was about to start its work, the Commission published the <u>Commission Recommendation of 24</u>. <u>August 2006 on the digitisation and online accessibility of cultural material and digital preservation (2006/585/EC)</u> in which it calls for concerted actions and for overviews on current and planned digitisation activities and quantitative targets for digitisation to lend greater coherence to the selection of material and lead to a more secure climate for companies investing in digitisation technologies. (Paragraph 6). Also, early in 2007, the Commission let a contract for <u>Numeric</u>, a study to develop and test a framework for collecting and analysing data relating to digitisation activities of materials held by libraries, archives and museums in the EU. The EDLproject team have maintained close links with the Numeric study team and the CENL Working Group will continue to make efforts to support the development of standardised reporting in the field of digitisation.

<sup>&</sup>lt;sup>1</sup> CENL's membership includes Council of Europe members

Holdings	2006	401.058.284		
Digital Objects	2006	4.874.163	2012:	17.103.051
Digital Pages	2006	125.340.850	2012	738.404.404
Books	2006	279.190	2012	1.537.913
Newspapers	2006	1.468.000	2012	6.293.000
Photographs	2006	1.005.000	2012	1.616.000
Maps	2006	217.000	2012	329.000
Engravings	2006	157.000	2012	269.000
Posters	2006	65.000	2012	300.000
Manuscripts	2006	9.000	2012	16.000
Other Archives	2006	1.520.000	2012	3.640.000

Some headline figures on materials already digitised and existing plans for respondent national libraries are as follows:

Although content digitisation in European national libraries will reach the estimates made for the future European Digital Library, the situation remains critical for books and printed research material, especially for 20th century material. This underlines the importance and the urgency of the work on Copyright issues and the suggestion for orphan works and out of print works being undertaken by the High Level Expert Group and also its work on public private partnership for cultural institutions.

Chapter 1 provides the policy background at EU level against which national libraries are carrying out their digitisation activities. It briefly describes key policy initiatives and how CENL developed the TEL project into The European Library and how that in turn is contributing to the development of the European Digital Library using the EDLnet network and the EDL Foundation. It explains the relationships between the various projects and entities and recent activities.

Chapter 2 begins by describing why CENL wanted to undertake the content survey and explains for non-experts the circumstances of most national libraries. It describes the methodology and the questionnaires used and gives some top level numbers about digitisation. The feedback to the questionnaires was very positive and all libraries with digitisation programs and projects responded. Responses were received from the following 36 countries: Austria, Belgium, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Liechtenstein, Lithuania, Luxembourg, Macedonia, Malta, Moldova, the Netherlands, Norway, Poland, Portugal, Russia, San Marino, Serbia, Slovenia, Spain, Sweden, Switzerland and the United Kingdom. Thanks to responses from both the Russian State Library and the Russian National Library, total responses are 37.

A first case study goes on to describe lessons learned about mass digitisation through a recent study undertaken by the Bibliothèque nationale de France. A second case study describes the state of the art of OCRing based mainly on a survey of European national libraries which are members of the TELplus project, 18 in all, at the end of 2007. The chapter ends by citing Austria, Estonia, Finland, France, Latvia, Norway, Slovenia, Slovakia, Spain, Russia and the United Kingdom as promising and successful examples of strategic planning which have already resulted in concrete implementation plans against a solid financial and institutional

background. Newspaper digitisation is a major focus and is still expanding. Mass digitisation of books is still related to specific projects in a few libraries. This situation is unlikely to change in the near future unless the EU or the Member States change their policies. Making this content available at European level and developing strategies for co-operation and integration with a European dimension in national and institutional strategies will be major tasks for the future.

Chapter 3 describes the digitisation strategies of all respondents, including their perception of the challenges ahead. The chapter is arranged in alphabetical order by country names. There were 37 responses out of a possible 47 responses.19 national libraries reported having a written digitisation strategy or plans to develop one in the near future. In all libraries with documented digitisation strategies access for users, long term preservation and building up a sustainable digital library infrastructure play a dominant role. Only a few libraries have an explicit and ongoing budget for digitisation (between 100.000 and 200.000 Euro per year). In most libraries the budget for digitisation is either dependent on projects or integrated in overall costs for digital library services and infrastructure. 33% of strategies and programmes focus on national heritage, 44% on European and national heritage, 19% on European content and 4% on non European content. Priority is given to newspaper digitisation (19 libraries), mass digitisation of printed material (14 libraries), unique and fragile resources, European and national heritage (19 libraries). Basic obstacles are seen in financing of mass digitisation at national and EU level (22 libraries), multilingual approach of the European digital library (including usability and standards (12 libraries), copyright (20<sup>th</sup> century) (9 libraries), and lack of infrastructure, technology and expertise (7 libraries).

Chapter 4 describes the digitisation projects undertaken up to 2007 or still in progress at that time in the 33 libraries that responded to this part of the survey. 184 completed questionnaires on individual projects were received and the projects are all described in this chapter which is again arranged in alphabetical order of countries. Projects vary greatly in their size and ambition.

Chapter 5 brings together a number of case studies contributed by various authors describing in some detail the issues around the development of their digitisation strategies. It includes contributions from the British Library, the Norwegian National Library, the National Library of the Netherlands and the Austrian National Library

Chapter 7 consists of the editor's conclusions from EDL report and recommendations.

European national libraries have been digitising parts of their collections for the last ten years.

Many Libraries have created national memory collections and have built up a digital library infrastructure.

The last ten years of digitisation at the European national libraries have created a valuable and enduring resource for scholars and the public alike.

### However:

If we look at the general outcomes from the survey on digitisation (chapters 4 & 5) it becomes clear, that although National Libraries are well prepared to meet the challenges of digitisation on a strategic level, the production of digital content is still on a very low level. Only 1-2 % of national libraries' collections will be digitally available in the near future. In terms of material types digital collections reflect the structure of analogue collections, with one basic and most critical exception: book digitisation and digitisation of printed research collections is far below expectations for the future European Digital Library. This will not change unless all stakeholders at EU, national and institutional level change their policies.

Due to successful private projects and initiatives, access to searchable digital books and printed collections are what users regard as the essential basis of a digital library. Currently a large part of digitised printed heritage is accessible in image mode only. Although the issue of full-text conversion (OCR) of digitised historical printed material is now being tackled by the European projects TELplus (eContentplus) and IMPACT (FP7) much more needs to be done to create a critical mass of full text for the European Digital Library. Due to this situation there is a high risk that users will not understand the concept of the European Digital Library based on a digital library infrastructure which guarantees sustainability, long term preservation, interoperability and access to multilingual collections, but without reaching a critical mass of digital content. Given the fact that mass digitisation of printed materials (with 100.000 volumes per year) (case studies of France and Norway) is not only still very cost- intensive but also timeconsuming and takes at least a preparation period of half a year it becomes clear that for the time being only very few libraries (Norway, France, UK, Netherlands) will be capable of doing large scale digitisation. On the other hand the example of Norway indicates that for the first time complete digitisation plans for national libraries can be based on a realistic background and that cost reduction in storage technologies and mass digitisation make it for the first time imaginable to "duplicate" the total collections of libraries in electronic format, to preserve them and to give virtual access to the complete content.

### Recommendations:

The first results of EDL Report on Digitisation were presented and discussed at the CENL annual meeting in Helsinki in October 2007. The members of CENL identified the following as the most critical issues: large scale digitisation of books and printed material; understanding of user needs; innovation and technological change; smarter marketing of the European Digital Library. A need for co-ordinated digitisation of content and reporting on digitisation on a regular basis (including statistical data) may be added to this list as a result of this EDL Report on Digitisation.

The following recommendations emerge from this discussion and from the conclusions of this EDL Report on Digitisation:

### Large scale digitisation of books and printed material

Recommendation: Given the massive size of the digitisation challenge which they face and the resource issues alluded to by many of them, national libraries should consider carefully whether and how public private sector partnerships might contribute to solving the problem.

- all efforts have to be undertaken to enlarge the scale of book digitisation and to give access to printed material, including 20<sup>th</sup> century documents
- all stakeholders should be involved, including appropriate public-private partnerships
- special attention should be paid to the work on orphan works and out of print works being undertaken by the High Level Expert Group

### Understanding of user needs

Material held in national libraries has been traditionally organised in a ways that help the trained librarian find information on the shelves. Content needs to be organised for a web audience. This means granularity of data to allow full text searching, the use of protocols allowing data to be crawled and harvested and the organisation of data to ensure maximum discovery.

For EDL to be a true success all the domains involved need to produce digitised, searchable, interoperable data

### Recommendation:

• The best way to meet user needs is to give direct access to digital content. CENL should support The European Library in developing a technological platform with central harvesting to create searchable and interoperable data

### Innovation and technological change

### Recommendation:

• Considerable attention should be paid to the OCRing of historical material, metadata extraction, content analysis tools and navigation applications

### Co-ordinated digitisation of content

- National libraries that have not yet done so should build up at institutional level a digital library infrastructure based on approved strategies and development plans
- National digitisation strategies should be based on a territorial approach and should focus on building large digital collections
- Partnerships with other organisations should be formed to create larger virtual collections.

### Report on digitisation an a regular basis (including statistical data)

Recommendation: CENL should continue to maintain close links with the Numeric study team and the CENL Content Working Group should continue to support the Commission's efforts to develop standardised reporting in the field of digitisation and raise the profile of the related issues with governments and other stakeholders.

• National Libraries should report on an annual basis their progress in digitisation in a way compatible with emerging European standards for such reports. (Numeric)

#### Improve awareness of The European Library and its role in building Europeana

There is not enough public awareness of The European Library. More effort is required to market this resource in each participating country. The crucial role that the national libraries are playing in the creation of Europeana should also be brought to public attention.

• CENL members should support the TEL-office in the marketing of The European Library as a resource and together with the European Commission make content providers from other domains aware of the importance of contributing material to Europeana

# Introduction

# 1.1 Recent European Union level initiatives influencing national libraries and their digitisation strategies

"The heritage of European libraries is unequalled in richness and diversity ... But if it is not digitised and made accessible online, this heritage could, tomorrow, fail to fill its just place in the future geography of knowledge".<sup>1</sup>

In April 2005 a letter was sent to European Council President Jean-Claude Juncker and European Commission President José Manuel Barroso, signed by the French, Italian, Spanish, German, Polish and Hungarian Heads of State and of Government formulating the above mentioned statement and including the proposal for the creation of a European Digital Library.

The six leaders noted that many libraries across Europe have started to digitize their collections, adding that *existing efforts should be coordinated* in order to develop a network and "create what could be called a European digital library". By coordinating and supporting current digitisation initiatives, the European Union could provide a framework for the project and contribute to the removal of any barriers, the letter said. Such an approach could among other things avoid redundancies in the scanning of works, allow for the development of common, state of the art search tools, and benefit from common studies and strategies to meet the expectations and needs of users.

The letter, which was also sent to all the members of the European Council, called for an EUwide debate of the proposed project. In the short term, the EU Member States' ministers in charge of cultural affairs and research should meet to discuss the feasibility and the modalities of the digital library. The ministerial discussions should be fed by a Commission Communication, the letter said, as well as by contributions by Member States.

In a press release issued on 3 May 2005, the European Commission welcomed the digital library proposal. "The message addressed to the Presidency of the European Union and the Commission by six EU heads of State and government is a positive sign of increased awareness of the need to bolster our policy in this field", commented Information Society and Media Commissioner Viviane Reding, adding that "digitising this heritage is also a marvellous opportunity which our industry should grasp".<sup>2</sup>

On 1 June 2005 the Commission presented the i2010 Initiative, which sets out its strategy for digitisation, online accessibility and digital preservation of Europe's collective memory and made digital libraries a key aspect of the i2010 initiative.

On 7 July 2005 Mr José Manuel Barroso, President of the European Commission, <u>replied to</u> <u>the letter</u> from 6 Heads of State and their respective Government's and welcomed the digital libraries initiative.

The national librarians of 19 EU Member States announced their support for the European digital library proposal and on 30 September 2005 the Commission adopted a

<sup>&</sup>lt;sup>1</sup> Letter of 28 April to EC President Barroso. Available at: http://ec.europa.eu/information\_society/activities/digital\_libraries/timeline/index\_en.htm

<sup>&</sup>lt;sup>2</sup> eGovernment News – 04 May 2005. Available at: http://europa.eu.int/idabc/en/document/4239/330

<u>Communication "i2010: Digital Libraries"</u> which outlines the vision underlying the digital libraries initiative and deals with Europe's <u>cultural heritage</u>.

### The "i2010: Digital Libraries initiative"

The "i2010 initiative included as a first concrete action an online consultation on digitisation, accessibility and preservation of digital content, which started in September 2005 and ended on 20<sup>th</sup> January 2006.

All interested individuals and organisations - from the private and from the public sector – were encouraged to provide their views on questions arising around the creation of a European Digital Library. The consultation closed on 20 January 2006 and received 225 replies. In general, the initiative was welcomed and seen as an opportunity to make Europe's cultural heritage more accessible and usable.

Most European national libraries responded to the commission and CENL, the Conference of European National Librarians, sent an answer formulating the position of Europe's national librarians:

"CENL welcomes with enthusiasm the Communication on digital libraries and greatly appreciates the opportunity to take part in the online consultation process. CENL shares the vision of a European Digital Library and has been working towards this goal by creating TEL, The European Library. Reading the Commissioner's Communication, however, gave the impulse to think in a much broader way of a more comprehensive European Digital Library.

A true European Digital Library (EDL) should serve all types of user needs: present and future ones, up-to-date knowledge or information and cultural heritage documents, science and humanities, education, research and everyday, "normal" information needs. It must comprise all types of media from all types of Europe's cultural heritage institutions: from libraries, museums and archives."<sup>1</sup>

CENL stressed that the European Digital Library should not be constructed from scratch, but build on existing initiatives and proposed The European Library [TEL] as a model platform and model organisational network for building the European Digital Library.

Many responses from other institutions and networks in the cultural heritage sector stated that it would be extremely difficult to organise a central, comprehensive super-structure for the European Digital Library and that a network of networks approach would be more realistic, allowing faster partners to proceed at higher speeds, and slower ones to participate at their own speed, but with all gaining from the process.

From the participating institutions' point of view it is the technical interoperability for limitless data exchange and high performance access that are primordial. From the users' point of view, cross-searching of all data and cross-services offered by all or most of the network-members is of great importance.

To become a real-world digital library, however, a network of networks needs a solid organisational, legal and technical structure. Therefore CENL's proposal to use The European Library, the platform of European national libraries as a starting point for the European Digital Library was supported by the responses to the online consultation. As Olaf Janssen, The European Library Office, Den Haag pointed out:

<sup>&</sup>lt;sup>1</sup> Answers to the Online Consultation of the European Commission concerning "i2010 digital libraries". CENL Archive.

"In fact, the results of the online consultation indicated that The European Library would be a very good starting point, because it not only offers a working technological platform based on common standards, but also provides a firm cooperative organisational framework, in which European national libraries already collaborate and have experimented with improving the online accessibility of their digital assets."<sup>1</sup>

The European Commission reacted immediately to the results of the consultation and decided on 27 February 2006 to set up a 20 member *High Level Expert Group on Digital Libraries*<sup>to</sup> advise the Commission on how to best address the organisational, legal and technical challenges at a European level and to contribute to a shared strategic vision for the European Digital Library.

On 24 August 2006 the Commission adopted a "<u>Recommendation on the digitisation and</u> <u>online accessibility of cultural material and digital preservation</u>"<sup>2</sup>, which aimed at defining the roles of the member states and indicating the priority actions the Member States need to undertake, to get Europe's digital heritage online.

In its first paragraph the recommendation states the importance of digital libraries for the i2010 initiative:

"(1) On 1 June 2005 the Commission presented the i2010 initiative, which seeks to optimise the benefits of the new information technologies for economic growth, job creation and the quality of life of European citizens. The Commission has made digital libraries a key aspect of i2010. In its Communication 'i2010: digital libraries' of 30 September 2005 (1), it set out its strategy for digitisation, online accessibility and digital preservation of Europe's collective memory. This collective memory includes print (books, journals, newspapers), photographs, museum objects, archival documents, audiovisual material (hereinafter 'cultural material')."

Paragraph 2 stresses the importance of the Internet for optimizing the economic and cultural potential of Europe's cultural heritage.

Therefore the "development of digitised material from libraries, archives and museums should be encouraged. The online accessibility of the material will make it possible for citizens throughout Europe to access and use it for leisure, studies or work. It will give Europe's diverse and multilingual heritage a clear profile on the Internet. Moreover, the digitised material can be re-used in industries such as tourism and the education industry, as well as in new creative efforts." (Paragraph 3)

In the Recommendation digitisation is seen as an important means of ensuring greater access to cultural material and the only means to make such material available for future generations. To overcome the fragmented efforts of many digitisation initiatives in the Member States, the Commission asks for concerted actions and for overviews on current and planned digitisation activities and quantitative targets for digitisation. This would lend greater coherence to the selection of material and lead to a more secure climate for companies investing in digitisation technologies. (Paragraph 6)

The Recommendation encourages public partnership, investments in new technologies and the development of large scale digitisation facilities, which can bring down costs of digitisation while maintaining or improving quality.

<sup>&</sup>lt;sup>1</sup> Olaf Janssen: Online Collaboration between European National Libraries Since 1995 – Building Towards a European Digital Library. Available at: http://libraries.theeuropeanlibrary.org/press/press\_en.html

<sup>&</sup>lt;sup>2</sup> COMMISSION RECOMMENDATION of 24 August 2006 on the digitisation and online accessibility of cultural material and digital preservation (2006/585/EC). Available at:

http://ec.europa.eu/information\_society/activities/digital\_libraries/index\_en.htm

Since only part of the material held by libraries, archives and museums is in the public domain and no longer protected by intellectual property rights the Recommendation suggests licensing mechanisms in areas such as orphan works (copyrighted works whose owners are difficult or even impossible to locate) and works that are out of print or distribution to facilitate rights clearance and consequently digitisation efforts and subsequent online accessibility. "Such mechanisms should therefore be encouraged in close cooperation with rightholders" (paragraph 10)

The development of effective means of digital preservation, collaboration between Member States in questions regarding legal obligations for the deposit of digital material and recognition of Web-harvesting is a new technique for collecting material from the Internet for preservation purposes are further topics mentioned in the Recommendation.

For national libraries most important, however, is paragraph 9, which clearly indicates that the European Digital Library will be based on The European Library (TEL) as proposed by CENL:

"(9) A common multilingual access point would make it possible to search Europe's distributed — that is to say, held in different places by different organisations — digital cultural heritage online. Such an access point would increase its visibility and underline common features. The access point should build on existing initiatives such as The European Library (TEL), in which Europe's libraries already collaborate. It should where possible closely associate private holders of rights in cultural material and all interested stakeholders. A strong commitment by the Member States and cultural institutions to arrive at such an access point should be encouraged."

Based on the Recommendation the Commission urges Member States to tackle three main areas:

Digitisation of content

by setting up large scale digitisation facilities.

by clearly indicating what they have already digitised and by quantifying their further plans. This will help to avoid overlap and to create collections with European added-value.

by encouraging partnerships between cultural institutions and the private sector.

### Online accessibility

by promoting the development of the European Digital Library as the multilingual access point to Europe's cultural heritage. This can be done, for example, by specifying conditions for granting funding to cultural institutions for digitisation work.

by looking at and working towards concrete solutions on copyright issues, for example mechanisms to deal with "orphan" works (copyrighted works whose owners are very difficult or impossible to locate) and works that are out of print.

### Digital preservation

by establishing national strategies and plans for the long-term preservation of and access to digital material.

by adapting legislation, where necessary, to allow multiple copying and migration for preservation purposes, and to tackle the issues of web-preservation and the deposit of digital material for preservation purposes.

In response to the Commission Recommendation, the Education, Youth and Culture Council adopted at its 13 November 2006 meeting conclusions showing the commitment of Member States to work together towards the European Digital Library.

It is worth mentioning that also the Council's resolution "underlines the instrumental work done at European level by CENL (the Conference of European National Librarians); — in organising and creating The European Library (TEL as a gateway to the collective resources of national libraries across Europe; — in carrying this work forward towards the creation of the European Digital Library"<sup>1</sup>

At the same time however "the goal of achieving a balanced cooperation between libraries, museums and archives" is formulated and projects and initiatives in this area like MICHAEL<sup>2</sup> are mentioned.

The reactions to the "i2010 digital libraries" initiative as a "flagship initiative" within the i2010 program and already called '*one of the greatest digital construction projects ever undertaken*' were overwhelming and not without contradictions.

Especially from the museums and archives communities reservations were made against an approach dominated by (national) libraries. Currently the principal fora for resolving interdomain issues relating to the European Digital Library are the <u>EDLnet network</u> and the <u>EDL</u> <u>Foundation</u>.

An ambitious timeline for building the European Digital Library is proposed:

1• *2006* - Full EU-wide collaboration between national libraries in the framework of The European Library and the CENL.

2• *2008* - Multilingual access to digital collections of national libraries through The European Library portal. The collections must be searchable and usable. A minimum of 2 million digital works (books, pictures, sound files etc.) should be accessible through the European Digital Library.

3• 2010 - The European Digital Library needs to have expanded to include collections of a number of archives, museums and other libraries, and possibly publishers. A minimum of 6 million digital works should be accessible through the European Digital Library. In practice, this number can be much higher, if cultural institutions of different types and at different levels (national, regional, local) participate.

### CENL, The European Library and the timeline for a European Digital Library

In September 2006, one month after the Commission published the "<u>Recommendation on the</u> <u>digitization and online accessibility of cultural material and digital preservation</u>" and communicated that the European Digital Library would be based on TEL as a model platform, the Conference of European National Librarians (CENL) was celebrating its 20th anniversary at the Annual Meeting in St. Petersburg.

<sup>&</sup>lt;sup>1</sup> Council Conclusions on the Digitisation and Online Accessibility of Cultural Material, and Digital Preservation (2006/C 297/01). Available at:

 $http://ec.europa.eu/information\_society/activities/digital\_libraries/index\_en.htm$ 

<sup>&</sup>lt;sup>2</sup> Multilingual Inventory of Cultural Heritage in Europe.

CENL has published its vision and mission statement and its policy towards digitisation on the Web:

"The Conference of European National Librarians (CENL) is a foundation under Dutch law with the aim of increasing and reinforcing the role of national libraries in Europe, in particular in respect of their responsibilities for maintaining the national cultural heritage and ensuring the accessibility of knowledge in that field.

Members are the national librarians of the member states of the Council of Europe. The conference currently consists of 47 members from 45 European countries forming the CENL Board. The CENL Board meets once a year and appoints the members of the Executive Committee for a period of at least three years. The Executive Committee consists of at least three members of the Board (chair, vice-chair and treasurer) and is charged with the management and the representation of the foundation.

CENL was founded in 1987 at the first meeting of 11 European national librarians in Lisbon. The following countries were represented: Denmark, France, Germany, Greece, Luxembourg, Netherlands, Norway, Portugal, Spain, United Kingdom and Vatican City. Topics on the agenda were the interconnection of computerized systems of the national libraries, acquisition policies, preservation and conservation, and financial issues of national libraries. The national librarians continued to meet annually and the group grew steadily."<sup>1</sup>

The vision of CENL is to "enable European national libraries to open up a world of information and knowledge and safeguard Europe's memory and cultural heritage".

Since 2005 building the European Digital Library has been defined as one of the key missions of CENL. At the CENL annual meeting in Luxemburg on 30<sup>th</sup> September 2005 the "CENL Luxembourg Resolution on digitisation of European Cultural Heritage" states:

"CENL, by creating The European Library offers a model platform and organisational network for co-operation between national libraries and for the co-ordination of national initiatives on a European level and avoids duplication of work. CENL is committed to the development of a network of digital activities as well as to the promotion and exchange of experiences regarding long-term preservation of digital and digitised objects. The European Library is an ideal point of entry to the digitised material of national libraries and other European cultural institutions.

CENL will engage actively with the proposals of the Information Society and Media Directorate stakeholder consultation and anticipates being a key player in any subsequent initiatives."<sup>2</sup>

Today CENL is a key player in the European Digital Library initiative and

- promotes a cooperative and widespread approach to digitisation by European national libraries and European libraries within the respective national networks
- continues to develop The European Library elements of which may play a role as models for building the new entity, the European Digital Library
- builds an organisational network for cooperation
- participates in the European Digital Library initiative
- acts as consultant to the EC regarding libraries

<sup>&</sup>lt;sup>1</sup> http://cenl.org

 $<sup>^2</sup>$  CENL Luxembourg Resolution on digitisation of European Cultural Heritage (Approved by the CENL Executive Committee, December 5<sup>th</sup> 2005). Available at: http://cenl.org

The European Library (TEL) originates from the TEL project funded by the European Commission, which was finished successfully on January 31, 2004. The key aim of the TEL project was to investigate the feasibility of establishing a new Pan-European service (named The European Library) which would ultimately give access to the combined resources of the national libraries of Europe.

After a preparation phase of nine months in which the results from of TEL Project were converted into an operational service, *The European Library portal* (v1.0) was launched on 17<sup>th</sup> March 2005.

Full Participants since 2005 are the national libraries of <u>Austria</u>, <u>Croatia</u>, <u>Czech Republic</u>, <u>Cyprus</u>, <u>Denmark</u>, <u>Estonia</u>, <u>Finland</u>, <u>France</u>, <u>Germany</u>, <u>Hungary</u>, <u>Italy</u>, <u>Latvia</u>, <u>Lithuania</u>, <u>Malta</u>, <u>Netherlands</u>, <u>Poland</u>, <u>Portugal</u>, <u>Serbia</u>, <u>Slovakia</u>, <u>Slovenia</u>, <u>Switzerland</u> and the <u>United</u> <u>Kingdom</u>, along with <u>ICCU</u> (the national central cataloguing institute from Italy) and <u>CENL</u> (the Conference of European National Librarians). The collections of the Full Participants were the first to be included (searchable) on the European Library website.

Basic Participants were the other 22 European national libraries, with their collections to be included in an ongoing process.

Within the framework of <u>EDL-project</u> since July 2007 Belgium, Greece, Iceland, Ireland, Liechtenstein, Luxembourg, Norway, Spain and Sweden have become full members of The European Library.

CENL and The European Library office are heavily engaged in European projects related to digital libraries. While the successful TEL-ME-MOR project which ended in January 2007 integrated the (then) New Member States to The European Library, the EDL-project, which started in September 2006 under the eContentplus Programme works towards the integration of the bibliographic catalogues and digital collections of the National Libraries of Belgium, Greece, Iceland, Ireland, Liechtenstein, Luxembourg, Norway, Spain and Sweden.<sup>1</sup> EDLproject also addresses the enhancement of multilingual capabilities of The European Library and other non-library cultural initiatives, and expands the marketing and communication activities of The European Library service.

Within the framework of the EDL-project two workshops were organised by the Austrian National Library on 27/28<sup>th</sup> November 2006 and on 31. January to 1 February 2007 in Vienna, which brought together representatives of projects funded by the eContentplus programme and a number Cultural Heritage organisations and networks to discuss practical collaboration and support the development of the "European Digital Library." Both workshops were precursors to the work currently being taken forward through EDLnet and EDL Foundation in encouraging cooperation between the archive, library, audio-visual archive and museum domains.

### EDLnet and the European Digital Library Foundation

On 28 November 2007 the European Commission endorsed the work of the European Digital Library Foundation when high level EC officials met members for the formal handover of their statutes. "Europe's citizens should all be able to enjoy our rich cultural heritage. This Foundation is a significant step towards making that ambition come true," Commissioner Viviane Reding, responsible for Information Society and Media commented. "It shows the commitment of Europe's cultural institutions to work together to make their collections available and searchable by the public through a common and multilingual access point online."

<sup>&</sup>lt;sup>1</sup> http://www.EDLproject.eu

Foundation members include the key European heritage and information associations. Their statutes commit members to work in partnership to:

- Provide access to Europe's cultural and scientific heritage though a cross-domain portal
- Co-operate in the delivery and sustainability of the joint portal
- Stimulate initiatives to bring together existing digital content
- Support digitisation of Europe's cultural and scientific heritage

The European digital library is developing its prototype site for launch in 2008. To coincide with the formal handover to the Commissioner, the Foundation announced the CITY as the first of the site's themes.

The CITY is a broad theme that will enable the prototype site to show the European urban experience from several perspectives. Emerging ideas include:

- cities of the future/cities of the past
- migration and diaspora
- trade and industry
- design, shopping and urban cool
- pox, cholera and the plague: the route to urban health
- archaeology and architecture
- utopias and cities of the imagination
- riot and disorder
- palaces and politics

The digital library project is gathering digitised content from European archives, museums, audio-visual collections and libraries. It will use maps, artefacts, photos, sound, film material, books, archival records and artworks to explore two millennia of connectivity between Europe's cities.

The prototype of the European digital library will give direct access to at least 2 million digitised books, photos, maps, sounds, films and archival records from Europe's libraries, archives, museums and audio-visual collections.

According to Wim van Drimmelen, foundation member and Director of the Koninklijke Bibliotheek, the Netherlands, which hosts the European digital library initiative "archives, museums, audio-visual collections and libraries are collaborating in order to guarantee that their resources can be brought together in a virtual world, regardless of where the original is held. It's what users expect. However, the benefits are not just for the users. The development work is providing an excellent forum for knowledge transfer between the domains, which increases our collective ability to respond to changing user needs and remain relevant in the fast-moving digital environment."<sup>1</sup>

<sup>&</sup>lt;sup>1</sup> European Digital Library Foundation Press Release, 28<sup>th</sup> November 2007. For more information see: <u>http://www.europeandigitallibrary.eu/edlnet/</u>

# 2 Content Digitisation in European National Libraries

# 2.1 Context of the EDL project's content survey

There is a need for more statistical information on content digitisation and digital content available in European national libraries.. CENL has established an annual reporting system for its members which aims at monitoring the development of digital libraries in European national libraries on a regular basis. The European Library, as an operational service provides an overview on the collection development of European national libraries through its portal. No statistical data however are available for digital content and digital content development. The CENL Executive Committee identified information exchange on digital content in European National Libraries as a key element for future strategic planning and set up a new CENL Working Group on Content at its 9<sup>th</sup> EC meeting in Frankfurt on 5<sup>th</sup> December 2005.It appointed Hans Petschar, CENL representative of the Austrian National Library, to lead this group and to engage the members as proposed: Austria, Norway for the Nordic countries, France, Hungary for the Central-Europe, Portugal and the United Kingdom. The working group will work on how content for mass digitisation in Europe can be selected and created for the European Digital Library.

The role of the Content Working Group is to act as an advisory board for CENL and to contribute to planning on a strategic level for EDL. The Content Working Group's role is to advise the CENL board on digitisation issues and work on recommendations, best practice and on advice how to proceed with digitisation activities. The CENL Content Working Group is represented in EDL- project via the Austrian National Library.

In its first two years the Working Group had a strong focus on creating awareness among CENL members for the necessity of strategic planning in digitisation and for monitoring the process. Information exchange and collecting information on available and planned digital content was been identified as a major task to be worked out within the framework of EDL-project.

The EDL project's content survey concentrated on material needing to be digitised, not born digital resources. Because of the nature of the holdings of most national libraries, the results relate mainly to paper based materials.

# 2.2 Role of National Libraries

European national libraries vary widely in their origins and in size and functions. At its meeting in Bangkok in 1999 the CDNL (Conference of Directors of National Libraries) defined a National Library as an:

"Institution, primarily funded (directly or indirectly) by the state, which is responsible for comprehensively collecting, bibliographically recording, preserving and making available the documentary heritage (primarily published materials of all types) emanating from or relating to its country; and which may also be responsible for furthering the effective and efficient functioning of the country's libraries through such tasks as the management of nationally significant collections, the provision of an infrastructure, the coordination of activities in the country's library and information system, international liaison, and the exercise of leadership. Normally these responsibilities are formally recognised, usually in law. For the purpose of this definition a country is defined as a sovereign independent state."<sup>1</sup>

<sup>&</sup>lt;sup>1</sup> International Encyclopedia of Information and Library Science, ed. By John Feather and Paul Surges.-London: Routledge 2003 (2<sup>nd</sup> Ed.). p. 445.

To summarize: a national library is responsible for acquiring and conserving copies of all relevant (published) documents in the country in which the library is located; it may function as a legal deposit library. Specific characteristics for European national libraries' collections are:

- the universal approach of many national libraries being successors of Royal Libraries
- cultural, geographical and linguistic diversity of collections reflecting historical changes in modern European history
- a strong focus on national heritage documentation, especially in the new member states of the EU

It goes without saying that all these aspects have a strong impact not only on collection development but also on retrospective digitisation of the collections.

### 2.3 Library Statistics and digital content

Traditionally, library statistics<sup>1</sup> have three principal functions:

- 1) to monitor operational effectiveness
- 2) to provide a base for strategic planning
- 3) to demonstrate the value obtained by users including the potential value to users of future generations.

These principal functions are valid also for statistical information on digital content. EDL content survey therefore has the overall objective to provide information on digital content production in European National Libraries:

- What is available now
- What will be available in the next 3-5 year
- How will this content be accessible

Monitoring the process on a regular basis and providing information for strategic planning are further objectives which should go beyond the lifetime of EDL-project.

To get an overview on digitisation activities in European national libraries collecting data is a necessary first step. *The focus for the EDL project's content survey lies in digital content production from physical library resources and is therefore directly related to collection development and resource statistics of European national libraries.* 

"A library collection is an assemblage of physical information sources combined with virtual access to selected and organized information sources."<sup>2</sup> Collection development focuses on the building of library collections, following guidelines in relation to the mission of individual libraries and the current or future needs of users.

For national libraries content digitisation and establishing a digital library have a strong impact on collection development policies and strategic planning. Cost reduction in storage technologies and mass digitisation make it for the first time imaginable to "duplicate" the total collections of libraries in electronic format, to preserve them and to give virtual access to the complete content.

<sup>&</sup>lt;sup>1</sup> International Encyclopedia of Information and Library Science, ed. By John Feather and Paul Surges.-London: Routledge 2003 (2<sup>nd</sup> Ed.). p. 619.

<sup>&</sup>lt;sup>2</sup> International Encyclopedia of Information and Library Science, ed. By John Feather and Paul Surges.-London: Routledge 2003 (2<sup>nd</sup> Ed.). p. 81.

But what do the complete holdings of a national library comprise, when it comes to digitisation?

Traditionally statistics on library collections, stock and acquisitions, comprise a variety of media types and formats: books, periodicals (current and back runs), newspapers, other printed materials, manuscripts, maps, music, audiovisual materials, museum objects etc. Microform and microfiche materials may be counted as books or as electronic formats. Periodicals (esp. back runs) are often counted as books. Newspapers are counted either as bibliographic entities (item per title) or as physical items (bounding entities). Bookstock (including periodicals and newspapers) therefore is often measured in length of shelving (linear meters) rather than in volumes to avoid misleading results, when slight leaflets and thick tomes count as equal units.

Legal deposit and delivery of more than one copy of a printed book (or other material) (2-5 copies) to a national Library may also have an influence not only on statistics but also on selection for digitisation.

### 2.4 Methodology

Accordingly, with the support of the CENL Content Working Group, two major questionnaires covering all aspects of digitised collections and on strategies and programs were prepared and sent to all 47 CENL members in April 2007. The results were analyzed and a report was made to the CENL meeting in Helsinki in September 2007 where it was discussed.

The feedback to the questionnaires was very positive and all libraries with digitisation programs and projects responded. Responses were received from the following 36 countries: Austria, Belgium, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Liechtenstein, Lithuania, Luxembourg, Macedonia, Malta, Moldova, the Netherlands, Norway, Poland, Portugal, Russia, San Marino, Serbia, Slovenia, Spain, Sweden, Switzerland and the United Kingdom. Thanks to responses from both the Russian State Library and the Russian National Library, total responses are 37.

The EDL project's content survey reflects great variety in collection development statistics but nevertheless tries to identify comparable data for content digitisation. When asking for statistical data on physical collections the EDL project's content survey asked for approved figures given by national libraries in their annual reports.

The EDL content survey questionnaire<sup>1</sup> consists of 3 parts:

- PART 1: Introduction, goals, background (this part): *DigitizationQuestionnairePart1.doc*
- PART 2: Programmes and Strategies: DigitizationQuestionnairePart2\_XXX.doc
- PART 3: Single Collections and Projects: DigitizationQuestionnairePart3\_XXX\_nn.doc

In Part 1 the aims of the survey are explained:

<sup>&</sup>lt;sup>1</sup> See Annex

The results of the survey should give an overview on institutional digitization programs and policies and provide measurable data for content digitization: What will be digitized when and how in European National Libraries in the next 5 years.

The Digitisation Roadmap which will be based on the survey will describe the current baseline: what is available in what form in the EU national libraries and the other CENL member libraries against the total holdings of these libraries. It will be used to make stakeholders aware of how enormous the potential is.

Part 2: Programmes and Strategies consists of 28 questions asking for information on

- National and institutional programs and strategies
- Funding entities and responsibilities
- Annual budget for digitisation
- Collections and materials to be digitised within the next 5 years
- Priorities in digitisation
- Obstacles for digitisation and availability of content

Part 3: Single Collections and Projects consists of 74 questions asking for information on

- Total number of physical objects of collections
- Single digital collections and projects already digitised
- Planned digitisation projects
- Budget
- Selection criteria
- Types, figures and timelines for materials to be digitised
- Thematic and geographic coverage of digital content
- Technical information on media types
- Metadata
- Availability of content
- Access and Services

Feedback and first results

The feedback to the questionnaires was very positive and all libraries with digitisation programs and projects responded:

- Questionnaires collected from 37 (out of 47) CENL libraries
- 37 completed questionnaires on programmes and strategies
- types of digitized objects 2012
- 184 completed questionnaires on individual projects

The total number of physical items in European National Libraries according to this information is: 401.058.284







In most libraries statistical data on collections are derived from bibliographic entities. When asking for digital objects EDL questionnaire was asking for the number of physical items digitised and for the number of pages digitised for printed material. Since most mass digitisation projects focus on material before 1930 and acquisitions of older material is rather low, the annual increase rate does not influence content digitisation to a large extent. (It has to be mentioned however that solutions to the 20<sup>th</sup> century hole and access to material under copyright have been identified as key success factors for the EDL.)

If we look at holdings and digitisation we get for libraries with reported digitisation projects and programs:

Holdings	2006	401.058.284		
Digital Objects	2006	4.874.163	2012:	17.103.051
Digital Pages	2006	125.340.850	2012	738.404.404
Books	2006	279.190	2012	1.537.913
Newspapers	2006	1.468.000	2012	6.293.000
Photographs	2006	1.005.000	2012	1.616.000
Maps	2006	217.000	2012	329.000
Engravings	2006	157.000	2012	269.000
Posters	2006	65.000	2012	300.000
Manuscripts	2006	9.000	2012	16.000
Other Archives	2006	1.520.000	2012	3.640.000

If we look at these figures it turns out that, although content digitisation in European national libraries will reach the estimates made for the future European Digital Library, the situation remains critical for books and printed research material, especially for 20th century material.<sup>1</sup>

<sup>&</sup>lt;sup>1</sup> Erland Kolding Nielsen, director of the Royal Library of Denkmark, when confronted with the first results of this report presented at the CENL annual meeting in Helsinki 26-28<sup>th</sup> September 2007, to state in his opening speech at the LIBER-EBLIDA workshop on digitisation of library material in Europe in Copenhagen 24 – 26 October 2007: "The European Digital Library will be a library without books".

The problem of mass digitisation of books and printed material and their availability in searchable form therefore deserves special attention.

National libraries are collaborating with the European Commission to coordinate efforts of the various cultural heritage domains to move towards an operational European Digital Library. Most of the national libraries of Europe have begun the process of giving access to their national cultural heritage via the web.

Since 2000 many European countries and national libraries have established a digitisation programme at national level. The Netherlands, Great Britain, and France provide examples of countries where national libraries are playing a key role in the digitisation of national cultural heritage. It is interesting and encouraging however to recognize that in recent years many European national libraries of all sizes in both old and new Member States have started digitisation programmes at institutional level. Many of the new Member States and candidate states have been successful in convincing their government to develop a national digitisation strategy and policy plan for their country. Membership of the European Union has had a strong impact on digitisation on a national level and almost everywhere the national library plays a leading role in successful implementation of national programs.

The European Digital Library initiative has become an important political project and has a strong impact on Information Society development in the Member States. The national libraries of Europe are key players in this process on a political and a practical level. To reach the ambitious goals that have been defined and above all to fulfil the needs and wishes of users and citizens and giving full access to the digital heritage of Europe they need to intensify their collaborative efforts and exchange information, knowledge and above all: produce and share digital content.

### 2.5 Book Digitisation Overview by Lucien Scotti, Bibliothèque nationale de France

In the context of the European digital library, the Bibliothèque nationale de France launched a study in order to define the feasibility and the organisation requirements for a mass digitisation, prior to implementing its plan to put online digitised books chosen from its collections.

This study was entrusted to IBM, who carried out several series of tests with the help of 3 firms (INFOTECHNIQUE, JOUVE and SAFIG) specialized in the digitisation and OCR-treatment of documents as well as with the support of BnF teams.

These tests have now been carried out on a sample comprising 299 books, representative of the difficulties inherent in the selected documents, using the presently available technical means. Their goals have been to provide factual elements in order to reveal information about the performance (output) and quality of the different production lines with the aim of supporting the recommendations for setting up an industrial-type digitisation and conversion line. They have been carried out according to a tests protocol, defined at the beginning of the study in order to describe the content and the realisation modalities of the tests and to allow the comparisons on the results given by the providers.

The observations carried out on the digitisation and conversion line led to a set of recommendations and orientations on each of the following fields:

- the organisation and implementation of quality control.



Quality control is a decisive element for the BnF goals and must be conceived globally, by first defining general objectives, then by inferring the objectives in each field while making clear the responsibilities of each actor. It must be accompanied by a defect management policy and take into account the time dimension in order to be able to take into consideration the predictable evolution of problems in regard to technological advances.

- the operations carried out beforehand at the BnF for organising and managing the digitisation of books: selection of contents, preparation of batches, dispatching.

Regarding the line organisation, the strategy of document selection and of batch preparation will have a strong impact on costs, particularly in determining the level of homogeneity of the batches (a high homogeneity is desirable) and the quality of the books to be treated. For maintaining a strong content selection policy, it will certainly be necessary to use several selection processes, in order to define a structuring "core" for the selected topic, representing 10 to 15% of the total number of books put online, completed by relevant books chosen by taking more into account the criteria which have an impact on costs. This can lead to a change in the present processes and organisation.

- Transport and preparation of digitisation (particularly classification of books and selection of a digitisation process).

Once the documents have been selected, the technical means used for identifying the books and trace their path must allow an automation of the line, at least partially. The use of bar codes or RFID labels is therefore recommended.

- the digitisation processes as such (image taking)

The digitisation tests have used automatic scanners (4DigitakBooks/I2S and Kirtas) and manual scanners (Digibook, Copibook particularly), in the most recent available versions. These tests have shown that 60 to 80% of the books could be treated on automatic scanners, with a flow of approximately 800 pages per hour and per scanner, or up to 3,200 pages per operator. The maturity level of the service providers on organisational aspects and mastery of equipment is satisfying and globally equivalent.

- the text extraction processes (OCR and related treatments)

In the field of OCR and of conversion of images into text, however, the results are less convincing<sup>1</sup>, a complete treatment of the sample within the time allowed, has turned out to be problematic. The treatments are still partially automated and a great number of tasks remain manual to a great extent. Different strategies have been explored by the providers. The quality objectives aimed at, however, are realistic and the improvement and industrialisation of the necessary processes are presently taking place.

- the reception of books and of the digitised products at the BnF.

The operations required for integrating the (digitised) books in the BNF information systems will also have to be adapted; the increase of the information volumes to be managed will raise new challenges, particularly regarding the search and navigation functions offered to users.

- the contractual management of relationships between the BnF and its providers.

In the same way, the management of contracts and disputes will have to be organised in order to be adapted to the volume of acts which will inevitably be generated by the digitisation of 150,000 to 200,000 books each year during 5 years.

Eventually, the factors more likely to have an impact on costs seem to be:

- the prior preparation by the BnF, decisive for the homogeneity of the batches,
- the publication date of the books (before 1920, before 1830),
- the quality level required for conversion,
- the staff costs for the manual operations necessary for conversion and quality control,
- and finally, the cost of the scanners.

As a conclusion, even if setting up the infrastructure required for a mass digitisation is in itself a sizeable project, which could take between 6 to 12 months, this study has shown that the installation of a mass digitisation organisation, with the capacity of digitising 200,000 books per year, can today be contemplated, with a cost per page between 0,08 and 0,15 euro.

Many factors, however, have an influence on the real performance of the lines, in which the OCR and text conversion treatments still represent more than half of the cost per page. The document selection strategy probably remains the main decisive element for reaching the outputs and costs aimed at.

### 2.6 The state of OCR Overview by Joachim Korb, Austrian National Library, <u>TELplus project</u>

What follows is based mainly on a survey of European national libraries which are members of the TELplus project, 18 in all, at the end of 2007.

<sup>&</sup>lt;sup>1</sup> The main difficulties are related to the age of the documents (for instance, the books published before 1830 have extremely various qualities and typographies), to the presence of different languages in one document or to the use of very small fonts.

Currently the majority of libraries do not OCR their digitised images. OCR-ing, especially of older material, is still a very new subject. Problems like deterioration of original material and special fonts are obstacles that need yet to be overcome to achieve good quality full-text of digitised materials. As opposed to modern office documents, which are usually printed in standard "Antiqua" type fonts (like the well known "Times New Roman", "Courier", and "Arial" fonts), the old documents the national libraries in Europe currently digitise are often printed in old fonts that are difficult to read, even for most modern readers. Due to the fact that they where mastered on handwriting using a pen with a nib, these fonts (especially gothic, black letter or Fraktur fonts) are made up of dots and lines. Many of these are so similar that even the smallest smudges may lead to confusion. Similar problems occur with hand-written texts.

Given the quality of the original material that was scanned – brittle, dirty, well used paper – these problems are very usual. Often whole parts of the text are unreadable or even missing. Here, where the human reader may successfully guess the missing passages, today's software still cannot match human guesswork. Especially for languages with few native speakers, there are no or insufficient dictionaries to aid this work. Even languages that have a large speaker base have been subject to change or have large numbers of dialects. Thus standard dictionaries are of little help.

Also, as the subject of OCR-ing older material is so new, methods of establishing the quality of OCR results are not yet standardised. Although there is no proper means of objectively comparing results, it can generally be said that results aren't as good one would wish.

The range of OCR-software fitted to the needs of libraries is rather limited. The fact that all TELplus project partners with any experience in OCR-ing use or have used some version of Abbyy's FineReader is an indicator of this. Besides versions 5.0 to 8.0, versions XIX (for Black Letter or Gothic (Fraktur) fonts) and the Recognition Server have been used.

Those libraries or subcontracted service providers that do not use Abbyy's products directly have used some sort of derivative. The National Library of Norway, for example is using docWorks a CCS product. docWorks in turn is based on a variety of FineReader versions. Similarly, the former service provider of the French National Library used its own product, which was in turn was based on a combination of OCR-products – FineReader among them.

Accordingly the results of full-text production are comparable and depend more on the choice of original material than on the quality of the software. The only exception to this is the Martynas Mažvydas National Library of Lithuania. Here the results of OCR-ing are revised by hand, giving almost perfect results. This practice is not feasible for most libraries, though.

### Choice of material to be OCRed within the TELplus project

The majority of material that is currently available with OCR is from newspapers, followed by books and journals. Most national libraries, as is to be expected, use material that is originally from their own country and in their own language. Also not surprisingly, some countries use material from countries they were once part of or linked to. For example, the National Széchényi Library of Hungary uses material from 6 different countries and in 4 different languages (including Latin and English, from modern text books). Similarly, the National Library of Poland is OCR-ing texts not only in Polish, but also in up to 6 "minority languages". By including newspapers and journals from the Faeroe Islands and Greenland, the National Library of Iceland extends the area where the material comes from almost across the Atlantic and Spain includes material from its old colonies in South America.

Just as the origin of the material is spread over a vast area geographically, so it is spread over several centuries in its timescale. In Slovenia the oldest book that is being OCRed is from 1500, while the youngest OCR-ed Icelandic newspaper is only 6 years old. The majority of the OCR-ed material is from the time between the early 19<sup>th</sup> to the mid 20<sup>th</sup> century, though.

### Outlook

As has been seen above, current state of the art OCR is not really ready for the needs of national libraries. In January a new FP7 project called <u>IMPACT</u> started. IMPACT as a network of centres of competence brings together fifteen national and regional libraries, research institutions and commercial suppliers. It aims to significantly improve access to historical text and to take away the barriers that stand in the way of the mass digitization of the European cultural heritage" Within IMPACT the OCR-challenges we are facing will be a major focus.

# 2.7 Summary

For national libraries content digitisation has a strong impact on collection development policies and strategic planning.

If we look at the most promising and successful examples of strategic planning which have already resulted in concrete implementation plans against a solid financial and institutional background in Austria, Estonia, Finland, France, Latvia, Norway, Slovenia, Slovakia, Spain, Russia and the United Kingdom it becomes clear that on an institutional level systematic content digitisation is daily practise in many European national libraries.

Making this content available at European level and developing strategies for co-operation and integration with a European dimension in national and institutional strategies will be major tasks for the future.

To summarize we can say that among the most successful digitisation programs newspaper digitisation is leading and grows constantly. Mass digitisation of books is still related to specific projects in a few libraries. This situation will not change in the near future unless the EU or the Member States change their policies.

# 3 Digitisation Strategies and programmes

### 3.1 Summary

For this report, 37 questionnaires from 37 European National Libraries (out of 47) were collected. 19 national libraries reported having a written digitisation strategy or plans to develop one in the near future. In all libraries with documented digitisation strategies access for users, long term preservation and building up a sustainable digital library infrastructure play a dominant role. Only a few libraries have an explicit and ongoing budget for digitisation (between €100.000 and €200.000 per year). In most libraries the budget for digitisation is either dependent on projects or integrated in overall costs for digital library services and infrastructure.



33% of strategies and programmes focus on national heritage, 44% on European and national heritage, 19% on European content and 4% on non European content.
Priority is given to newspaper digitisation (19 libraries), mass digitisation of printed material (14 libraries), unique and fragile resources, European and national heritage (19 libraries).
Basic obstacles are seen in financing of mass digitisation national and EU level (22 libraries), multilingual approach of the European digital library (including usability and standards (12 libraries), copyright (20<sup>th</sup> century) (9 libraries), and lack of infrastructure, technology and expertise (7 libraries).



If we compare the statistical data with major obstacles given by the National Libraries we may detect three major operational fields that will be crucial for the success of EDL:

Building up a digital library infrastructure based on approved development plans Co-ordinated digitisation of content on European Level Exchange of information on digital content

Based on this evidence the following recommendations can be given:

- National Libraries should build up at institutional level a digital library infrastructure based on approved strategies and development plans
- National digitisation strategies should be based on a territorial approach and should focus on building large digital collections
- Partnerships with other organisations should be formed to create larger virtual collections.
- CENL should continue to maintain close links with the Numeric study team and the CENL Working Group should continue to support the Commission's efforts to develop standardised reporting in the field of digitisation and raise the profile of the related issues with governments and other stakeholders.

# 3.2 Austria:

In Austria at national level the definition of a co-ordinated digitisation strategy is still under development. At institutional level the Austrian National Library has developed a written strategy in which the setting up of a special department and the acquisition and further development of special archiving software (the digitool) are major points. The delivery of online publications, archiving of online media on the basis of an expected amendment to the Law on Media, realisation of web archiving on the basis of a feasibility study and long-term archiving of digital material produced by the ANL itself are the points on which the focus lies.

Also the ANL aims at having by 2011 an electronic search engine that includes all of its holdings. For this purpose the new overall catalogue is to be standardised, free of duplicate copies, lendable, and compatible with The European Library.

Current projects at the ANL include:

- ANNO Austrian Newspapers online
- ALEX portal on historical legal texts
- Picture Archive Austria
- Portrait collection of Picture Archive (digitisation of 200.000 historical portraits)
- Ariadne (Women's movement)
- Audiovisual media (preserving project)
- Collection 1848 (broadsheet collection concerning the "year of revolution" in Vienna)
- Incunabula-Project (140 Bibles)
- Papyri-project (10.000 published papyri)

Over the next five year the ANL plans to digitise:

- Newspapers: 3,5 million pages
- Books: 2 million pages
- Images: 300.000 pages
- Papyri: 10.000 pieces
- Audio files: 4000 (hours)

Among its own material the ANB sees its priorities with historical newspapers, legislative material, incunabula, portraits, papyri and historical audio-records. For the EDL it sees printed European heritage and unique sources of European cultural heritage (books, manuscripts, pictures; audio) as most important material for digitisation.

For the Austrian National Library copyright restrictions, a lack of financial support at national level as well as a lack of financial support for mass-digitisation from the EU are major obstacles for the EDL.

# 3.3 Belgium:

Under the Belgian Digitisation Programme of the Federal Scientific Institutions 10 subprojects have been launched to ensure the start of the digitisation of some collections in each of the institutes. The participants in these programmes are the National Library, The State Archive, 4 museums, 3 research institutes, and the Institute of Art Heritage. The Royal Library of Belgium participates in two of the subprojects: the digitisation of Belgian newspapers and the retro-cataloguing of the catalogue of printed books.

The National Library of Belgium currently has a digitisation strategy under review.

The current Projects at the National Library are:

- o Digitisation of Belgian newspapers
- Digitisation of the first cadastral survey of Belgium
- Digitisation of the manuscripts of the dukes of Burgundy
- Digitisation of parts of the maps collection
- Digitisation of parts of the engravings collection
- Digitisation of parts of the Belgian medals collection
- Digitisation of parts of the manuscripts collection
- Digitisation of 19th century printed brochures

Depending on the outcome of the discussions on the digitisation strategy, the National Library of Belgium expects to digitise the following:

- Belgian newspapers
- Engravings
- o Maps
- Coins and Medals
- Manuscripts
- o Music

When choosing material for digitisation, the National Library gives priority to material in bad condition and heavily-used objects. Digitisation efforts for the EDL should, according to the National Library of Belgium, focus on the specific European cultural and documentary heritage.

Lack of financial resources, too little experience with digital preservation, and the principle of subsidiarity concerning cultural policy are the main obstacles on the way to the EDL.

# 3.4 Bulgaria:

In Bulgaria there is currently no national programme for digitalisation in place. Also the National Library has no written strategy, but a pilot project to set up a digitisation facility is in progress.

Depending on the outcome of this, the BNL is planning on digitising the following:

- manuscripts (Slavonic and Oriental)
- old printed books
- archives (Bulgarian and Oriental)
- photographs
- o newspapers

Manuscript documents, archive documents, photographs and some newspapers have priority in the Bulgarian National Library's choice of material for digitisation. The EDL's priorities should lie with unique documents that are national monuments of culture and a common European significance of the documents digitised.

For the representative of the Bulgarian National Library the greatest obstacles on the way to the EDL are irregular investment in the process of digitization in particular countries and difficulties in the standardisation of criteria for the selection of materials for digitisation in particular countries.

## 3.5 Croatia:

The Croatian National Heritage Programme<sup>1</sup> is a national programme under the responsibility of the Ministry of Culture. It aims at achieving coherence amongst all the country's memory institutions, a national infrastructure which supports storage, long term preservation and searching and retrieval of digital content. By influencing funding policies and programmes and by co-operating with other heritage institutions and stakeholders of digital content, as well as the education sector, systematic co-ordination and benchmarking, the setup of competence centres and as well as a national digital archive are to be put into place. Also a portal of the digitised cultural heritage is to be set up.

The National Library of Croatia (NBC) has a written strategy for digital content<sup>2</sup>. Under it the NBC wants to increase access to and usage of its collection, enhance the use of frequently demanded materials, preserve valuable originals and integrate digitised reproductions into the e-learning programmes. Also the NBC plans to co-operate with national and international digitisation projects in order to create virtual collections.

Current projects include:

- o the Collection of Manuscripts and Rare Books
- o the Print Collection
- the Maps and Atlases Collection
- the Collection of Music and Audio Material
- the old newspapers collection

Over the next five years the following material is expected to be digitised:

(continued)

- o the Collection of Manuscripts and Rare Books
- the Prints Collection
- the Maps and Atlases Collection
- the Collection of Music and Audio Material
- $\circ$  the Croatian historical newspapers collection

(new)

- Croatian scientific journals
- Croatian books published after 1835

The National Library sees frequently used and fragile items, in particular Croatian old newspapers, as its priority when choosing material for digitisation.

<sup>&</sup>lt;sup>1</sup> <u>http://daz.hr/bastina/</u> (Croatian)

<sup>&</sup>lt;sup>2</sup> <u>http://www.nsk.hr/HeritageUnits.aspx?id=197</u> (Croatian)

# 3.6 Cyprus:

The Cyprus Cultural Portal<sup>1</sup> is a national programme, under the control of the Cultural Services of the Ministry of Education and Culture and Department of Antiquities, that focuses mainly on cultural artefacts. Its objective is to include systematic and rescue excavations as well as archaeological surveys, the establishment, management and operation of archaeological museums, the conservation, restoration, protection and promotion of Ancient Monuments in the First and Second Schedule of the Antiquities Law, of archaeological sites and of monuments of architectural heritage. Hence the National Library is not part of this programme.

The National Library itself does not currently have a digitisation programme, nor does it have a written strategy for the future.

A number of Cypriot institutions have material worth digitising, but as yet there is no timescale for their digitisation.

For the National Library newspapers and old books should have priority when choosing material for digitisation.

On the European level the National Library sees that libraries, museums and archives would have different priorities in the material that is digitised, but proposes that libraries, with mainly researchers in mind, should concentrate on the digitisation of material that is old or out of print.

When building the EDL, the National Library of Cyprus sees its contribution hampered by the absence of technical support and infrastructure, inadequate personnel, inadequate funding, and lack of proper building facilities.

<sup>&</sup>lt;sup>1</sup> <u>http://www.mcw.gov.cy/mcw/da/da.nsf/DMLindex\_en/DMLindex\_en?OpenDocument</u> (Greek / English)

# 3.7 Czech Republic

In the Czech Republic no explicit national programme is currently in place, but the state through the Ministry of Culture funds the National Library's efforts.

The National Library itself has a written strategy<sup>1</sup> which focuses on technical standards. Its objectives are digitisation for better access and indirect preservation (through exclusion of works from physical services).

Currently the National Library of the Czech Republic has two major projects:

- Memoria (Memoriae Mundi Series Bohemica; Manuscriptorium Digital Library)
- o manuscripts
- rare printed books
- historical maps

Kramerius

- o acid-paper library materials
- o newspapers
- o journals
- late 19<sup>th</sup> century monographs

Over the next five years around 500.000 pages of Memoria material will be digitised. Among these, acid-paper materials will have priority.

According to the Czech National Library for the European Digital Library mass digitization should have priority.

As obstacles in building the EDL the National Library's representative sees that the principles on which it should be build are not yet known. He believes that as for now, global talk about EDL is not concrete and may be sometimes naïve or rather too general and political, but that the EDL will might become similar to the European Library in many respects. Some seamless access will be possible where it is not in contradiction to the licensing policies of individual digital libraries.

<sup>&</sup>lt;sup>1</sup> <u>http://digit.nkp.cz/techstandards.html</u> (Czech / English)

## 3.8 Denmark:

In future Danish digitisation efforts will be coordinated by the Danish Ministry of Culture. The ministry has appointed a committee to work on a national strategy for digitisation. The work is focusing on digitisation for improved access, digitisation for preservation, and copyright and IPR-issues in regard to digitisation.

The Danish National Library currently has no written strategy on digitisation, but the following material is being digitised:

- o monographs
- o manuscripts
- $\circ$  rare books
- o music
- photographs
- o maps
- $\circ$  serials

Over the next five years the following material is expected to be digitised:

0	monographs (50 books):	10.000 pages
0	manuscripts and rare books:	34.000 pages
0	music:	17.000 pages
0	photographs and maps:	19.000 items
0	serials:	26.000 pages

These are also the collections that need priority in digitisation.
# 3.9 Estonia:

The Estonian Ministry of Culture has put the Digital Cultural Heritage 2007-2010<sup>1</sup> programme in place. The programme's objectives are to develop a coordination network to lessen double investment, while

finding additional financial resources. Technical points of the programme lie in establishing common criteria and quality requirements,

developing common solutions for a long-term storage and

establishing common rules and developing necessary facilities for the use of digital content . Also there is the aim of integration into EU development projects.

The Estonian National Library's digitalisation strategy<sup>2</sup> was written to achieve the creation of a digital archive of digitised cultural heritage. Through a national cultural heritage portal the material will be made accessible via the internet.

Currently the Estonian National Library is working on two digitisation programmes:

- DEA (digiteeritud eesti ajalehed = Estonian newspapers)
- DIGAR (Digitisation of Cultural Heritage)

For the next five year the following material is expected to be digitised:

	5 6 1	U
0	Old maps:	350
0	Old Estonian newspapers(until 1945):	100 titles
0	Old Estonian magazines (until 1945):	420 titles
0	Old Estonian books, published before 1917:	3.300 titles

For the National Library of Estonia multilingual access and preservation should be the priorities in digitisation for the European Digital Library, while common standards, cooperation with other cultural institutions and financing are its greatest obstacles.

<sup>&</sup>lt;sup>1</sup> <u>http://www.kul.ee/webeditor/files/Digi\_Kult\_AK\_loplik.pdf</u> (Estonian)

<sup>&</sup>lt;sup>2</sup> <u>http://www.nlib.ee/67367</u> (Estonian)

#### 3.10 Finland:

In Finland the National Digitisation Centre<sup>1</sup> has been coordinating ALM (archive, library and museum) collaboration eKAM since the beginning of 2003. It thus continues the work begun by the Finnish cultural heritage committee (KULDI). Set up by Ministry of Education it has thematically structured sub-groups: e.g. the selection group is working on a pilot project around common thematics. The National Digitisation Centre coordinates and leads the Digitisation Programme for the Libraries<sup>2</sup> that is based on the findings of an inquiry into the digitisation needs of the scientific and public libraries.

Due to this structure the National Library has no written strategy of its own. But current efforts centre on the following material:

- Historical Newspaper Library
- Scholarly Journals
- Ephemera

Due to funding issues no proper estimates for the next five years can be given, but the goal is to produce more than 1.5 million pages and 4000 digitised sound recordings in one year. Material to be digitised during the next five years could include the following:

	0	Newspapers 1890-1910:	10 million pages			
Books:						
	0	church & religion:	3,2 million pages			
	0	school & University:	0,5 million pages			
	0	history:	0,5 million pages			
	0	collections & series:	1,2 million pages			
	0	novels:	9,8 million pages			
	0	language & poetry etc:	2,0 million pages			

For the Finnish National Library historical Finnish newspapers, national Collections in bad condition and scientific journals have priority in digitisation.

That national digitisation is funded on an adequate level, common solutions are developed in Europe and co-operation is enhanced should be the priorities when building the EDL. Lack of funding and co-ordination are seen as obstacles on this way.

<sup>&</sup>lt;sup>1</sup> <u>http://www.minedu.fi/OPM/Julkaisut/2003/kulttuuriperinto\_tietoyhteiskunnassa\_strategiset\_tavoitteet\_ja?la\_ng=fi&extra\_locale=en</u> (Finish / English)

<sup>&</sup>lt;sup>2</sup> <u>http://www.kirjastot.fi/fi-FI/kirjastoala/julkaisut/digitaalinen\_tietopankki/</u> (Finish)

#### 3.11 France:

France has, since 1996, embarked on a vast digitisation plan of its collections. This coordinated program concerns all of the domains of intervention of the Ministry of Culture: archives, heritage, architecture, museums, arts, music, books and audiovisual. With its National Digitalisation Plan' France is a precursor of the EDL project. The French National Library developed the prototype Europeana for the future European Digital Library and is actively engaged in European projects, esp. inTELplus.

The Gallica Digital Library Charter<sup>2</sup> is the French National Library's written strategy. Launched in 1997, the Gallica Digital Library is one of the most extensive collections in Europe. Other priority programmes have been defined, among them a mock-up and prototype for the European Digital Library: Europeana. Gallica is mostly constructed from existing library collections. However, some partnerships and collaborations with French institutions allow having many shared digitisation programmes.

Besides the ongoing work and the further evolution of Gallica, current programmes of the French National Library are:

- the Newspaper Digitisation Project
- o the contribution to the European Digital Library
- the coordination of the Réseau Francophone des bibliothèques nationales numériques (French-speaking Network of the Digital National Libraries)
- o the long-term preservation programme

For the next years 100,000 printed documents in image and text mode per year are planned.

For its own project as well as for the EDL the French National Library sees the conservation of national heritage in the tradition of Gallica as priority. When choosing material for digitisation international relations in Europe, the circulation of people and ideas within the European space (including morals, migrations, tourism, transports, economy and trade), the relations of Europe with other areas as well as language related topics (translations, collections of works in a countries language) should be part of the thematic focus.

Financial issues are seen as the EDL's biggest problem, esp. the financing of mass digitisation of books.

<sup>&</sup>lt;sup>1</sup> <u>http://www.numerique.culture.fr/mpf/pub-fr/index.html</u> (French)

<sup>&</sup>lt;sup>2</sup> <u>http://www.bnf.fr/pages/zNavigat/frame/version\_anglaise.htm?ancre=english.htm</u> (French, English, Spanish)

# 3.12 Germany:

In Germany digitisation programmes are left to the federal states, but a number of national agencies have digitisation programmes. For example the DFG (Deutsche Forschungsgemeinschaft) has recently published a comprehensive manual<sup>1</sup> comprising detailed rules and regulations applicable to digitisation projects that receive DFG funding. The use of open standards and non-proprietary software is an important point of these recommendations which support archives, libraries and museums in their digitisation efforts. Next to standards for actual digital images, a variety of international metadata standards are in use. The DFG projects emphasis is on digitising historical book, newspapers, prints etc.

The German National Library, while not yet having a written strategy (though one is in the making), has its focus in the 20<sup>th</sup> century. Current projects are:

- o digitised periodicals and newspapers of the German Exile archive 1933-1945
- digitised Legal Sources of the Era of the Allied Control Council for the Territory of the later Federal Republic of Germany: 1945-1949
- o Jewish periodicals during the NS regime in Germany

For the next five years digitisation of the following is planned:

- The collection of the music information centre of the (former) German Democratic Republic
- Books
- Newspapers
- Audio (music and speech)

For the German National Library the establishment of high-throughput digitisation workflows while maintaining the heritage's quality level of metadata annotation, digitisation of multimedia content, automatic asset quality processing, automatic metadata generation, the use of both automatically and intellectually generated metadata to optimise asset retrieval, using means of semantic web and interaction with users by opening knowledge networks should be priorities when building the European Digital Library.

Finding appropriate technologies for knowledge retrieval, establishing a business plan to attract as many users as possible and finding a compromise for users' access to cultural heritage and protecting the holder of a copyright are seen as the main obstacles on the way to building the EDL.

<sup>&</sup>lt;sup>1</sup> <u>http://www.dfg.de/forschungsfoerderung/formulare/download/12\_151.pdf</u> (German)

### 3.13 Greece:

Digitisation efforts in Greece concentrate on archaeological items and items from museums. The Hellenic Ministry of Culture sets up a main portal<sup>1</sup> to Hellenic cultural content. A similar one is planned by the National Audiovisual Archive for audiovisual content.

The Greek National Library does not have a written strategy for digital content, but it does have several digitisation programmes:

- Digitisation of Greek newspapers published in Greece and one published in Egypt, where there was a very big Greek community.
- Digitisation of the Year Books of the Philosophical Schools of the Universities of Athens and Thessaloniki 1925-1977
- o Digitisation of catalogues of manuscript codices

For the next five year the Greek National Library plans the digitisation of newspapers and manuscripts, with old newspapers as its main priority.

<sup>&</sup>lt;sup>1</sup> <u>http://www.culture.gr</u> (Greek, English)

#### 3.14 Hungary:

The Hungarian Ministry of Education and Culture's National Digitisation Plan for Libraries<sup>1</sup> treats the notion of library digitisation in its widest sense: besides digitisation in the narrow sense it includes also questions regarding the collection, long-term preservation, processing and publishing of digital objects. The plan considers digital documents (images, sound, texts, and motion pictures) as new information carriers, which have special attributes diverging from those of traditional library documents. In 2007 two competence centres are planned to be set up: one in the National Library functioning as a national knowledge centre for library digitisation, the second in the Neumann Digital Library concentrating on digitisation of audiovisual (broadcast media), museum and archival material.

The National Széchényi Library's (OSZK) aim<sup>2</sup> is to construct a digital archive of off-line and on-line electronic materials falling into its collecting scope and belonging to cultural heritage, and to selectively digitise its main collection and special collections. The digital collection should be processed for, reflected and retrievable in the library's main catalogue. The solutions applied in the process of digitisation should correspond to standards in order to ensure high quality.

The National Library's digitalisation programmes focus mainly on old and valuable books as well as on chronologically digitising certain object types (e.g. postcards, musical manuscripts, leaflets of the Hungarian Revolution).

For the next five years digitisation of the following is planned:

- Obituary notifications 1711-2003: ca. 600.000 objects
- Old Hungarian books to 1850: 10.000 books, 15.000 pages
- Old Hungarian periodicals, newspaper and social sciences journals to 1920
- Postcards collection by counties of the Greater Hungary before 1920: 200.000 images
- Collection of Hungarian posters before 1914.
- Legacies of famous writers, musicians, manuscripts.
- Old Hungarian codices
- Most endangered and highly-used books

The most endangered and highly-used books are also of high priority when choosing material for digitisation, along with old Hungarian books (to 1850), periodicals, newspapers and social sciences journals (to 1920).

The EDLs priorities should lie with diversity of documents from the European culture and diversity of languages from the European culture as well as with useful and attractive content.

As the main obstacle for the establishment of the EDL the Hungarian National Library sees copyright regulations and financial sustainability.

<sup>&</sup>lt;sup>1</sup> <u>http://www.ki.oszk.hu/107/download.php?list.45</u> (Hungarian / English)

<sup>&</sup>lt;sup>2</sup> <u>http://www.oszk.hu/hun/szakmai/tervek/digitstrat/digitstrat\_terv\_hu.htm</u> (Hungarian)

### 3.15 Iceland:

Iceland has two national digitisation programmes, one under the auspices of the Prime Minister's office, called Resources to serve everyone<sup>1</sup>, and one by the Ministry of Culture, called Risk with responsibility<sup>2</sup>. While the first is a more general programme, the second has the educational sector as its focus. It supports giving open access to computers and internet in schools and libraries, states the need to have libraries provide the public with effective assistance in the use of digital resources, and for all teachers to acquire the requisite skills to use ICT at work.

In the first programme the National Library (NB) plays an important part, so its strategy<sup>3</sup> follows the ideas put forward in that programme closely. A main point of this strategy is to publicise and facilitate access to the Icelandic written cultural heritage, by making it available beyond the walls of the Library. This includes web harvesting as well as systematic digitalisation of all of the NB's material. This is done in full cooperation with those that hold the copyright to the material in question. The preservation of physical as well as digital material are as important to the NB as its service to the users and making Icelandic material accessible on the Internet.

Current projects of the National Library include:

Stafrænt þjóðbókasafn (Icelandic National Digital Library):

- Forn Íslandskort (Antique maps of Iceland)
- Sagnanet (Saganet)
- Blöð og tímarit (Journals and Newspapers)

Planned for the next five years is the digitisation of the following:

0	National newspapers 1920-2000 (about 10 titles):	1.250.000 pages
0	Icelandic books 1530 – 1900 (about 1.600 titles):	400.000 pages
0	Icelandic books 1900 – 1924 (about 4.300 titles):	870.000 pages
0	Manuscript collections:	about 300.000 pages
0	Local Newspapers 1920 – 2001:	about 400.000 pages
0	Icelandic journals:	about 250.000 pages
In add	lition the following is planned depending on funds:	
0	Icelandic journals 1920 – 2000:	about 5.000.000 pages
0	Icelandic music 1900 – 2000:	about 4000 records/CD's
0	Maps published 1900-2000	
0	Other manuscript collections (13,000 volumes):	about 3.300.000 pages
0	Posters	

• Books 1900 - 2000

Newspapers, old books, manuscripts, sound recordings, and journals are the Iceland National Library's priority in digitisation.

<sup>&</sup>lt;sup>1</sup> <u>http://eng.forsaetisraduneyti.is/media/English/IT\_Policy2004.pdf</u> (English)

<sup>&</sup>lt;sup>2</sup> <u>http://bella.stjr.is/utgafur/risk.pdf</u> (English)

<sup>&</sup>lt;sup>3</sup> <u>http://www.landsbokasafn.is/id/1020710</u> (page is currently empty)

In general old material, heavily used material and material that is difficult to access should be prioritised in building the EDL. The focus should be placed on material like manuscripts, newspapers and journals, graphic material, sound recordings, and books.

As the EDLs main obstacles the National Library sees funding the digitisation process and providing the necessary metadata as well as copyright issues.

# 3.16 Ireland:

Ireland currently has no national programmes for digitisation, but its National Library has a digitisation strategy under review.

Digitisation work done at the NB consists of sample digital images of items from the Prints and Drawings and the Photographs collections, to accompany item records in the online catalogue.

Plans for the future are still under review so neither material nor concrete figures have been given in the survey.

According to the Irish National Library when building the EDL, there should be a focus on unique material. Also the following should have priority: A plan for digital preservation, the creation of adequate metadata to recognised standards, the interoperability of digitisation activities across Europe, and the identification and engagement of users and ongoing usability testing/user evaluation/measurement. The EDL should become an access portal designed for usability and in line with user expectations; leveraging new technologies to add value to user experience and enhance interactivity.

As the EDLs main obstacles the National Library sees a lack of shared standards and consequent interoperability issues, the disparity in level and nature of digitisation activity across Europe, and the inadequate promotion of initiative to participants and users.

# 3.17 Italy:

Italy has no single national programme at the moment, but the sector of heritage has a national reference point in SIGEC, the General Information System of the Catalogue prepared by the Central Institute for Catalogues and Documentation – ICCD<sup>1</sup>. The Ministry for Cultural Heritage and Activities oversees a set of programmes known as Biblioteca Digitale Italiana (BDI)<sup>2</sup>, in which also the National Library in Rome takes part.

Of the two National Libraries in Italy data was only received from the Biblioteca Nazionale Centrale – Roma. While it does not have a written strategy, it carried out its programmes in compliance with the BDI national programmes.

Currently the National Library in Rome is working on:

- Biblioteca Nonantolana Virtuale: Reconstruction and digitisation of the manuscript library of the mediaeval abbey of Nonantula
- Progetto Siciliani: the collection (books, scores and sound records) of Francesco Siciliani, a distinguished XX century musician.
- RIMI digitalizzazione: fragile and often consulted XIX and XX century Italian serials

In the next five years the National Library in Rome plans to digitise sound recordings and XIX and XX century serials. Here its priority lies with old serials and periodicals, because they are fragile and often consulted.

In building the European Digital Library the National Library's main priority would be the exchange of knowledge, coordination and cooperation, while shortage of staff and funding, and the problems of maintaining of a digital library as obstacles to this process.

<sup>1 &</sup>lt;u>http://www.iccd.beniculturali.it</u> (Italian / English)

<sup>&</sup>lt;sup>2</sup> <u>http://www.iccu.sbn.it/genera.jsp?s=18</u> (Italian / English)

# 3.18 Latvia:

The main document which provides guidelines for state growth for the next seven years is the Latvian National Development Plan 2007-2013. This is a very far reaching plan in which the Latvian National Library (NB) plays a leading role. During the plan's execution the NB will receive a completely new building and Latvia will improve and develop the existing state information systems by integrating their mutual operation in a unified state system, to ensure the online availability of public services, to facilitate the implementation of e-government and the optimisation of state and local governments' work. Part of this change in the IT infrastructure will be the Latvian Library Information System. The aim of the project is to create a coordinated information system of state and public libraries, create a Latvian library portal, provide library staff training in order to enhance their skills and quality of library services, and develop internet connections in all libraries.

The Latvian National Library has a digitisation strategy that focuses on common standards and common quality criteria. Its aims are the preservation of Latvian cultural heritage for future generations, making it widely accessible and to support for life-long learning and education.

The Latvian National Library currently digitises:

- Heritage 1: Latvian Newspapers 1822-1945
- Latvian Poster
- Latvian Postcards
- The Picture Collection of the Baltic Central Library
- History of Latvian culture in images. Series A. : Portraits
- Latvian Graphics: prints and book plates
- Latvia on maps
- Latvian symphony

And for the National Digital Library of Latvia Letonica

- $\circ$  all of the above from 2007
- Jazeps Vitols pilot project
- In Search of Latvia Lost (to begin in June 2007)

In the next five years digitisation of the following is planned:

All of the above will be continued. Several to be begun:

- Parliamentary (Saeima) Transcripts (text, to be OCR-ed)
- o In Search of Latvia Lost (mostly images-photographs, postcards etc.)
- History of Latvian Cinema (text, images, video files)
- Latvian Sound recordings (audio)
- History of Latvian Libraries and Library Science (text, images)
- Riga Through the Centuries (maps, plans, images, books)
- Digitization on Demand service (mostly books and manuscripts)

At the Latvian National Library the collections that need priority are Latvian Parliament (Saeima) Transcripts, The Presidential Library (full texts, images (later – audio and video) files of speeches, interviews, and other documents of Latvian presidents), In search of Latvia Lost (images), Latvian Newspapers, and Latvian maps (geo-referenced maps).

<sup>&</sup>lt;sup>1</sup> <u>http://www.lndb.lv</u> (not yet online)

For the EDL priority should be given to collections that reflect common European identity and cultural heritage and large volumes of indexed full text as well as strong support for educational content and lobbying for copyright law on the EU level.

As the main obstacle in building the EDL the National Library sees the slow adoption process of relevant standards (e.g. metadata), a lacking of environment for implementing the standards, unsolved copyright matters, and the slow adoption of multilanguage support. Also the National Library believes that other memory institutions (such as museums and archives) are finding their own solutions.

### 3.19 Liechtenstein:

Liechtenstein currently has no national programme for a digital library and the National Library has no strategy on digitisation.

The National Library has one project called intelligentCapture. It aims are the scanning of tables of contents and index enhancement through computer-aided indexing supported by Thesauri.

In the next five years the National Library of Liechtenstein plans to digitise the following material:

- o Liechtenstein literature
- Annuals of the Historical Association of Liechtenstein (Historischer Verein für das Fürstentum Liechtenstein)
- graphic collection
- o dissertations

The Annuals of the Historical Association of Liechtenstein (Historischer Verein für das Fürstentum Liechtenstein) and the graphic collection need priority in the National Library's choice of material for digitisation.

According to the National Library of Liechtenstein instant easy access to information (historical, cultural) which is not publicly available should be made the priority for building the EDL. Legal reasons and/or financial reasons are seen as obstacles to the EDL.

#### 3.20 Lithuania:

The written strategy of the Lithuanian National Library follows the Government approved Conceptual framework of Digitisation of Lithuanian Cultural Heritage and the National *Bibliography.* Besides the selection of the most significant published and manuscript cultural heritage objects to be digitised, cooperation among memory institutions in further developing the established Integral Virtual Library Information System is of central importance. Through it compatibility of metadata, management of the digital content and its long-term preservation and access interoperability are pursued.

Currently the NL is working on:

- Creation of the Integral Virtual Library System
- Digitisation of old Lithuanian shellac records

For the next five years digitisation of the following is planned:

- books in Lithuanian from 1547-1917:
- around 5.000 objects around 3.000 objects • periodicals in Lithuanian from 1823-1940: • periodicals in other languages published in Lithuania: around 250 objects  $\circ$  books in Polish from the 16th – 18th c.: around 3.000 objects  $\circ$  books in Latin from the 16th – 17th c.: around 250 objects • Lithuanian shellac records: around 1.100 objects

In the Lithuanian National Library's opinion uniqueness, content and worth as well as physical condition and age should be the priorities in digitisation to build the European Digital Library, while multilingual and too complicated access are seen as its main obstacles.

# 3.21 Luxembourg:

In Luxembourg the Ministry of Culture is currently working on a national programme for digitisation.

The National Library has a written strategy<sup>1</sup>. Its goals are ease of access, if possible remotely via the internet and the promotion of the printed heritage. To achieve this the NL will start with historically valuable and fragile originals and then include collections that have a large public interest. It will digitise in the best possible quality and structure the content extensively, so that best possible indexes and searchability is guaranteed.

Currently the National Library works on Luxemburgensia-Online, for which it digitises postcards, periodicals and monographs.

During the next five year the following will be digitised:

• Newspapers and periodicals: 100-200k pages a	
	0

• Monographs: 10-50k pages a year

Among these fragile newspapers should have priority.

According to the National Library of Luxemburg provision of excellent structural metadata, OCR and search engines to enable search across and the inclusion of fairly recent material (at least up to 1950) to foster public interest should be made priorities in building the EDL. The standardisation and availability of high-quality structural metadata, search across several languages and the availability of actual content due to rights issues are the main obstacles to the EDL, as is the problem of funding.

<sup>&</sup>lt;sup>1</sup> <u>http://www.luxemburgensia.bnl.lu/project1\_2/project.html</u> (French)

#### 3.22 Macedonia:

In Macedonia a team of librarians and IT personnel is working on the priorities for digitisation of the collections that are part of library holdings. No digitisation work has been done as of yet, but for the next five years the digitisation of old Slavonic and oriental manuscripts and of the music collection is planned. Of these the collection of old Slavonic manuscripts and the music collection have priority.

For the Macedonian National Library the construction of a good infrastructure and its unification on a European level (network, servers, telecommunication, data transfer speed) and a common agreement on selection of material that should be digitized should have priority while the EDL is being built, while European cultural heritage and European publishing production with scientific content should be the material of choice.

The National Library sees the need to be a European Union country and missing financial sources as its main obstacles in helping to build the EDL.

#### 3.23 Malta:

In Malta no digitisation work has yet been done. No national programme is in place, but, depending on funding, the National Library plans to digitise the following over the next five year:

0	Manuscripts and Books:	3,000
0	Maps and Plans:	200
0	Newspapers	1,000

For the National Library of Malta preserving materials of high importance, reduce damage to original material and increasing access to collections have priority in building the EDL, while a lack of funds and of trained staff, non-standard equipment and deteriorating collections are seen as its main obstacles to partaking in the process.

#### 3.24 Moldova:

In Moldova the Information Technologies Department is working on a programme called Electronic Moldova<sup>1</sup>.

The National Library has as yet not done any digitisation work, but its representative says that it has a written strategy<sup>2</sup>.

For the next five years digitisation of the following is planned:

- $\circ$  books
- o series
- $\circ$  images
- o manuscripts
- $\circ$  audio-visuals
- o documents
- o etc.

The National Library of Moldova sees the many European languages as an obstacle to building the EDL.

<sup>&</sup>lt;sup>1</sup> <u>http://Gov.md/content/ro</u> (link seems to be dead)

<sup>&</sup>lt;sup>2</sup> www.bnrm.md (link leads to the National Library's main web-site)

### 3.25 The Netherlands:

In the Netherlands the National Library is the coordination partner in two ambitious programmes, with others being worked on by other institutions. The programmes in question are called *Memory of the Netherlands* and *National Program for the mass digitisation for the humanities* respectively. The Memory of the Netherlands contains 55 digitised collections from 50 different heritage institutions, while the second national program mentioned above is still under development but already contains: Digitisation Parliamentary Papers, Digitisation Dutch Newspapers, Digitisation Special Collections.

The National Library's strategy for digitisation is to digitise most used, most important materials.

For the next five year digitisation of the following is planned:

• newspapers and pages of pre-1800 books:

millions of pages 1.3 million pages

- press communication from the Dutch press agency:
- 'best-of' digitisation project with highlights from its collection

In digitisation as well as in building the EDL the National Library of the Netherlands gives priority to collections whose material is much used, has a large and diverse target audience (both cultural as well as research), and has sufficient metadata.

The National Library sees the lack of digitised objects and insufficient metadata as the main obstacles to the EDL.

### 3.26 Norway:

The Norwegian government's policy aims for IT development and how it is planned to achieve these aims are set out in the report eNorge 2009. The interim report states that "The digitization of the sources of culture and knowledge is an important step in the democratization of society. The creation of access to the materials includes both the actual conversion of these to a digitized format and the establishment of standards for searches and presentations.

In addition, the use of restricted materials should be clarified with the rights holders. The work is demanding, and there will be a need for considerable efforts over the next several years, both in terms of equipment and organization. Comprehensive programmes of digitization have been initiated under the auspices of the National Library and the Arkivverket1.

The Norwegian Archive, Library and Museum Authority (ABM-utvikling) plays a key role in the ongoing development of a national strategy for work on digitization in this domain.

In summer 2006, The Norwegian Archive, Library and Museum Authority published the report *A Cultural Heritage for all*<sup>1</sup>. The report is about digitization, digital preservation and digital dissemination in the archive, library and museum sector and was made by a Working Group on Digitization and is currently under discussion.

Currently the National Library is digitising the following types of material:

- $\circ$  Books
- Newspapers
- o Photos
- Sound recordings

For the next five year digitisation is planned of:

- Books: 97.000
- Newspapers: 8.000.000 pages
- o Journals: 299.800
- Photo: 260.000
- Radio: 310.000 hours
- o Manuscripts: 149.000 pages
- Posters: 45.000

To the Norwegian National Library, in building the EDL it is important to obtain a mix of older out-of-copyright material and newer copyright protected material. The national libraries should share experiences with making agreements for access with the copyright holders. On the other hand multi-linguality is seen as a major challenge. Also, it is seen as important to obtain a critical mass of accessible digital contents.

<sup>&</sup>lt;sup>1</sup> <u>http://www.abm-utvikling.no/publisert/ABM- skrift/2006/32\_eng\_web.pdf</u> (English)

# 3.27 Poland:

In June 2006 the Minister of Culture and National Heritage established the Committee for Digitization. This assembly, consisting of high quality specialists representing different memory institutions, is responsible for coordinating and governing the process of digitisation in Poland.

The task of the Committee is to gather information about current and planned digitisation initiatives of all kinds of cultural material in different institutions, and to create overviews of such digitisation in order to prevent duplication of efforts and being able to indicate the predictable increase of digitised material in future.

Secondly, it develops quantitative and qualitative targets for digitisation, minding that all digitizing efforts in memory institutions should be purposeful, useful and efficient in order to ensure the preservation of digitized cultural heritage and its use for the benefit of the whole society and each of its members. Important role in the Committee works plays the National Library, which has rich digitisation experience and is the host of the National Digital Library Polona.

The National Library itself has a written strategy which includes participation in preparing the national program of digitisation, participation in developing and implementing internationally accepted metadata standards for all digital objects, the coordination of all activities connected with digitisation in Polish libraries and the creation of a database of digital documents and objects to be digitised in future.

In the next five years the digitisation of the following is planned:

- o newspapers
- o journals
- photographs
- o drawings
- $\circ$  postcards
- o notes
- music files
- Polish literature, history and culture

The overall plan is to digitise 250.000 objects.

The newspapers and images collections need priority in digitisation at the Polish National Library.

For the National Library materials relating to culture and history of particular countries, national literatures and its translations into European languages and documents not accessible in the libraries due to security considerations should be the priorities in digitisation to build the European Digital Library, while financing and copyright issues as seen as its main obstacles.

# 3.28 Portugal:

Between February and April 2006 the Operational Programme for the Knowledge Information Society<sup>1</sup> (which manages the FEDER European Funds for this area) launched a contest regarding the development of content projects by using broadband. Several projects were then supported in making accessible digital library contents. Among them the Digital National Library, a project coordinated by the Portuguese National Library.

The National Library does not have a written strategy on digitisation, but digitises a varied range of printed, manuscript, cartographic and other pictorial materials.

For the next five years the digitisation of the following is planned:

- o manuscripts
- o pictorial materials (especially oversize formats)
- o old printed books
- 19<sup>th</sup> century newspapers

At the Portuguese National Library it is believed that unique and rare materials and European language materials should be the priorities in digitisation to build the European Digital Library. Financial constraints of the NLs for mass digitization processes and for building large scale digital archives, a lack of generalised technical expertise on digitization, and a lack of know how for large IT infrastructures to support local digital archives are seen as the main obstacles to progress.

<sup>&</sup>lt;sup>1</sup> <u>http://www.posc.mctes.pt/</u> (Portugese)

### 3.29 Russia:

Russia has the ambitious National Digital Library<sup>1</sup> programme by the Federal Agency for Culture and Cinematography in place. The programme targets the creation of a trusted digital repository for e-documents (digitised or born-digitally) submitted to it by the National Library of Russia (St. Petersburg), the Russian State Library (Moscow) or other interested parties (including private persons) and to make them available to all users. The National Digital Library's tasks are to promote shared technology architectures and standards, to develop and implement digital library standards, to provide a reliable and robust infrastructure for information technology throughout the Russian libraries, and ensure the continuity and evolution of software assets by building and using modular, reusable, documented software components.

Both National Libraries have written digitisation strategies<sup>2</sup> that work hand-in-hand with the national programme.

Beside the national programme both National Libraries have separate programmes:

RSL:

- Retrospective conversion of card catalogues
- OREL (Open Russian Electronic Library: digitization of all types of materials without copyright restrictions)
- Electronic Leninka (digitisation of materials that fall under copyright restriction and digital copies of which are accessible locally or via special subscription)

NLR:

• Digital History of Russia

Over the next five years digitisation of the following is planned:

RSL:

- Annual output is about 25.000 theses.
- Annual projections for digitization of books are 1.600 titles per year.
- Projected annual output for a number of exposures is 1.000.000.000 for other types of materials (maps, for example, that require several exposures for one document). Other projects are totally dependent on funding that is most unpredictable.

NLR:

- Collection of country councils (zemstvo) publications (1864-1917): approx. 100.000 items
- Publications about history of the Russian nobility, published before 1917
- The First Russian maps collection of the 18th century maps and atlases
- Collection of regional eparchies newspapers (1860-1917): approx. 3.000.000 pages
- Thematic collections of old prints, engravings and photographs
- Additions to «Complete statute book of Russian Empire» (1649-1913)
- Collection of Russian dissertation theses

<sup>&</sup>lt;sup>1</sup> <u>www.runeb.ru</u> or <u>http://www.rusneb.ru/</u> (Russian)

<sup>&</sup>lt;sup>2</sup> NRL: <u>http://www.nlr.ru/e-res/</u> (Russian)

In both National Libraries collections related to Russian history and especially the Collection of Russian dissertation theses have priority in digitisation.

At the Russian National Libraries it is believed that the priorities in digitisation to build the European Digital Library should be national and complete collections of all kinds of materials representations (heterogeneity) on the one hand and valuable or much used collections on the other.

As the main obstacles problems with financing and with the EDL's multi-lingual approach are seen, as are the differences of formats, protocols and standards in use on both the national and the libraries' levels. On the technical side search and retrieval are viewed as among the greatest problems.

#### 3.30 San Marino:

San Marino has no national programme on digitisation, but the National Library of San Marino is currently in the process of planning a digitalisation project. The focus is on the acquisition of adequate machinery for digitalisation and on the choice of adequate material for digitisation.

Depending on this programme in progress, the National Library plans to digitise the following:

- $\circ$  periodicals
- o manuscripts
- Pagine di Bibliografia Sammarinese
- o photographs
- autograph documents

At the National Library of San Marino the standardisation of digitalisation processes and strategies is seen as the point needing priority in building the EDL, while financial problems and the difficult bureaucracy are seen as the main obstacles.

### 3.31 Serbia:

Serbia has no national programme on digitisation.

The Serbian National Library has a written strategy made according to UNESCO - IFLA recommendations. Its aims are the preservation of national cultural heritage, presentation of Serbian culture, cooperation with other institutions (museums and archives) and stakeholders in Serbia to form a National Centre for Digitisation and the application of international standards in metadata and preservation.

Currently the National Library has two projects:

- Digital NBS, digitising the following:
- Serbian children's digital library
- Cyrillic manuscripts
- Old and rare books
- Newspapers and magazines
- Cartographic materials
- Printed music and sound records
- Photographic documents
- Engravings and art material
- Posters and documentary material

DOI Serbia, a full-text archive consisting of 27 locally published scientific journals

Plans for the next five years include the following:

- Serbian literature (out of copyright (first or original or oldest editions: around 10.000 books
- old and rare books from the National Library (printed between 1500 and 1867): around 5.500 books
- Serbian authors' manuscripts from the Special collections of the National Library: around 35.000 items

In the National Library the Collection of Serbian old and rare books need priority in digitisation.

Items with values belonging to the European heritage should have priority in digitisation when building the European Digital Library.

Conflicts between the big library players and a kind of "digitisation divide" among different European national libraries are seen by the Serbian National Library as the main obstacles on the way to the EDL.

### 3.32 Slovakia:

Slovakia has two national programmes: The Programme for the Computerisation of Libraries<sup>1</sup> by the Ministry of Culture and the Informatisation of Society Operational Programme<sup>2</sup> by the Government with EU structural funding to be managed by the Ministry of Finance. The first programme will build a common library database of over 40 libraries, includes the Slovak Library portal to catalogues and collections of Slovak Libraries. It will also set up the Virtual Library of Slovakia - a pilot digital library that will later be developed into the *Memoria slovaca* a portal to cultural heritage with museum and archives collections.

The National Library of Slovakia has a written strategy aiming at digitisation coordinated on national level by the Board of ICT Implementation in Culture at the Ministry of Culture. An industrial approach to mass digitisation of printed bound materials using high capacity digitisation robots will be starting production in 2008. Furthermore the National Library will be connecting the content with users (education, research), will undertake digital preservation and migration. For this digital content management tools for organisation, processing and delivery will be used.

The National Library is part both of the national programmes and thus still in the phase of setting up digitalisation projects.

Depending on the realisation of the national programmes the National Library will digitise the following over the next five years:

350 000 books: estimated 120 million pages
10 000 newspaper titles and other materials: estimated 25 million pages
The NL plans to process gradually to sound recordings and visual materials.

At the National Library of Slovakia it is believed that the EDL should consist of all European digital materials in any form and provided with metadata. The overall aim should be to support knowledge management processes to build a global knowledge society by providing well organised heterogeneous content to the relevant users. It is up to each country which collections and materials from that country will be retrievable from EDL, the important thing is that the content is usable (i.e. has good metadata, the right format etc.) As the main obstacles to the EDL problem with interoperability, IPR and multilingualism are seen.

<sup>&</sup>lt;sup>1</sup> <u>https://www.kis3g.sk/building\_snk/grsnk/projekt%20SDK/SDK\_borz\_eng.pdf</u> (English)

<sup>&</sup>lt;sup>2</sup> <u>http://www.roopis.vlada.gov.sk/data/files/2573.pdf</u> (English)

### 3.33 Slovenia:

A previous strategic document on the information society dating from 2003, the Strategy of the Republic of Slovenia in the Information Society will be replaced in 2007 by the Strategija Republike Slovenije v informacijski družbi - si2010. The aim of the new document is to bring Slovenia in line with the EU information society initiatives arising from the revised Lisbon strategy and the i2010 initiative. The resultant programme will be under the auspices of the Ministry of Culture Ministry of Higher Education, Science and Technology.

The National Library has a Strategic plan for Digital Library of Slovenia<sup>1</sup>. Its aims are to develop a systematic, consistent and comprehensive collection of the nation's published output (called »slovenica«), immediate processing of acquired material with no backlogs to ensure the longevity and security of the written cultural heritage housed in and maintained by NUK. Its goals are the conservation, preservation and storage of digital collections with texts of national importance and the development and maintenance of databases from the library material housed in NUK. Further focus has been put on national and international cooperation and adequate training of staff.

Currently the NL digitises the following material:

- Scientific Articles
- Articles
- o Photographs
- Manuscripts
- o Maps
- o Posters
- Sound recordings
- Slovenian older serials
- Slovenian older literature (19th century books)
- Slovenian old postcards
- Emigrant serials
- Scientific papers
- Scientific articles (before 2000)

Over the next five year the Slovenian National Library expects to digitise:

- Slovenian older serials
- Slovenian older literature
- Slovenian old postcards
- Emigrant serials
- Scientific papers
- Scientific articles (before 2000)

#### 3.34 Spain:

At the moment Spain has no national programme for digitization, but there are a lot of different regional programmes and the Ministry of Culture has recently set up a National

<sup>&</sup>lt;sup>1</sup> <u>http://www.nuk.uni-lj.si/nukeng1.asp?id=150069826</u> (English)

Commission to design a national programme on the digitization and online accessibility of cultural content and digital preservation.

With the Biblioteca Digital Hispánica the Spanish National Library has written strategy for a national digital library. Its aims are replacement of its present system with one that fulfils the OAI protocol and incorporates metadata (at least unclassified in the Dublin Core standard). It plans to give access to the digital versions of the bibliographic treasures of the Library, with a first selection of 30 pieces that permit offering the public a complete presentation. It further plans to include those unique pieces whose present access is now restricted and that can only be consulted in situ. For this a first selection of the 150 most important titles has been made by a committee of experts to ensure their interest and impact for research and for the dissemination of Spanish culture.

- Currently the National Library has the following projects: Obras maestras de la Biblioteca Digital Hispánica (Masterpieces of the Hispanic Digital Library)
- Prensa contemporánea (Spanish contemporary newspapers)
- Hemeroteca digital (Spanish old newspapers and magazines)
- Quijote in the National Library
- Carteles de la República y de la Guerra Civil (Posters of the Spanish Civil War)
- Cartografía española de los siglos XVI al XIX (Spanish maps from XVI to XIX Century)
- o Iconografía hispana (Spanish portraits)
- Goya en la Biblioteca Nacional (Goya in the National Library)

For the next five years the NL plans:

- o systematic digitalization of medieval manuscripts
- systematic digitalization of old maps
- systematic digitalization of old engravings and drawings
- massive digitalization of contemporary Spanish newspapers (more than 400 titles form 1999)
- o massive digitalization of old Spanish newspapers and magazines
- o mass digitisation of books from XVIII century to 1930.

In its digitisation efforts, the National Library of Spain is going to give priority to massive digitisation of old Spanish newspapers and magazines and massive digitisation of books from the 18th century to 1930.

When building the EDL it is believed at the National Library that the priorities in digitisation should be massive digitisation of books in order to be competitive with other providers, subject selection based on works related to Europe and the relationship between European countries, works related to the history and culture of each country and works related to the history of ideas, philosophy, scientific knowledge, etc., and systematic digitization of fragile materials.

A clear organizational model is seen as important to build the European Digital Library, because there are a lot of initiatives in every country and it is very difficult to coordinate them. The lack of funding resources is seen as a main obstacle on the way to the EDL.

### 3.35 Sweden:

The Swedish Arts Council has put in place the Access Programme. The aim of this programme is to make it possible for cultural institutions to improve access to their collections.

In the library field grants have been allocated for digitization of various collections.

For the National Library a digitization strategy is under development.

The National Library currently digitises:

The Open Collections:

- Travelling through the ages
- Swedish classics/Litteraturbanken
- Codex Gigas
- o Suecia antique et hodierna

Over the next five years the NL plans to digitise:

- Newspapers (retrospective digitization of Swedish newspapers)
- Popular journals
- Some manuscripts and other materials

According to the National Library of Sweden, the digitisation of newspapers needs priority on a national as well as on the European level. When building the EDL there is seen to be a need for coordination of digitization of printed materials – to help to avoid duplication. Interoperability issues and the lack of the content are made out by the National Library as the main obstacles on the way to the EDL.

# 3.36 Switzerland:

Switzerland currently has no national digitisation programmes, nor has the National Library a written strategy on digitisation.

The Swiss National Library currently digitises the following collections:

- Dürrenmatt: Inventar der Bildwerke : digitised drawings and paintings
- Posters
- Journal de Genève: digitised newspaper
- Fotodatenbank: digitised photos
- Imvocs: digitised sound recordings

The NL currently has no digitisation plans for the next five years.

In Switzerland as well as for the EDL, in the choice of material for digitisation the National Library believes that newspapers need to be give priority. Rights management, money, infrastructure and workflow are seen as the main obstacles to building the European Digital Library.

### 3.37 United Kingdom:

The UK has no single national programme, but a number of programmes are underway. Among them a number of lottery funded programmes and e.g. the JISC digitisation programme<sup>1</sup> which is investing in digitising content from some of the UK's greatest collections.

The British Library is probably the most important organisation in the field of digitisation in the UK and one of the most active ones in Europe. In its strategy it states that the BL will use digitisation to maximise use of the collections by facilitating a greater volume of networked access, both in the reading room and remote, and by providing the enhanced functionality intrinsic to the digitised items. It wants to reflect national and international priorities for wider access to, and enhanced use of, integrated collections of digitised educational, cultural or scientific materials which have a combined value greater than their component parts. Further aims are to contribute to the conservation of original analogue materials for future users by substituting digital surrogates and to generate income from those products with market appeal that can be exploited commercially by a partner, or the British Library itself, consistent with the aim of maximising accessibility to the collection. The Library's priority is mass digitisation e.g. newspapers, sound.

The British Library Digitisation Programme focuses on:

- 18<sup>th</sup> & 19<sup>th</sup> Century Newspapers
- Sound Recordings
- 19<sup>th</sup> Century Books
- IDP (International Dunhuang Project. The Silk Road Online)

Over the next five years the British Library will digitise:

- 19<sup>th</sup> century books: 25 million pages
- Newspapers: 4 million pages
- Sound: 36,000 items

At the British Library, newspapers and sound files need to be given priority as material for digitisation.

On grounds of access, conservation and pan-European interest pre-1900 newspapers should be the priorities in digitisation to build the European Digital Library, while lack of funding at European and national level for digitisation and post-digitisation activities, and IPR issues (with a special focus on orphan works) are the main obstacles when building the EDL.

<sup>&</sup>lt;sup>1</sup> <u>http://www.jisc.ac.uk/whatwedo/programmes/programme\_digitisation</u> (English)

# 4 Digitisation projects in European National Libraries

#### 4.1 Austria:

The Österreichische Nationalbibliothek has several digitisation projects.

#### Digitisation of Bibles in the Collection of Incunabula of Austrian National Library

The project's aim is the digitisation of 143 Bible copies (about 200 volumes with a total of app. 170.000 pages) of the period from 1460-1500. The project is planned to run from 2007 to 2010.

The collection will be accessible via the Inkunabelzensus Österreich portal.

#### ALEX - Historical legislative and parliamentary Documents online

The project digitises historical law gazettes, jurisdictional and parliamentary material from the Habsburg Monarchy and Austria from 1780-2000. The Project started in 2006 and is currently open ended with 800.000 pages already scanned.

The collection is accessible through The European Library via the OAI-protocol.

#### Digitisation of analogue sound documents

For the project acoustic historical documents, many of them unique objects, get digitized for the reason of long term preserving and saving the contents as well as for the improvement of access to them. The recordings are from between 1898 and 1999. The collection comprises 23 cylinders, 25 records, 160 shellac discs and 300 tapes that will be digitised into 993 digital sound files. The project is expected to produce a total of 4.000 hours recorded time.

The collection will be made accessible through The European Library via the OAI and the Z39.50 protocols.

#### HaLeV (Hebrew Library Vienna)

This project aims at retro-cataloguing of the Hebraica and at the digitisation of its key pages. The Hebraica-collection contains document of the time from 1503-1929. The project started in 2007 is has currently digitised approximately 2.300 books. It ends in 2008.

HaLeV is accessible through The European Library.

#### ANNO – AustriaN Newspapers Online

The project digitises historical newspapers from the Habsburg Monarchy and Austria from 1703-1936. The project started in 2003 has already produced digitisations of 250.000 newspapers and journals with a total of around 3.7 mil pages.

ANNO is accessible through The European Library via the OAI-protocol.

#### Digitisation of the published items of the Papyrus Collection of the Austrian National Library

The project, which is scheduled to run from 2006 to 2011, aims at the digitisation of around 8.000 papyrus documents from 1500 BC - 1600 AD, with some 1.500 already objects scanned.

The digitised papyrus collection will be made accessible through The European Library via the OAI-protocol.

#### Austrian Picture Archive

The Austrian Picture Archive is, with more than 2 million objects, Austria's largest documentation centre of historical photographs and graphics. The core of the archives is made up of over a million photographic documents with topics stretching from portraits to architectural photographs, documents on current events, everyday pictures, and artistic photos.

The digitisation programme of the Austrian Picture Archive has started in 2000 and has been installed as an operational service since 2003. Today the Picture Archive is the centre for reproduction services of the Austrian National Library. The catalogue records for more than 1, 5 Mio pictorial objects can be searched via Internet and ordered as analogue or digital copies. The digital repository of the Austrian Picture Archive holds more than 60.000 photographic objects and more than 130.000 historical portraits at the moment.

The Archive is accessible through The European Library via the OAI-protocol.

### 4.2 Belgium:

The Koninklijke Bibliotheek van België (Royal Library of Belgium) has a list of nine digitisation projects.

#### **Belgian Newspapers**

For this project 30 Belgian newspapers (with 3.000.000 pages) will be digitised completely (scanning, OCR, segmentation, metadata, indexing). The project that works on  $19^{th} - 20^{th}$  Century content will run from October 2007 to December 2008.

The newspapers will be made accessible through The European Library via the OAI-protocol.

#### The first Cadastral survey of Belgium

For the project maps and textbooks of the first Cadastral survey of Belgium (1800) will be digitised between 2006 and 2008. Currently 1.200 maps have been digitised to one image each. By the end of the project 2.300 maps and 2.300 textbooks with a total of 69.000 pages are expected to have been digitised.

Documents from the first Cadastral survey will be made accessible through The European Library via the OAI-protocol by the end of 2007.

#### Manuscripts of the Dukes of Burgundy

Documents for this project come from the  $14^{th} - 15^{th}$  century. Of the 275 manuscripts that are expected to be digitised for this project, 120 have already been finished producing a total of 24.000 images.

#### Maps

For this project 16 maps (miscellanies) have already been scanned producing a total of 734 images. Starting in 2004, this ongoing project digitises maps from the  $16^{th} - 18^{th}$  century.

Maps from this collection will be made accessible through The European Library by 2007 and through DIGIMAP.

#### Engravings

For this project, which stated in 2006 and is ongoing, 3.805 engravings from the  $16^{th} - 20^{th}$  century have already been digitised to one image each.

The engravings will be made accessible through The European Library via the OAI-protocol by the end of 2007.

#### Historical medals

Of the 37.000 medals that are expected to be digitised for this project, 2.503 have already been finished producing one image each. The project has been running since 2006 and produces digital copies of medals from 1790 - 1840.

The medals will be made accessible through The European Library via the OAI-protocol by the end of 2007.

#### Manuscripts

For this project 1.500 pages out of 32 manuscripts have been digitised. The project started in 2004 and is ongoing.

#### Printed 19<sup>th</sup> century brochures

In 1990 19.000 printed 19<sup>th</sup> century brochures were digitised for this project producing 950.000 images.

The brochures will be made accessible through The European Library via the OAI-protocol by the end of 2008.

#### Photographs

From 2003 - 2004 3.010 photographs dating from the  $19^{th} - 20^{th}$  century where digitised. They are accessible through The European Library via the OAI-protocol.
# 4.3 Bulgaria:

Bulgaria's "St. St. Cyril and Methodius" National Library currently has two projects aimed at establishing a basis for digitisation.

## Digital library of the Bulgarian National library

This pilot project provides technical equipment for the digital centre. It includes a sub-project for the digitisation of a certain amount of documents from different unique collections of the library (manuscripts, archives, photographs) and their presentation in Internet. The project is scheduled to run from 2006 to 2008. It has currently digitised 2 books with a total of 463 pages, 8 manuscripts with 1.813 pages, 5 objects of (unspecified) archival with a total of 1.470 pages, and 207 photographs. By 2008 the project will have added to these digital copies of 2 books, 1 manuscript, and 19 objects of archival material. All digitised objects will be from the  $11^{\text{th}} - 19^{\text{th}}$  century.

The National Library plans to make the collection accessible through The European Library.

# Digital, Archives and Documents: Preservation, Information, Scholarship (DAPIS)

Scheduled to start in September 2007, the DAPIS-project is directed at restoration and digital methods of preservation for special collections in the National library. They include Slavonic, Oriental and other foreign language manuscripts and archive documents. Until 2009 the project is expected to produce digital copies of 10 manuscripts and 90 objects of (unspecified) archival material from the  $11^{th} - 20^{th}$  century.

# 4.4 Croatia:

The National and University Library in Zagreb collects all of its digitisation efforts in one project.

# Digitised Heritage

The project's aim is to produce digital reproductions of the valuable and unique items of Croatian cultural, historical and scientific heritage. Items are selected from the special collections: the Collection of Manuscripts and Rare Books, the Print Collection, the Maps and Atlases Collection, the Collection of Music and Audio Material, and the Old Newspapers Collection. Digitised Heritage comprises objects from 1483 to the present day. Thus far the project that was begun in 2004 has produced digital copies of 79 books with a total of 15.720 pages, 3 newspaper titles amounting to 4.727 issues or 37.722 pages, 47 manuscripts with 2.328 pages, 23 music scores and manuscripts with a total of 1.285 pages, 406 maps, 1.170 photographs, and 300 engravings. Until 2013 plans are to digitise around 400.000 objects and a total of 4.000.000 pages of varying content.

The Croatian Digitised Heritage content is or will be made available through The European Library via the Z39.50 protocol.

# 4.5 Czech Republic:

The National Library of the Czech Republic has the following two digitisation projects:

#### Systém Kramerius

Running since 2001 the project digitises Bohemian monographs and periodicals of 19th and 20th centuries. The project has already digitised 9.000 book pages from 25 volumes, and 2.300.000 newspaper pages from 4.600 volumes. The project is open-ended.

Systém Kramerius is available through The European Library (TEL) and the National Uniform Information Gateway via the OAI protocol.

#### Manuscriptorium

The Manuscriptorium is a digital library giving access to content of more than thirty institutions from CZ and abroad via a shared catalogue of historical collections. Beginning in 1992 and running routinely since 1994, this ongoing project's content comes mainly from 8th – 18th century with the exception of some rare recent documents and plans to digitise even older material, e.g. papyri. Currently the project has digitised about 300 old and rare books with 200.000 pages (1.000.000 images as there are 5 images to each page), 1.800 manuscripts with 600.000 pages (or 3.000.000 images), 300 maps, and a few autographs. Plans are to digitise 30.000.000 books, 15.000.000 manuscripts, and several thousand maps during this open-ended project.

# 4.6 Denmark:

The Royal Danish Library has 5 different projects that are grouped by content.

#### Archive for Danish Literature and rare books

This project digitises selected rare books from ca. 1495 to 1900 from the library's collections and books from the Archive for Danish Literature (about 70 Danish classics, collected works + single works by important authors) with books from 1100 to 1930. Running since 1996 has already digitised 2.300 books with 310.000 pages with plans to digitise a further 2.350 books or 320.000 pages within the next 3 years.

#### Manuscripts and rare books

This project is subdivided by period and origin as follows:

*Digitised Western Manuscripts*: Medieval and Renaissance ("Codices Latini Haunienses", "Fragmenta Latina Hauniensia"), 17th and 18th Centuries, 19th and 20th Centuries and Modern Danish Literary Manuscripts, and Rare Books

*From the Oriental and Judaica Collections*: Manuscripts, Rare Books, and Visual. The project that has been running since 1996 has an overall time span from 1000 to 1900. 350 manuscripts and rare books with a total of 70.893 pages have already been added to this collection and another 105.000 pages are expected to be added over the next 3 years.

## Music

This project, with works from 1700 to the present time, is sub-divided into three digital collections: Rischel and Birket Schmidt's digital collection of music for the guitar (complete), Giedde's e-collection of 18th century (mainly instrumental) music (complete), and selected digital works by Danish composers ca. 1750- ca. 1900 (mss. and prints). Running since 1996 the project has already produced 3.599 volumes or 106.247 pages of sheet music and is expected to add another 123.000 pages over the next 3 years.

## Maps and atlases

Containing works from 1500 to the present time, this project is sub-divided into the following collections:

Danish and Foreign digital maps and atlases (containing digitised works from the library's collection of important Danish maps and atlases from the 17th - 19th century), *The Danish National Digital Picture Collection, The Danish Portrait Data Base, The Müller Pinakotek,* (containing the complete digitised collection of 4.433 copper engravings (1500-1795) illustrating the history of Norway and Denmark in original portraits, topographical views, historical pictures, and a few maps. The collection was established by the merchant F. A. Müller. In the collection are several unique prints in a very good quality. This collection was the first to be scanned in the National Picture Database project from 1995), and *The Collection of Topographical Drawings by Rawert* (the complete collection of 1.635 original drawings and watercolour from Denmark and several countries in Europe made by Ole Jørgen Rawert between 1805 and 1849, mainly showing landscapes and buildings). Running

since 1996 this project has already produced digitisations of 148.503 maps and photographs, with another 167.500 maps and photographs to follow over the next 3 years.

## Serials

This project digitises selected serials (1850-1930) from the collections of the Danish National Library. Running since 1996 the project has already produced 315.880 pages and is expected to add a further 342.000 pages over the next 3 years.

All digital collections of the Royal Danish Library are accessible through The European Library, but other services are also free to give access to these collections. Access is possible via the OAI and Z39.50 protocols and via ftp.

# 4.7 Estonia:

The National Library of Estonia currently has two digitisation programmes.

**DEA** (digiteeritud eesti ajalehed or digital Estonian newspapers; to be renamed as Digital Estonian Journalism in 2007)

Started in 2003, DEA was the first large-scale activity of digitising selected material to make it accessible in web. During the pilot phase (2001-2002) the technology and workflow for further digitisation activity were worked out. In order to develop a digital image database, a set of digital images are linked to the database with the required bibliographical, administrative and technical descriptions. The collection includes material digitised in cooperation with the Estonian Literary Museum and covers the most important newspapers published in Estonia from 1821-1944. The project is due to end in 2012 has already produced 82 newspaper titles with a total of 511.880 pages.

The DEA is accessible through The European Library via the Z39.50 protocol with access via other protocols being planned.

# Cultural Heritage of Estonia

This project digitises publications issued in Estonia or concerning Estonia and spans the time since 1535 till today. Started in 2005, the project has already produce digitisations of 549 books with 22.040 pages, 12 newspaper titles with 36.691 pages, 159 serial titles with 56.200 pages, 10 booklets in 21 files, 38 maps in 148 files, and 1.183 book illustrations with one file per illustration. Over the next three years (2007-2010) a further 3.300 books, 100 newspaper titles, 420 serial titles, 50 booklets, 350 maps, and 3.000 book illustrations are expected to be digitised.

The Cultural Heritage of Estonia project's digital material should be accessible through The European Library via the OAI and Z39.50 protocols.

# 4.8 Finland:

The National Library of Finland has already finished 3 digitisation projects (with the last one due to end in 2007).

#### Newspapers

The projects aim is the digitisation of all the earliest newspapers published in Finland 1771-1890. Running since 1998, the project is due to end this in 2007 with 1.000.000 newspaper pages digitised.

## Scholarly Journals

This project's aim was the digitisation of three scholarly journals (1841–1999) in cooperation with the Finnish copyright organisation Kopiosto. Running from 2000-2006, the project has produced 250.000 pages.

## Ephemera

The aim of this project was the digitisation of industrial price lists and catalogues 1809–1944 describing the industrial development in Finland. Running from 2004-2005, the project has produced 130.000 pages.

The Finnish digital collections are already accessible through the national NELLI PORTAL and will by 2008 also be accessible through The European Library, both via the OAI and SRU protocols.

# 4.9 France:

The Bibliothèque nationale de France's flagship in digitisation is of course Gallica, but the BnF has 3 other projects.

# Gallica

Gallica is an encyclopaedic digital library constructed from existing library collections. Those collections consist of printed materials (books, journals, newspapers, printed music, and others), graphic materials (engravings, maps, photographs, etc.), and sound recordings from the 14th century to 1937.

Running since the beginning of the 1990, the Gallica project has produce 117.000 pages of monographs and periodicals digitised in black and white (for the periodicals, this is the number of the issues), of which only 90.000 are online. The other ones are under copyright and only accessible in the reading rooms. Also 250,000 still images have been digitised in colour (photographs collections (Atget), architect drawings (Boulet, Lequeux), maps (d'Anville), and illuminations of the Mandragore database). Of these 80,000 are online.

## Digitisation Programme for the European Digital Library

This project aims to digitise and OCRise first 30,000 then 100,000 books per year during 3 years (2007-2009). 4.000 books (of which 1,750 are currently online) with a total of 1.2 million pages have already be digitised. 330,000 books with a total of 99 million pages are due to be digitised until the end of the project.

## Newspaper Digitisation Programme

Over the period of 5 years, the Newspaper Digitisation Programme aims to digitise, over 5 years, more than 3 million pages of 1 Regional (Ouest éclair) and 26 national newspapers (including Le Figaro, Le Temps, La Croix, and L'Humanité). The project started in 2005 and is due to end in 2008. It has already produced digital copies of 137.000 newspaper issues (with only 134.000 on-line due to copy right reasons) and it is expected that 3,2 million pages will have been produced by the end of the project.

#### Iconographic materials and manuscripts Preservation programmes

Starting in 2005, this combined project aims to replace the traditional preservation programmes (microforms). 35,000 illuminations in 3,000 manuscripts (of the Mandragore project), 20,000 images of specific manuscripts collections (Clémenceau within a national project called IDA), 68,000 images of the Peliot collection (digitised within the Dun Huang digitisation project), 8.000 maps, 3.000 glass plates (plaques de verre; 2.000 maps and 1.000 engravings),: 10,000 images of the Gaignière papiers peints collection, 800 daguerreotypes, and 800 autochromes plates have already be digitised. By the end of the project it is expected that 100.000 manuscript pages (plus 18,000 images of Curie, Flaubert and Valéry from a national project called Optima), 16.000 engravings, and 40.000 drawings will have been digitised.

All Bibliothèque nationale de France material is or will be accessible through The European Library via the OAI protocol. Some material will be accessible through the Mandragore database also via the OAI protocol.

# 4.10 Germany:

The Deutsche Nationalbibliothek has 3 finished projects.

## Exilpresse digital – Deutsche Exilzeitschriften 1933-1945

Running from 1998-2003, this project's aim was to digitise periodicals and newspapers of the German Exile archive 1933-1945. 37 newspaper titles with a total of 86.249 pages have been digitised during the project time.

#### Verteilte digitale Rechtsquellen

Running from 1997-1999, this project's aim was to digitise Legal Sources of the Era of the Allied Control Council for the Territory of the later Federal Republic of Germany 1945-1949. 16 law gazettes have been digitised.

#### Jüdische Zeitschriften

Jewish periodicals during the NS regime in Germany (1933 -1943 (some 1928-1945)) have been digitised from 2004 to 2006 by this project. Digital images of 25 newspapers with a total of 27.649 have been produced.

By 2008 all of the material from the Deutsche Nationalbibliothek is expected to be accessible through The European Library via the OAI protocol and via ftp.

# 4.11 Greece:

The National Library of Greece has 3 finished projects.

#### Newspapers

The digitisation of 4 Greek newspapers (1893–1977) was the aim of this project that ran from 2003-2004. 5 newspaper titles with a total of 220.006 pages have been digitised.

## Year Books of the Philosophical Schools of the Universities of Athens and Thessaloniki

The aim of this project was the digitisation of the Year Books of the Philosophical Schools of the Universities of Athens and Thessaloniki 1925–1977. No images have been produced.

#### Catalogues of Manuscript Codices

This projects aim was the digitisation of 16 printed catalogues of manuscript codices. From 2001-2002 digital copies of 16 printed catalogues of manuscript codices have been produced.

All of the material produced by the National Library of Greece is accessible via the Z39.50 protocol.

# 4.12 Hungary:

The National Széchényi Library has a list of no less than 39 individual digitisation projects which will here be grouped by original material for a better overview.

# Audio

This project group contains the following sub-projects (with, in brackets, their running time) : Collection of Historical Interviews (2007- on-going); Digitised Radio Programmes (2003-2012); HEDERVARY (2006-2007); ANDOR FOLDES piano recordings (2006), Szemethy (2007). This group combines digitised sound recordings from 1930-2006. Currently a total of 17.298 sound files have been produced. By the end of all projects in this group 30.109 sound files are expected to have been produced.

Currently none of this material is accessible for external services.

## Apponyi engravings

The engravings in question belong to the Apponyi Hungarica collection, which comprises of books, pamphlets and engravings printed all over Europe. Approximately half of the engraving collection will be digitised (and described) in this project, the engravings originating from 1501-1700. Running from 2005 to 2007, this project has already produced 519 digitised engravings and is expected to finish with a total of 621 engravings.

Currently the engravings are not accessible for external services.

## Leaflets

A combination of the Collection of 16-17th century leaflets and leaflets of the Hungarian Revolution (1956), this group (running time 2004-2005 and 2006 respectively) has produced a total of 13.500 digitised leaflets.

All leaflets are accessible through The National Digital Archive and The European Library via the OAI and Z39.50 protocols. In addition the 16-17<sup>th</sup> century leaflets are also going to be accessible through <u>http://www.europeana.eu</u>.

#### Manuscripts

In this group manuscript material from about 1500-1884 has been collected. It is comprised of three individual projects. The Collection of album amicorum with 16-18th century peregrination albums, the Collection of Imre Madách's works (a 19th century writer), and Manuscripts from the National Theatre in Budapest (with composer Erkel Ferenc). Running from 2007-2010, 2005-2008, and in 2004 respectively, the projects in the manuscript group have produced a total of 427 digitised manuscripts already and will finish with 570 manuscript overall.

Except for the National Theatre Collection for which no plans seem to exist, the manuscripts are due to be made accessible through The National Digital Archive and The European Library via the OAI and Z39.50 protocols.

#### Maps

In this group there are 2 projects, the Early Maps of Hungary and Transylvania Collection, and the Manuscript Maps and Atlases Collection. Ages given for the maps range from 1474-1900. The projects ran from 2006-2007 and from 2000-2002 respectively and have produced about 3.390 maps and atlas pages on approximately 7.350 images.

All maps are accessible through The European Library via the OAI and Z39.50 protocols.

#### ERKEL

This project which ran from 2002-2004 was aimed at the production of digital copies of musicals manuscripts from between 1840-1880. The projects have produced a total of 57.144 pages.

The ERKEL Collection is currently not accessible for external services.

#### Newspapers

Newspapers in this group are all from between 1900 and 2002. They were digitised in 4 projects, Pénzügyőrségi szaklapok, Tapolcai lapok, Iparjogi Szemle, Newspapers of the Hungarian Revolution (1956). The projects that ran between 2005 and 2007 have produced 115 copies of newspapers. The majority of these are from the Hungarian Revolution period.

The digital newspapers are currently not accessible for external services.

#### Photo and Postcards

The Photos and Postcards group contains 8 projects whose aim it was to digitise photographs and photographic postcards from 1860 to the present day. The group consists of Photos of the famous actors and singers of the 19th century, Postcards collection by counties of the Greater Hungary (before 1920), Postcards and photos from the Theatre History Collection, Picture postcards of Hungarian settlements with theatres, Photo Collection of the 1st World War, Photo Collection of Károly ESCHER, Collection of Contemporary Hungarian Photographers, and the Photo archive of newspaper Magyar Hírlap. The projects have different running times, but are scheduled between 2004 and 2012. They have so far digitised 1.588 photos and 1.328 postcards. When all projects in this group have ended, the collection is expected to consist of digital images of approximately 19.650 photos, ca. 30.000-31.000 negatives, and 1.686 postcards.

Only the World War I collection and the yet to be digitised negatives from the Escher collections are currently scheduled to be made accessible through The European Library.

#### Posters

This group consists of 2 projects whose aim it was to digitise posters from the time between 1837 and 1915. While the first (2003-2006) only digitised Posters of the Hungarian National Theatre in Budapest between 1837 and 1850, the second (2008-2009) of digitising the

Collection of Hungarian posters from before 1914. 6.247 have already been digitised and a further 758 will be added by the second project.

Only the theatre posters are currently scheduled to be accessible through The European Library.

## Theatre Set Designs

The Theatre Set Design group consists of 2 projects that have digitised drawings from between 1920 and 1950. While the first project focused only on the designs of one person (Oláh Gusztáv), the second digitised designs of a number of people. Together the projects have produced digital representations of 2.155 drawings (with about 2 digital images for every drawing of Oláh Gusztáv).

The theatre set designs are currently not accessible to external services.

## Video

This group consists of 2 projects. The first digitises news programs of five national TV channels, while the second focuses on the Hungarian Motion Picture Treasure Collection. The projects have run since 2001 and 2004 respectively and are both ongoing. 11.627 news programmes have already been digitised and a further 1.825 are expected to be digitised every year. 1.476 Hungarian motion picture films have already been digitised and the project is expected to produce a total of 10.290 digitised films.

Only the news programmes are currently accessible via ftp and http.

## SZER

This is an individual project that has digitised 3 microfilm rolls containing news from Radio Liberty from 1956.

This material is currently not accessible to external services.

#### Corvina collection

Bibliotheca Corviniana was one of the most famous European renaissance libraries in the 15th century. In the National Széchényi Library there are 35 pieces codices. Between 2001 and 2002 all of these have been digitised, producing a total of 10.500 digital images.

The codices are due to be made accessible through The National Digital Archive and The European Library via the OAI and Z39.50 protocols by 2008.

## Collection of earliest documents in Hungarian language

This project has been running since 2005 and has already produced digitisations of 26 documents with a total of 6.000. This number will rise to 50 documents by 2010.

The documents are due to be made accessible through The National Digital Archive, The European Library, and through a special portal via the OAI and Z39.50 protocols by 2008/2009.

## Obituary notifications 1711-2003

This project has a running time since 2003 to 2008. Its aim is to produce digital representations of obituary notifications for people who died between 1711 and 2003 in Hungary. 600.000 of these have already been produced, giving a total of 602.500 images.

This material is currently not accessible to external services.

#### Biblioteca italiana-ungherese

This technical co-operation, running since 2007, is a Hungarian-Italian project. It has no digital output.

#### Digital Library of Hungarian Studies

The project aim was the digitisation and online publishing the most important publications (including sound recordings) concerned with the Hungarian culture and social sciences. Running since 2004, it has already produced digitisations of 106 books with 106.000 pages, 3 serial titles with 12.000 pages, and 44 sound recordings. The number of digitised books is expected to rise to 3.800, while the number of digitised serial titles is supposed to rise to 5.000.

This material is going to be accessible through <u>http://www.europeana.eu</u> and through The European Library via the OAI and Z39.50 protocols.

# 4.13 Iceland:

The National and University Library of Iceland has 3 projects listed.

# Forn Íslandskort (Antique maps of Iceland)

All antique maps of Iceland (older than 1900) that are in the collections of the National and University Library of Iceland and the Central Bank of Iceland have been converted to a digital format and are accessible here. A short historical description in Icelandic and English is available for most of the maps. The maps where digitised between 1996 and 1997 with new maps being added where appropriate. Approximately 500 maps, dating from between 1500-1900, have been digitised into one digital image each.

The maps will be made accessible through The European Library via the OAI protocol by 2008.

# Sagnanet

Sagnanet gives access via the Web to digital images of about 240.000 manuscript pages and 153.000 printed pages of books with related material. The material, dating from about 1200-1900, consists of the entire range of Icelandic family sagas. It also includes a very large portion of Germanic/Nordic mythology (the Eddas), the history of Norwegian kings, contemporary sagas and tales from the European age of chivalry. A great number of manuscripts contain Icelandic ballads, poetry or epigrams. These Collections are kept in The National and University Library of Iceland, The Árni Magnússon Institute in Iceland and in the Fiske Icelandic Collection at Cornell University. All manuscripts, on vellum and paper, and printed editions and translations of the Sagas as well as relevant critical studies published before 1900 are included and available through the Internet. Starting in 1997 the project has produced digital representations of 452 books with a total of 151.000 pages, and 910 manuscript volumes with 220.000 pages (containing about 4.400 different works). Though the project has officially ended in 2001, the number of digitised manuscripts is expected to rise to a total of 1.000 volumes with an estimated total of 300.000 pages.

The Sagnanet material is expected to be accessible through The European Library, ENRICH and CERL via the OAI protocol by 2008 or 2009.

## Journals and Newspapers

This project, starting in 1999 and ending in 2010, produces digital copies of newspapers and periodicals from Iceland, the Faroe Islands and Greenland that are available to all through the Web. Users can search the database by titles and countries and eventually a full text search will be provided. They can also browse through the material and print selected pages. The digital library will steadily accumulate more titles from the participating countries, i.e. The Faroe Islands, Greenland and Iceland. The material dates from 1773-2001. About 70.000 issues of newspapers with 935.000 page images and about 40.000 issues of journals with 225.000 page images have already been digitised. These numbers are expected to rise to about 90.000 issues of newspapers and about 20.000 issues of journals by the end of the project.

This material will be made accessible through The European Library via the OAI protocol.

# 4.14 Ireland:

The National Library of Ireland is still in the early stages of its digitisation efforts, but has already had two projects to produce sample images.

## Photographs Collection

A representative selection of photographs from significant collections has been digitised on an ongoing basis to accompany the item records available in the online catalogue. The project is now complete and a strategic review of digitisation activity in the National Library is underway. During this project, which ran from 1998 to 2007, 8.000 photographs have been digitised with three digital images to each original producing a total of 24.000 digital images. The originals date from 1855-2000.

## Prints & Drawings Collection

A representative selection of prints/drawings from significant collections has been digitised on an ongoing basis to accompany the item records available in the online catalogue. The project is now complete and a strategic review of digitisation activity in the National Library is underway. During this project, which ran from 1998 to 2006, 5.000 prints and drawings have been digitised with three digital images to each original producing a total of 15.000 digital images. The originals date from 1700-2000.

Both collections will be made accessible through The European Library via the Z39.50 protocol by 2007.

# 4.15 Italy:

The Biblioteca Nazionale Centrale – Roma lists three digitisation projects.

## Biblioteca Nonantolana Virtuale

The aim of this project, which ran in 2004, was to digitise 45 mediaeval manuscripts coming from the Benedictine abbey of Nonantula (Modena). The 6<sup>th</sup>-13<sup>th</sup> century manuscripts have been digitised into 16.208 images files.

The manuscripts are due to be published on the library web-site (http protocol) by 2007.

# Progetto Siciliani

For this project, running from 2006-2008, the collection of a distinguished 20th century Italian musician and bibliophile, Francesco Siciliani, will be digitised. It includes books, music scores and sound records. 554 books (including music scores) with a total of 120.000 pages have already been digitised. The sound files from 6.000 long playing music records are expected to be added until the end of the project. All material in the collection comes from the 19<sup>th</sup> and 20<sup>th</sup> centuries.

This material is currently not accessible to external services.

## RIMI - digitalizzazione

The project aims at making available the complete collection of some important Italian serials published between 1896 and 1966. Running from 2004-2007, the projects has already produced digital copies of 1 newspaper title with 4.015 issues and 14.000 images of pages, and 4 periodical titles with 4.108 issues and 35.000 images of pages. The number of digitised periodical pages is expected to rise to 65.00 by the end of the project.

This collection will be made accessible through The European Library via the OAI protocol by 2010.

# 4.16 Latvia:

The National Library of Latvia has a list of ten digitisation projects.

# Jazeps Vitols Digital Collection

Running from 2006 to 2009, this project's aim is the digitisation a collection of works, correspondence, photographs, sound recordings, and other material on the founder of Latvian academic school of music, Vitols. All first editions of his musical scores will be digitized. Several biographical essays, bibliographies, and a video film on Vitols have been written for this particular project. The earliest material in this collection dates from 1863. Since the start of the project 1 book with 495 pages, 233 with 1916 pages, 54 autographs with 62 pages, 22 musical manuscripts (scores) with 278 pages, 5 concert programmes with 20 pages, 1 map, 51 photographs, 1 drawing, 3 sound recordings, 2 film or video recordings, and 1 Sibelius score have been digitised. It is expected that the number of printed musical scores will rise to 826, the number of autographs will rise to 100, the number of musical manuscript scores will rise to 35, the number of concert programmes will rise to about 30, the number of photographs will rise to about 90, the number of sound recordings will rise to about 70 by the end of the project.

This Vitols Collection is accessible through The European Library via the SRU protocol, with access via OAI protocol being planned.

# Heritage 1: Latvian Newspapers 1822-1945

The aim of this project, which started in 1999 and is ongoing, is to provide a full digital copy of regional and national newspapers published in Latvia between 1822 and 1945. Later an OCRed and indexed text will be provided. Currently 98 titles have been digitised into 271.000 images with 130.000 to be added by the end of 2007.

The Latvian Newspapers are accessible through The European Library via the OAI and SRU protocols.

## Latvian Posters

The collection of Latvian posters of the National Library of Latvia is among the largest and most comprehensive collections in the world. Amongst others, it consists of approx. 7000 items from the Art Nouveau period as well as several rare loose-leaf materials printed between the beginning of 19th century and the present. All major Latvian artists working in the field of poster are represented. The goal of the project, which started in 2004 and is ongoing, is to digitise the whole collection, which includes items that date from 1823 to the present day. The collection already consists of 1.207 digitised items and is expected to accumulate a total around 7.000 posters.

The collection is expected to be made accessible through The European Library via the OAI and SRU protocols by 2009, with the Z39.50 protocol to be added at a later date.

## Latvian Postcards

The collection of Latvian postcards of the National Library of Latvia, about 80.000 items, includes unique pieces printed between the 1890s and 1914. The range of subjects is very wide, including but not limited to landscapes, views, portraits, and greeting cards. The goal of the project is to digitise all postcards printed between the 1890s and 1945. Running since 2004, this ongoing project has already digitised 880 postcards. The expected final total is approximately 80.000 digitised postcards.

The collection is expected to be made accessible through The European Library via the OAI and SRU protocols by 2010-2011, with the Z39.50 protocol to be added at a later date.

## Latvian Graphics

The collection of Latvian graphics consists of two separate collections: prints (more then 1.100 items) and book plates – (currently about 2.000-2.500 items). The graphics collection represents all classical printing techniques and styles. The range of subjects is quite wide: landscapes, still lives, portraits, illustrations, etc. All major Latvian graphic artists are represented. The goal of the project, which started in 2004 and is ongoing, is to digitize all items. The prints date from the 1920s to the 1990s, the Book plates date from the 1880s to the present day. The collection is expected to have a final total of approximately 4.000 digital images. Currently 217 prints and 1.293 book plates have been digitised, giving a current total of 1.510 files.

The collection is expected to be made accessible through The European Library via the OAI and SRU protocols by 2010, with the Z39.50 protocol to be added at a later date.

## Latvian History of Culture in Images, Part A: Portraits

The database contains portraits from the collection of the Latvian National Library and the State Historical Archives, including drawings, silhouettes, portraits, daguerreotypes, and photographs from the beginning of 19th century – beginning of 20th century. Project part A was started in 2003 and is ongoing. It has already digitised 3.672 photographs. That number is expected to rise to 6.000.

The collection is accessible through The European Library via the OAI protocol, with the SRU protocol to be added at a later date.

## The Picture Collection of the Baltic Central Library

Running from 2006 to 2010, this projects aim is to digitise pictures of Riga from the end of the 19th century until the end of World War II. Of the final 1.300, 50 photographs have already been digitised.

The collection is expected to be made accessible through The European Library via the OAI and SRU protocols by 2011.

#### Latvian Maps

The digitized maps actually represent several collections: maps of Ancient Livonia, Latvian topographical maps, and maps by Latvian cartographer Mantnieks. The priority of the project lies in the digitization of those Latvian maps that where published between 1573 and 1940. Since 2000 this ongoing project has produced digitisations 4.952 (with a total of 5.086 images) of the 6.787 maps that are to be in the final collection.

The collection is expected to be made accessible through The European Library via the OAI and SRU protocols by 2010, with the Z39.50 protocol to be added at a later date.

## Latvian Symphony

This project was part of "Year 2005 of Latvian Symphony". Full scores and orchestra sets that were available only as a single copy or in handwriting, that were only available in damaged copies, but were played in concerts on a regular basis, were digitized. 7 music scores from 1880 to 2005 where digitised into 634 digital images. To round up this collection one more score is expected to be digitised in the future.

The collection is expected to be made accessible via the OAI and SRU protocols by 2008, with the Z39.50 protocol to be added at a later date.

# 4.17 Liechtenstein:

The National Library of Liechtenstein currently has one digitisation project.

#### Scanning of tables of contents

Running from 2005 to 2007, this project's aim is to scan, OCR and index Liechtenstein literature and non-fiction for catalogue enrichment. The original material for the project dates from 1900 to the present day, with new books to be added continually. Of the 40.000 books whose tables of content planned to be scanned for the project 34.000 have already been finished and PDFs have been produced.

The collection is expected to be made accessible through The European Library via the OAI protocol by 2007. The book titles with scanned tables of contents will also be made accessible through <u>http://www.dandelon.com</u>.

# 4.18 Lithuania:

The Martynas Mažvydas National Library of Lithuania has 2 projects listed.

## Creation of Integrated Virtual Library Information System

The implementation of the project began on 1st September, 2005 by its main beneficiary, the National Martynas Mažyydas Library of Lithuania and the partners – the Lithuanian Museum of the Arts and the Archives Department of Lithuania. A databank of digitised cultural heritage objects from libraries, museums and archives is being created. It will contain over 3.605.000 digitised pages of original manuscripts, newspapers, old books, metrics and chronicles as well as other valuable documents of historical importance. The project digitises material from 1413–1940. It has already digitised 663 books with a total of 67.895 pages, 121 newspaper titles with 592.870 pages, 74 serial titles with 312.083 pages, 3.000 manuscripts with 20.00 pages, 43 items of archival material (births, marriage and death records, books of the Roman Catholic churches in Lithuania (1599-1907), and other archival documents) with 116.245 pages, 1.221 photographs (producing 1.845 digital images), 7.024 (producing 10.325 digital images), 8.467 digital images of 4.976 drawings, 624 digital images of 309 postcards, 3.016 digital images of fabrics and textiles, and 8.015 digital images of paintings. When the project ends in 2008 at total of 5.000 books, 200 serial titles, 3.000 manuscripts, 70 items of archival material, 5.000 photographs, 13.024 engravings, 5.976 drawings, 309 postcards, 6.016 items of fabrics and materials, and 10.382 paintings will have been digitised.

The collection is expected to be made accessible through The European Library via the SRU protocol by 2008.

## Collection of Lithuanian Shellac records

Since 1961 the National Library has been collecting and preserving the archive of published Lithuanian sonic documents as a part of the national archive of published documents. The main criteria for its formation are comprehensiveness, a guarantee for preservation, accessibility and a foundation for discography, discology, musicology and scientific research of the sonic heritage. The national archive of sound documents contains more than 7.000 titles, 1.560 of these represent old Lithuanian shellac records. After winning the grants from the Fund for the Support of Culture and Sports of the Republic of Lithuania in 2003, 2004 and 2007 the Department of Music of the National Library has been conducting the continuous project of Lithuanian shellac records digitisation. The Department of Music is planning to convert all remaining old shellac records into digital format and to form a cumulative stock of digital shellac records on CDs. The recordings are from between 1907 and 1950. 1.286 shellac records have already been digitised, representing 2.572 musical works or around 9.231 minutes of audio data. This digitisation of a total of 1.800 shellac records is planned for this project.

Currently none of this material is accessible for external services.

# 4.19 Luxembourg:

The Bibliothèque nationale Luxembourg lists 2 projects.

#### Historical Postcards

This 2006 projects aim was the digitisation of 18.000 historical postcards of Luxembourg from 19th century until the 1950s. Scans are full-colour and include front and back side. Full-resolution images are only available in-house, as an estimated 10.000 postcards are orphan works, for which rights could not be secured.

#### Periodicals and Monographs

Starting in 2003, this project's aim is the digitisation of 12 Luxembourg periodicals 5 Luxembourgish reference works. Included are already digitised but not online yet titles. The original material is from between 1704 and 1952. For the project 48 books with 10.500 pages, 31.000 newspapers with 157.000 pages, and 6.500 serial issues with 56.300 pages have already been digitised. The projects final goal is the digitisation of 43.000 newspapers with 229.500 pages, 7.800 serial issues with 161.300 pages, and an unnamed number of books by the end of 2007.

The collection is expected to be made accessible through The European Library via the OAI protocol by 2007.

# 4.20 Macedonia:

The National and University Library "St Kliment Ohridski" in Skopje currently has 2 projects underway.

#### Strategy development for the digitization of special collections

The aim of the project is to develop a digitization strategy and to provide a methodological framework based on international experience in order to build and preserve a digital collection at the NUL "St Kliment Ohridski".

The project will consist of three phases: 1. The exchange of know-how between the Slovenian, Austrian and Macedonian partners. 2. The development of a complete strategy and methodology for the digitization of Macedonian special collections; 3. Dissemination of the project results on national and international level.

The project is scheduled to run between June 2007 and May 2008. It will not produce digitised data.

#### Digitisation of the Archeographic collection

The digitisation project began in 1999. The Archeographic collection contains about 450 manuscripts from the 13th to 18th centuries. 58 of these have already been digitised producing 3.000 digital images. The project is expected to produce a total of 20.000 digital images.

The collection is expected to be made accessible through The European Library via the ftp protocol by 2008/2009.

# 4.21 Malta:

The National Library of Malta currently has no digitisation projects, but has produced digital copies of 3 maps and 10 postcards. These have been made accessible through The European Library via the Z39.50 protocol.

# 4.22 Moldova:

The National Library of the Republic of Moldova currently has a single project that combines all of its digitisation efforts.

## National electronic library the of Republic of Moldova

Starting in 2007, the project is expected to produce digital copies of 30.000 original objects. Of these 1 book with 800 pages, 1 serial title with 3.000 pages have already been digitised. 3.000 books, 400 serial issues, 48 manuscripts, 40 maps, 300 photographs, 200 postcards, and 26.000 sound recordings are to follow. The material is from the time between 1643 and 1945.

The collection is expected to be made accessible through The European Library via the Z39.50 and the SRU protocols.

# 4.23 The Netherlands:

The Koninklijke Bibliotheek has a list of 5 individual digitisation projects.

## Geheugen van Nederland / Memory of the Netherlands

This is actually a national programme for the digitisation of Dutch cultural heritage, providing funding for digitisation projects in ALM-sector, expertise in the form of guidelines, project coordination, and quality control. The programme is also providing tiff-archive for long term preservation and access of master formats and services to encourage re-use of digital material. The programme is divided into a first phase from 2000-2004 and a second phase starting 2005 and ending in 2008, but negotiations are on the way to prolong this programme after that. Under the programme some 350.000 items have already been digitised. These are 600 books with 80.000 pages, 1.775 serial issues with 30.000 pages, 5.300 manuscript pages, 5.150 maps, 48.795 drawings, 38.205 posters, 2.500 postcards, 19.610 three-dimensional objects, 100 hours of sound recordings, 250 hours of film and video recordings, and 33.700 paintings. Future plans depend on funding. The oldest objects of this collection are pre-historic and the newest from present times.

The collection is expected to be made accessible via the SRU protocol, with the OAI protocol to be added at a later date.

## ANP Nieuwsberichten Digitaal

This projects aim is the digitisation of the news bulletins of the Dutch News Agency ANP that were broadcast on radio in the period 1937-1989. This project will make app. 1,5 million pages (each containing several news items) of typewritten material electronically accessible through scan and OCR. The projects runtime is scheduled to start in 2007 and end in 2007.

The collection is expected to be made accessible through The European Library via the SRU protocol by 2009. Negotiations are also on the way to make this collection available to the ANP Photo-archive website where the bulletins will be linked to the historical photographs.

## Digitalisering Bijzondere Collecties / Digitization Special Collections

The aim of the project is providing free access to 5.200 digitized books from 1781-1800 that are preserved in Dutch scientific collections. The project is due to start in 2007 and end in 2010.

The collection is expected to be made accessible through The European Library via the SRU protocol by 2009, possibly with other protocols to be added at a later date.

## Databank of Digital Daily newspapers

Running from 2006-2011, this project will cover a selection of Dutch newspapers dating from 1618 to 1995. An estimated 8 million pages of text material will be digitised and OCRed.

The collection is expected to be made accessible through The European Library via the SRU protocol by 2009.

#### Staten-Generaal Digitaal

Running from 2004 to 2010, this project's aim is the digitisation of the documents of the Dutch national parliament from 1814-1995. A total of 2.3 million pages of text, tables and images will be digitised during project time, with 412.884 pages already scanned providing 2.010.965 digital files (412.884 tiff, 367.852 jpg, 777.280 xml, 40.065 PDF).

The collection is expected to be made accessible via the SRU protocol.

# 4.24 Norway:

The National Library of Norway combines its digitisation efforts into one programme.

## The National Library Digitisation Programme

The objective of the programme is to digitise the complete collection at the National library of Norway. This contains a large number of objects from a variety of media types. Currently digital representations of approximately 1.000.000 objects are available from the digital library. Some of the objects are, however, only available when accessed inside the Library's buildings, or when access has been granted for given persons upon specific request. The material dates from 1400 to the present day. 20.000 books with 4.500.000 pages, 200.000 newspaper issues with 1.000.000 pages, 200 serial issues with 10.000 pages, 1.000 pamphlets, 1.000 manuscript pages, 45 maps, 320.000 postcards, 5.000 posters, 150.000 radio sound recordings (with digital representations in 3 file formats each), 4.000 music or other sound recordings, and 570 film and video recordings have been digitised since 2006. Until 2026 digital copies of a total of 450.000 books, 4.700.000 newspaper issues, 2.000.000 serial issues, 1.800.000 pamphlets, 200 000 sheets of music, 4.000.000 manuscript pages, 180.000 maps, 1.050.000 postcards, 60.000 posters, 250.000 postcards, 1.000.000 radio sound recordings, 75.000 music or other sound recordings, and 245.000 film and video recordings are expected to have been produced.

By October 2007 the free part of the collection is expected to be made accessible through The European Library via the OAI and Z39.50 protocols. By 2008 it is supposed to be searchable by all search engines.

# 4.25 Poland:

The Polish National Library has a list of 10 digitisation projects.

## Polona

The aim of this project is the digitisation of a collection of Polish literature and scientific materials, historical documents, journals, graphics, photography, musical scores, and maps from the time between 1100 and 1945. Running since 2006, the project has up to now digitised 2.900 books, 18 manuscripts, and 40 engravings. Until 2010 digital copies of a total of 40.000 books, 60.000 serial issues, 1.000 manuscripts, 5.000 maps, 50.000 photographs, 50.000 engravings, 10.000 drawings, and 20.000 postcards are expected to have been digitised.

# Adam Mickiewicz

This projects aim is the digitisation of a collection of the most interesting editions of the works by Adam Mickiewicz (1798-1855), considered the greatest Polish Romantic poet of the 19th century. Running since 2006, the project has up to now digitised 90 books with a total of 12.000 pages. By the end of the project in 2007 a total of 400 books, 5 manuscripts and 5 autographs, 60 engravings and 10 drawings, as well as 20 postcards are expected to have been digitised. The original material was published between 1820 and 1960.

## Children's and youth's literature

This project's aim is the digitisation of a collection of literary works and textbooks for children and youths. It includes among others works by Hans Christian Andersen, Władysław Bełza, Zofia Bukowiecka, Stanisław Jachowicz, Maria Konopnicka, Artur Oppman, Jadwiga Teresa Papi and first Polish encyclopaedias for children and abc-books. The collection is made up material dating from 1790-1937. Starting in 2007, the project has already produced 150 books with a total of 12.000 pages. By the end of the project in 2008 a total of 1.000 books and 400 serial issues are expected to have been digitised.

# Cyprian Kamil Norwid

This project's aim is the digitisation of a collection of manuscripts, published works and drawings by Cyprian Kamil Norwid (1821-1883), a nationally esteemed Polish poet, dramatist and painter. The material in the collection was published between 1850 and 1960. Since starting in 2007 a total of 20 books with 8.000 pages and 3 autographs have been digitised. By the end of the project in 2008 digital copies of 250 books, 15 manuscripts, 10 autographs, 25 engravings, and 300 drawings are expected to have been produced.

## Jan Kochanowski

This project's aim is the digitisation of a collection of the most important editions of the works by Polish Renaissance poet Jan Kochanowski (1530-1584). He is regarded as "the father of Polish literary language" and the greatest Polish poet prior to the 19th century. The collection contains material published between 1560 and 1960. The project is scheduled to

start and finish in 2007. A total of 190 books, 50 early printed books, 1 autograph, and 40 engravings will have been digitised.

## January Uprising

This project's aim was the digitisation of a collection of classified periodicals, appeals, declarations, photographs, postcards, and leaflets as well as other material from 1861-1864 concerned with the January Uprising. Its goal is to provide access to material which is not easily obtainable in its traditional format because of its substantial value, uniqueness and preservation conditions. Running from 2004 to 2007, the project has produced digital copies of 150 with a total of 1.200 pages, 50 photographs, and 50 leaflets (80 digital images).

## Juliusz Słowacki

This project's aim is the digitisation of a collection of the most important editions of works by Polish Romantic poet and dramatist Juliusz Słowacki (1809-1849). The material in this collection was published between 1820 and 1960. The project is scheduled to start and finish in 2007. A total of 400 books, 5 manuscripts, 5 autographs, 60 engravings, 10 drawings, and 20 postcards will have been digitised.

## Mikołaj Rej

This project's aim is the digitisation of a collection of the most important editions of the works by Polish Renaissance poet and writer Mikołaj Rej (1505-1569). The material in this collection was published between 1550 and 1960. The project is scheduled to start and finish in 2007. A total of 150 books and 30 engravings will have been digitised.

## Treasures of the National Library

This project's aim is the digitisation of a collection of the most precious and valuable documents from the collection of the National Library. The material in this collection was published between 1100 and 1939. Running from 2007 to 2010, the project has already produced digital copies of 10 with a total of 1.000 pages, 12 manuscripts with a total of 1.200 pages, and 1 autograph with 2 pages. The final production is expected to be digital copies of 100 books, 190 manuscripts, 10 autographs, 50 maps, 100 engravings, and 50 drawings.

## Underground publications from World War II

This project's aim is the digitisation of a collection of antifascist underground literate activity from years 1939–1945. These publications gathered in the collection are part of the most precious 19<sup>th</sup> and 20<sup>th</sup> century recourses of the National Library, so called cimelia, rare and valuable documents, under special protection, that are not available to readers. Of the 500 books that are to be digitised for this project between 2007 and 2008, digital copies of 135 books with 5.000 pages are already finished.

All collections of the National Library of Poland are accessible through The European Library via the OAI protocol.

# 4.26 Portugal:

The National Library of Portugal combines all its digitisation efforts in 1 project.

## The National Digital Library

Starting in 2002, the project is a continuing initiative with no definite target figures. To date some 490.000 digital images have been produced. These come from around 2.000 books, 34 newspaper and serial titles, ca 500 manuscripts and autographs, around 600 items of archival material, ca 1.500 maps, and from 4.974 engravings, drawings, posters, and postcards. The originals digitised in this project date from since 1500.

The collection of the National Library of Portugal is accessible through The European Library, PORBASE, CERL and Google Scholar via the OAI, Z39.50 and SRU protocols.

# 4.27 Russia:

Russia has two separate National Libraries that are treated here individually.

## Russian State Library, Moscow:

#### Library Science Virtual Hall

The Library Science Virtual Hall is a collection of full-text electronic documents relating library science. The Virtual Hall was begun in 2001 and is ongoing. It contains documents from the 19<sup>th</sup> and 20<sup>th</sup> centuries as well as current ones. The collection contains articles from around 370 books and about 30 serials, with 3.000 new articles added every year. It also contains 18 posters.

The Virtual Hall is accessible via the Z39.50 protocol.

#### Libmuz (all-Russian Musical Electronic Library)

The purpose of Libmuz is the digitization of musical sound recordings for preservation. The project has been running since 2006 and is ongoing. It contains about 300 recordings are from between 1950-1990 with a further 300 to be added every year.

The collection is currently only accessible on-site, but may be accessible via the Z39.50 protocol in the future.

#### Meetings at Frontiers: Russia-USA, Russia-Kazakhstan

These collections deal with emigration, transmigration, development of new areas as well as the socialist economic and cultural development in Kazakhstan during first five-year-plan periods (1917-1944). There are also rare and unique books on history of Russia and Kazakhstan, on ethnography and archaeology of Kazakhstan during the period of 1829-1916. "Meetings at Frontiers" contains documents from the 14<sup>th</sup> century to the present time. They ran from 1997 to 2001 (USA) and from 2003 to 2004 (Kazakhstan), comprising 335 books, 114 maps, 15 manuscripts, 8 albums and 11 atlases.

"Meetings at Frontiers" Hall is accessible via the Z39.50 protocol and may be accessible through The European Library in future.

#### **Electronic Library of Dissertations**

Starting in 2003 with dissertations on more often requested topics as economic sciences, jurisprudence, pedagogy, psychology and philosophy, the RSU has been digitising all dissertations except for those in the fields of medicine and pharmacy since 2004. Dating from 1998 to today, 174.026 documents, including 33.594 abstracts of dissertations and 140.432 dissertations have already been digitised.

The Electronic Library of Dissertations is accessible via the Z39.50 protocol.

## Russian State Symbols – History and Reality

The Russian State Symbols digital collection gives open access to text materials, images and sound records devoted to state symbols of Russia beginning with Kiev Russia (9<sup>th</sup> century) up to nowadays. Running since 2006 as ongoing work, the collection currently contains 107 books, 184 drawings, 2 posters, and 10 sound recordings.

The Russian State Symbols are accessible via the Z39.50 protocol and may be accessible through TEL in future.

# Slavic-Russian hand-written books from the 14<sup>th</sup> to 19<sup>th</sup> century from the Trinity-Sergiev monastery

This collection contains handwritten books, archive from the Trinity-Sergiev Monastery and the Moscow theological academy. Running from 2002 to 2015, this project has already digitised 1.200 Autographs and 517 units of unspecified archival material, dating the 14<sup>th</sup> to 19<sup>th</sup> century.

This collection is accessible via the Z39.50 protocol and may be accessible through TEL in future.

#### National Library of Russia, St. Petersburg:

#### Celebrating the 300th anniversary of Saint-Petersburg

This collection has the following sub-collections: Saint-Petersburg: Postcards with city views; Saint-Petersburg suburbs: Postcards collection; Maps of the City of Saint Petersburg; History and Culture of Saint Petersburg in the Graphic Arts. Running from 2002 to 2003, the collection now contains 600 maps, 500 engravings, and 900 postcards dating from the 18<sup>th</sup> to the 20<sup>th</sup> century.

The collection is accessible through The European Library via the OAI protocol.

#### The First Russian maps

The First Russian Maps is a collection of 18th century maps and atlases printed in Russia. Running from 2002 to 2009, the collection contains 900 pieces of unspecified printed material as well as 300 maps from between 1700 and 1800.

The collection is accessible through The European Library via the OAI protocol.

#### Newspapers in the Bukhar-Jewish language, in the Khanty and Mansi languages

For this project, newspapers in languages of the Finno-Ugric group (1930-1950) were digitised in 2006. The collection contains around 6.000 items.

The collection is accessible through The European Library via the OAI protocol.

#### Books in languages of the Finno-Ugric group

Running from 2003 to 2006, this projects aim was the digitisation of books in Khanty, Mansi, Mari, Udmurt, Karelian, and other languages of the Finno-Ugric group. The collection contains about 2.500 books dating from 1800 to 1917.

The collection is accessible through The European Library via the OAI protocol.

#### The art of 5<sup>th</sup> to 16<sup>th</sup> century European manuscripts

For this project approximately 1.500 manuscripts from the 5<sup>th</sup> to 16<sup>th</sup> century where digitised in 2003.

The collection is accessible through The European Library via the OAI protocol.

#### The 100th Anniversary of the State Duma

This collection consists of 250 photos (500 files) of members of the State Duma (Russian Parliament). All images were digitized in 2006. The originals date from 1906 – 1917.

The collection is accessible through The European Library via the OAI protocol.

## Works of Karl Bulla - "godfather of photo journalists"

Running in 2005, this project has digitised approximately 250 photos (500 files) dating from the beginning of the 20<sup>th</sup> century.

The collection is accessible through The European Library via the OAI protocol.

#### Russian/Soviet posters

Dating from the beginning of the 20<sup>th</sup> century, approximately 250 posters (500 files) where digitised in2003.

The collection is accessible through The European Library via the OAI protocol.

#### Collection of Russian dissertation theses

Running since 2000, this project aims at producing a complete collection of all new Russian language dissertation theses, printed in the territory of the Russian Federation. Currently the collection consists of 180.000 theses or 4.680.000 images of pages.

The collection is accessible via the Z93.50 and OAI protocols.

# Complete statute book of the Russian Empire

Running between 2006 and 2007, this projects aim is a complete digitisation of Russian legislation from 1649 till the end of 1913. The collection consists of 235 books or over 235.000 pages.

The collection is accessible through The European Library via the OAI protocol.
# 4.28 Serbia:

The National Library of Serbia has a list of 10 digitisation projects.

# Serbian children's digital library

The Serbian Children's Digital Library contains about 130 most beautiful books of Serbian children's literature in digital format. It is a part of a big non-profit research project, the International Children's Digital Library. A national editorial board consisting of children's literature professors, publishers, librarians, made a choice of books taking care of contents, artistic, historical and literature values. All digital books are colour digital copies of analogue ones. This project ran between 2003 and 2004 and has produced 14.184 images of books published between 1954 and 2003.

This collection is accessible through The European Library and ICDL (<u>http://www.icdlbooks.org/</u>) via the OAI protocol.

# Cyrillic manuscripts

This projects aim is the digitisation of a collection of selected folios of 290 Serbian Cyrillic manuscripts from the treasure of the National Library. The manuscripts date from 13th to 18th century and present first monuments of Serbian written culture. Running from 2005 to 2007, this project has produced 116.850 images.

This collection is accessible through The European Library by 2008 and the ENRICH project by 2009 via the OAI protocol.

# Old and rare books

This projects aim is the digitisation of a collection of books in the Serbian language printed in Venice, Vienna, Pest, Buda, Rimnik, and Leipzig. Examples like the Stematography of Žefarović, the work of Orfelin or Rajić, books of Dositej, Sterija or Vuk, illustrate perfectly the development of Serbian social culture. Almanacs and yearbooks from the 19th century are also included. All the material dates from between 1500 and 1867. Between 2007 and 2009 at total of 3.647 books are to be digitalised. &3 of these have already been digitalised, producing 7.480 digital images.

This collection is accessible through The European Library by 2009 via the OAI protocol.

# Politika

Politika is Serbia's oldest daily newspaper representing Serbian history and culture from 1904 up to the present day. The project to digitise the greater part of this collection started in 2003 and has already produced digital copies of 28.000 issues with 550.700 images. Ending in 2007 the project is expected to produce a total of 35.000 digitised newspaper issues.

This collection is accessible through The European Library and <u>http://www.politika.co.yu/</u>. (Politika is available online only for the period: 1904-1941)

# Cartographic materials

This project's aim is the digitisation of a collection of cartographic material consisting of several sub-collections: Old atlases, old maps of Belgrade, and first military maps of Serbia printed in the country. The material dates from the 16<sup>th</sup> to the 19<sup>th</sup> Centuries. Running from 2006 to 2007, the project has produced digitisations of 900 maps or 1.000 digital images.

This collection is accessible via the OAI protocol.

#### Printed music and sound records

This project's aim is the digitisation of a collection of gramophone records on 78 rpm and includes mainly about 900 audio recordings of Serbian and Yugoslav music from the first half of 20th century (1900-1960), but also recorded world classics performed by famous musicians. Running since 2004, the project has already produced digital copies of 450 gramophone records with 1.200 individual sound recordings. Ending in 2007, the project is expected to produce a total of 600 digital copies of gramophone records.

This collection will be accessible through The European Library by 2008 via the OAI protocol.

#### Photograph collection

This project's aim is the digitisation of a collection of selected materials from the photograph collection. These are photographs, albums and postcards dating from 19th century and the first half of 20th century. In its runtime, 2006-2007, the project has produced digitisations of 550 photographs and 150 postcards.

This collection will be accessible through The European Library by 2010 via the OAI protocol.

#### Engravings and art material

The engravings are collected from several sub-collections: Graphics from Mount Athos, graphics by Anastas Jovanovic, the first Serbian photographer, and graphics of Belgrade from between the 16th and 19th centuries. The art material contains an album of seven linocuts (1921-1922) by Mihailo S. Petrov, a pioneer of Serbian contemporary graphics. Running from 2006 to 2007, the project has already produced digital copies of 1 of the 10 designated linocuts and of all 230 engravings.

Currently only the sub-collection containing the graphics from Mount Athos is accessible through The European Library via the OAI protocol.

#### Posters and documentary material

This project's aim was the digitisation of a collection of playbills covering Serbia and the Kingdom of Yugoslavia during the second half of 19th and the first half of 20th century, old

stock shares from the 19th and 20th century, and proclamations and declarations of the Serbian Royal Dynasties. Running from 2006 to 2007, the project has digitised 670 posters and 380 stock shares.

This collection is accessible through The European Library via the OAI protocol.

# DOI Serbia

This project's aim is the digitisation of a collection of 27 locally published scientific journals. The full-text archive includes journals from since 2002 on a "cover to cover" basis. All indexed journals are published by professional societies, and some of them are indexed in well-known indexes such as World of Science or CA. Metadata, as well as full-text, are formed according OAI-PMH, and may be downloaded free of charge. This collection will be expanded with new titles during next years. The project started in 2004 and has up to now digitised 4.200 individual journal articles.

This collection is accessible through The European Library via the OAI protocol.

# 4.29 Slovakia:

The Slovak National Library combines all its digitisation efforts in 1 project.

# The Slovak Digital Library Collection

Started in 2003, the collection currently contains digitised cultural and scientific heritage held by the Slovak National Library and its organisational units: photographs, postcards, musical manuscripts, paintings, drawings, book illustrations, regalia, literary museum objects, text with restricted access, and materials by and about the 1st Czech-and Slovak Republic's Prime Minister. In the future (starting in 2008) this collection will include output from mass digitisation of books and newspapers and will be integrated into a higher level national resource. The material's age varies, the oldest dates back to the Middle Ages, the newest is from the present day. Currently about 100 books, 1.500 serial issues, ca. 500 fragments of manuscripts (producing the same amount of images files), about 1.000 Paintings (1 digital image per painting), around 50 boxes of archival material, 2.000 photographs, 2.000 drawings, 500 postcards, ca. 100 bookplates, 3 three-dimensional objects, close to 20 pieces of fabrics and textiles, and 45 files of sound recording (< 60 hours) have been digitised. The number of digitised books is expected rise to around 350.000 titles (with ca. 120 million pages) and around 10.000 newspaper titles (with ca. 20 million pages) are expected to be added to the collection.

The collection is accessible through The European Library via the OAI and Z39.50 and ftp protocols.

# 4.30 Slovenia:

The National and University Library of Slovenia has a list of 13 individual projects. For a better overview the projects dealing with scientific papers or articles and the projects dealing with serials have been grouped together.

# Scientific Articles and Papers

The group consists of three individual projects, the first of which started to digitise current scientific articles in 2006. This project is ongoing. The other two will be from 2007 - 2010 and be digitising older scientific articles (1945 - 2000) and scientific papers (1918 onwards) respectively. For the first project 710 articles with 2.840 pages have already been scanned. This collection is expected to grow by approximately 3.000 pages a year. The other two projects are expected to produce digital copies of 75.000 scientific articles and 15.000 scientific papers respectively.

These collections will be made accessible through The European Library via the OAI protocol by 2008 and 2010.

# Journal and Newspaper Articles

For this project a special older articles database providing full-text search and search of bibliographic data has been created. The project has been running since 2002 and is ongoing. 29.813 have already been digitised, producing 120.000. A total of 500.000 articles are planned to be digitised by the end of the project. Articles in this collection date from 1797 - 1945.

The collection will be made accessible through The European Library via the OAI protocol by 2009.

# Photographs

Through the portal <u>http://www.dLib.si</u> one can access more than 10.000 photographs and drawings, which have been made by distinguished Slovene photographers and artists. The collection may be searched by bibliographic information. For this project 10.800 photographs have been digitised between 1998 and 2007. They date from 1890 onwards.

The collection is accessible through The European Library via the OAI protocol.

# Manuscripts

The collection consists of manuscripts from the editorial archive of the musical serial Novi akordi. From 2006 to 2007 542 manuscripts (with one image each), dating from 1900 to 1918, have been digitised.

The collection will be made accessible through The European Library via the OAI protocol by 2009.

# Maps of the Slovenian Territory from 1548 to 1924

The project entitled Digitization of maps of the Slovenian territory from 1548 to 1924 encompasses the most representative historical maps of the Slovenian territory and city plans of Ljubljana in a digital form. The majority of the maps are essential for national identity and cultural heritage since they represent historical development of the Slovenian territory. Scheduled to run from 2000 to 2010 the project has already digitised all of the 83 maps it was planned for at one image each.

The maps of the Slovenian territory are accessible through The European Library via the OAI protocol.

# Sound Recordings

The collection consists of 69 digitized songs. The project has made possible the revitalization and access to forgotten and technically unreachable Slovenian art heritage. This acoustic documentation is invaluable for knowledge and studies of music history. The songs, dating from 1918 – 1925, were digitised between 2001 and 2007.

The sound recordings are accessible via the OAI protocol.

#### Posters

Posters are a mass media and communication tool and a mirror of the society, space and time in which they were created. They help us to follow the development of Slovene design. They have captured the political history, social life and economic development of a certain period. The portal <u>http://www.dLib.si</u> provides access from one point via the Posters database. You can search by theme, author or year of creation. They cover the period from the 1850 to 1992 and their authors are famous Slovene designers and artists. This collection will contain 10.000 posters by 2010, of which 2.030 (at one image each) have already been digitised since 2001.

The posters will be made accessible through The European Library via the OAI protocol by 2008.

# Serials

This group consists of one project each for the digitisation of emigrant serials and Slovenian older serials. Both will run from 2007 to 2010 and will digitise 75.000 and 500.000 articles respectively. The older articles date from 1800 to 1918 while the emigrant articles date from 1945 onwards.

Both collections will be made accessible through The European Library via the OAI protocol by 2010.

# Slovenian older literature

Running from 2007 to 2010, this project is expected to digitise 6.600 books dating from 1700 – 1918. 400 of these have already been digitised, producing at total of 100.000 images.

This collection will be made accessible through The European Library via the OAI protocol by 2008.

#### Slovenian old postcards

Running from 2007 - 2010, this project will produce digital copies of 1.000 postcards dating from 1880 - 1945.

The postcards will be made accessible through The European Library via the OAI protocol by 2010.

# 4.31 Spain:

The Biblioteca Nacional de España has a list of 8 digitisation projects.

# Obras maestras de la Biblioteca Digital Hispánica (Masterpieces of the Hispanic Digital Library)

Running from 2006 to 2007, the projects aim was the digitisation of unique pieces whose present access is restricted and that can only be consulted in situ. To start this task, a first selection of the 150 most important titles has been made by a committee of experts to ensure their interest and impact for research and for the dissemination of Spanish culture. For the project 36 manuscripts, 4 incunabula, 65 books (18.726 digitalized pages), 7 drawings, 10 engravings, 7 photographs, 7 maps, 7 musical scores, and 7 booklets have been digitised.

This collection will be made accessible through The European Library via the OAI protocol by 2008.

# Carteles de la República y de la Guerra Civil (Posters of the Spanish Civil War)

Running from 2002 to 2008, this project has digitised 456 posters (on images each) dating from 1931 - 1939.

This collection will be made accessible through The European Library via the OAI protocol by 2007.

# Goya en la Biblioteca Nacional (Goya in the National Library)

This 1997 project digitised 411 engravings by Goya (one image each). The engravings date from 1771-1828.

The engravings will be made accessible through The European Library via the OAI protocol by 2007.

# Hemeroteca digital (Spanish old newspapers and magazines)

This project's aim is to offer access to the digital collection of historic Spanish press contained at the Library, initially composed of 143 titles, published between 1772 and 1933. Started in 2007, the project is ongoing and will produce digital copies of around 2.000 titles. The current 143 titles amount to 39.740 articles or 448.383 images of pages.

The Hemeroteca digital will be made accessible through The European Library via the OAI protocol by 2009.

The project's aim is the digitisation of portraits (engravings, drawings and photographs) of Spanish writers, politicians, kings and queens, etc. from the 1st to the 20th century. The project's running time is from 2002 - 2008. 4.890 engravings have already been digitised and 3.000 further portraits are expected to be added to this sum by the end of the project.

The Iconografía hispana will be made accessible through The European Library via the OAI protocol by 2007.

#### Prensa contemporánea (Spanish contemporary newspapers)

For this project 400 titles (amounting to 50.000.000 pages) of Spanish contemporary newspapers have been digitised since 1999. The project is ongoing and 4.500.000 digitised pages are expected to be produced per year.

# El Quijote en la Biblioteca Nacional: ediciones digitalizadas (Quijote in the National Library: digitised editions)

Running from 2005 to 2006, this project digitised 26 editions selected from the main editions of Quijote, including the first one. These 54 books (14.963 images of 29.926 digitalized pages) date from 1605-1853.

This collection will be made accessible through The European Library via the OAI protocol by 2007.

# 4.32 Sweden:

The National Library of Sweden has a list of 4 digitisation projects.

#### Suecia antique et hodierna

The Suecia antiqua et hodierna (1660-1715) is an extensively illustrated work from Sweden's period as a great power. Its digitisation is an attempt to preserve the original, yet making it even more accessible than previously. Committed to high standards of quality, the volume of Suecia antiqua et hodierna used for the digitisation process was supervised throughout by conservationists from the Preservation Section at the National Library of Sweden.

Running from 2001 to 2004, the project digitised 353 pages of the Suecia plus 8 images for the binding and 13 images fort he list of illustrations, producing a total of 374 images. The aim of the project was to also to provide samples of copper engravings and etchings as well as the union of these two onto a single sheet. A total of 1.870 digital images were thus produced.

The collection is accessible via the Z39.50 protocol.

#### Swedish Historical Newspaper Archive

The Swedish archive has so far concentrated on a collection of newspapers called "Posttidningar", but will be expanded to include other titles. The digitized collection covers the period 1645-1721. It will also consist of material from the 19th century. The database is based on free text search capabilities with a possibility to view the image of the microfilm that has been digitized to a tiff format. Running since 1998, the project has so far produced 17.400 images of newspaper pages and will have produced a total of 150.000 images of newspaper pages by 2010.

The collection will be accessible through TEL by 2010.

# Travelling Trough the Ages

"Travelling through the Ages" aims to present the diversity of the National Library collections. The project showcases selected travel material: Manuscripts, diaries, maps, illustrations, posters, brochures, books and ephemera. Running from May 2006 to December 2007, has thus far produced digitals copies of 43 books (with 6.000 pages), 7 maps (with 2 images each), 338 photographs (with 2 images each), 500 drawings (with 2 images each), 40 posters (with 2 images each), and 77 ephemera. In the end the collection will consist of digital copies of 50 books, 20 maps, 362 photographs, 500 drawings, 40 posters, 469 ephemera, and 7 manuscript letters. The collection covers the time 1782 – 1920.

The collection will be accessible through TEL via the Z39.50 and OAI protocols.

# Codex Gigas, 13th century Bohemian manuscript

The projects aim was the digitisation of a manuscript, made accessible via tri-lingual web site with content aimed both at the research community and the general public. The codex will be exhibited in Prague in the autumn of 2007; the web site will open at the same time. The Codex Gigas manuscript consists of 612 pages which have been digitised into 1228 master files TIFF (614 raking light, 614 diffused light) and 614 webmaster files (Pyramide TIFF), and a large number of detail images. The projects runtime was between 2006 and 2007.

The collection will be accessible through its own web-site and through The European Library.

# 4.33 Switzerland:

The Swiss National Library has a list of five digitisation projects.

### **Swiss Poster Collection**

This collection comprises a wide variety of digitised images and catalogue descriptions of posters created by artists from Switzerland and elsewhere. The materials span decades and provide a snapshot from the end of the 19th Century until today. They cover tourism, exhibitions, publicity, cultural activities, sport, and political campaigns among many other topics. Beginning in 2000, this ongoing project has already digitised an estimated 45.000 individual posters and aims at a total of 250.000 posters when it finishes.

The collection is currently accessible through TEL via the Z39.50 protocol. There are plans to also make it accessible via the OAI protocol in the near future.

#### Numérisation du Journal de Genève

Running from 2006 - 2008, the goal of the project is to digitise, OCR and make available via the web the complete collection of the Journal de Genève (172 years / 1826 - 1998). By the end of the project around 100.000 items will have been digitised, producing about 600.00 digital pages.

The digitised version of the Journal de Genève is not accessible to other services and there are currently not plans to make it accessible.

#### Friedrich Dürrenmatt: Index of his paintings

The aim of this project is to produce digitised photographs of Dürrenmatt's paintings and drawings. Starting in 2002, this ongoing project has already produced 414 digital objects of drawings dating from 1930 - 1990.

The collection is accessible through TEL.

# Images et Voix de la Culture Suisse (IMVOCS)

Starting in 2002, this project aims at digitizing records, videos and films of the Swiss Literary Archives, the Max-Frisch-Archives and the Swiss TV stations SF DRS, TSR and TSI. Currently the resulting collection comprises digital versions of 1.525 sound recordings and 148 film and video recordings dating from 1950 - 2007.

The collection is accessible through TEL.

#### The photographs of Annemarie Schwarzenbach

Running from 2005 - 2007, this projects aim was the digitisation of around 7.000 photographs from the estate of Annemarie Schwarzenbach. The images date from 1933 - 1942.

# 4.34 United Kingdom:

The British Library has four large digitisation projects.

#### **British Newspapers**

Running from 2004 - 2009, the project will digitise 4 million pages of pre-20th Century newspapers. 3 million page files and approx 9 million article files have already been digitised.

The collection will be accessible via OAI protocol.

# Archival Sound Recordings

Running from 2004 -2009, this project will deliver thousands of hours of segmented recordings from oral history, field and location recordings of traditional and improvised music, rare or deleted classical and popular music recordings, soundscape and educational material. The material dates from 1920 – 2000. 12.000 of the expected 36.000 recordings have already been digitised.

The collection will be accessible via OAI protocol.

# Microsoft Digitisation Project

Running from 2006 - 2009, this project is undertaken in partnership with Microsoft to digitise out of copyright 19<sup>th</sup> century English literature books. The project has already digitised 300.000 of the expected 25.000.000 pages (approximately 100,000 books).

This collection will not be accessible through other services.

#### International Dunhuang

Running since 1997, this ongoing project is an international collaboration based at the BL to make information and images available of more than 100,000 items from Dunhuang and other Silk Road sites available on the internet. 50.000 items (mainly manuscripts) dating 100 BC - 1200 AD, have already been digitised.

This collection will not be accessible through other services.

# 5 Case Studies

# 5.1 British Library Digitisation Strategy

November 24, 2005

# Definition

Digitisation is the process of taking a digital photograph (photography or scanning) of a physical item. The result of this process is a digital image that can be displayed on a computer. If the digital image contains printed text or handwriting a process know as OCR (Optical Character Recognition) can be applied to the image to extract the underlying text. The OCR process generally does not produce perfect results but a good quality image of clear uniform printed text can yield very good results (95%+) however a poor quality image of a handwritten work can produce either no or completely incorrect results. If the underlying text is extracted from a document with enough precision through OCR this text can be used in a key-word search or linked to other documents to create a 'network' of similar documents. In addition further treatment of the image and associated text from a physical item can be used to extract individual stories from newspapers, identify photographs within documents, locate metadata such as subject, author, publication date & etc. None of these advanced processes can claim perfect results however on homogenous collections very high accuracy for all these techniques has been demonstrated.

This document is concerned with the issues surrounding digitisation of physical items and the collection, preservation and access to these digital surrogates. Minimal consideration will be given to items that are created in a digital environment (born digital). The digitisation strategy at the British Library does not exist as a stand alone piece of work. Broader strategies focused on content management and associated considerations, as well as, the overall British Library strategy take precedence.

#### Background

The British Library has been digitising parts of its collection for at least the last ten years. Each digitisation project has tended to focus on a specific collection item or set of closely related items and been managed as an individual and discrete ('one-off') project. These projects fall into a three broad and overlapping categories:

- A. Heritage items where the primary motivation is broader access to an item which normally could only be viewed in the treasuries gallery one page at a time. In some cases the resulting images are used in a 'turning-the-pages' project that allows a user to virtually turn the pages of a digital version of the manuscript, and adds a commentary, sound track and navigation. Heritage items that have been digitised include Beowulf (2,500), Guttenberg Bible (2,610), Anglo-Saxon Charters (398), Lindisfarne Gospels (about 20). (Images created).
- B. Larger scale digitisation projects where the motivation is to create a more comprehensive resource that can be viewed as a single collection. In general specific software applications have been created to aid navigation and viewing collections of this type. These projects have been primarily been funded externally with specific quality, quantity and timing objectives set for the project. Projects in this category include IDP (18,000), Collect Britain (100,000), Burney Newspapers (1,000,000), 19<sup>th</sup> Century Newspapers\* (2,000,000), JISC Funded Audio\* (3,900 Hours). (\*Funded by not completed).

C. Commercially funded projects that digitise parts of the BL collection and then sell or license the resultant digital collection. Typically these collections have some type of initial usage restrictions limited by a period of time (multiple years). Examples include Early English Books Online (Books published between 1475 and 1700 sold by ProQuest); Eighteenth Century Collections Online (sold by Gale).

In addition other digitisation projects with a range of motivations have also been completed. These include projects to digitise unique BL items funded by third parties, individual curator interest, and some for purely conservation reasons.

# What Has The British Library Digitised?

The results of the last ten years of digitisation at the British Library have created a valuable and enduring resource for scholars and the public alike. A limited funding environment has meant that small internally funded projects combined with larger externally funded projects have created a range of relatively disparate resources. In total we have around 10,000,000 digitised images which represents about 0.2% of our entire collection (150,000,000 items or approximately 5,000,000,000 pages).

#### Rationale for Digitisation

For almost all material in the British Library collections several common benefits for digitisation can be identified. These are:

#### 1. Digital Delivery

Access to an electronic surrogate, by an appropriate user interface and content management system, is much easier to deliver to a user's desktop than the physical original. Providing access to material in this way also saves the cost of retrieval from and return to the basements or shipping costs from other BL sites. In addition both local and remote access is equally simple however copyright and control issues have to be addressed for delivery outside of the reading rooms.

#### 2. Resource Discovery

Text-based items that have been digitised can have their underlying text extracted through an OCR process. This underlying text of the document (full-text) can be directly searched providing an additional method of resource discovery. This full-text searching is the basis for most popular internet search engines (Google, MSN, Yahoo, &etc).

#### 3. Reduction In Handling Of Original Physical Items

By using an electronic surrogate rather than the original document or manuscript it is likely that less direct use will be made of the original physical item improving its preservation prospects.

# 4. Building Digital Collections - New Scholarship

Bringing together a large collection of digital items (both born-digital and digitised) creates the possibility of cross-linking these items. These linkages can provide direct access to documents that are cited by other documents, documents that refer to similar subject areas, documents that have the same author and etc. This linkage process can create new areas of scholarship as cross-referencing of electronic resources can be achieved in a fraction of the time and cost that a manual process would using the physical originals.

#### 5. Global Collections

Digital collections can also be established with external organisations. With careful adherence to standards for interoperability larger international collections are possible and can be used

for a number of purposes including virtual re-unification of separated physical items as in the International Dunhaung Project.

### Costs & Issues with Digitisation

Although digitisation projects have many benefits there are a number of issues that all digitisation projects have to resolve to be successful. The primary issues are:

#### Cost

One method of comparing the costs of digitisation projects with each other is to compute the overall cost of producing each image that the project creates. This is simply computed by dividing the total cost of the project by the number of images produced. Using this methodology the larger BL projects have cost between £50 and £1 per image. The total project cost consists of many factors including the cost of the actual digitisation process which can represent between 30% and 80% of the project cost.

The BL projects have primarily used a manual process to perform the digitisation process. A number of organisations have now purchased custom build machines that dramatically speed up this process for certain types of uniform material (books especially) with a cost of around 5p per image having been reported for these machines. In addition some organisations have shipped materials to countries with low labour costs further reducing the cost of the overall project.

As technologies and methodologies improve it seems likely that the raw cost of digitisation will continue to decline with the prospect that digitisation of uniform material will ultimately reach the 1p per image range for large projects. However, projects that require special handling of the physical items or have a range of physical dimensions are likely to retain a high cost per image profile.

#### 1. Content Management

The act of digitisation creates a digital surrogate which alone is just a digital file that contains coded digital information (bits). To take full advantage of the content contained within these digital files additional information (metadata), computer hardware, software applications, and communications networks are required. Metadata is required to describe various properties of the digital file and of the content it contains. Computer hardware is required to store the files and metadata while software applications are required to process, display, search and access the content. Finally communications networks are required to transport the content from a central storage location to the ultimate user's computer.

Each digitisation project must consider the full IT infrastructure that is required to provide access to the results of the digitisation process. This IT infrastructure can represent a significant proportion of the cost of a project both at initial deployment and sustaining the resource over a number of years.

# 2. Copyright

The copyright law is a formidable barrier to the use of 'in copyright' material for digitisation. Although the act of digitisation does not in itself create a tangible copy of the physical item that is being digitised, once the digital image is printed or rendered onto a computer screen a 'copy' is deemed to have been made.

Generally digitisation projects avoid copyright issues by assuming a 100 year rule (life of author plus 70 years) although copyright material can be licensed if the copyright holder can be identified. Recently legislation has been introduced to clarify the situation with

unpublished manuscripts where copyright expires in 2039 or life of author plus 70 years which ever is later.

### 3. Digital Preservation

Sustaining digital collections over the long term (3 years and beyond) requires special consideration with respect to the content management system used to store, preserve and render the content. At some point in the future these systems will no longer be supported by the hardware and software providers thus requiring some form of upgrade. This upgrade process will probably require a format conversion (migration) of the digital files created by the digitisation process. Due to the improvement in technology over time it is assumed that this format conversion will cause no loss of information or degradation of the content contain in the digital files. However, the exact frequency and cost of this process is hard to estimate but it is assumed that this cost will be built into the normal operational budget of the Library and no special financial provision need be made for this process.

Although digitisation projects have a number of costs and issues to resolve these can be generally quantified and managed. The largest barrier to digitising the BL collections is the copyright law which greatly reduces the content that can be selected for the digitisation process.

# What Should The British Library Digitise?

A number of different approaches could be taken to decide which content the British Library should select for future digitisation projects. It is assumed that commercial or cost justified projects will proceed if the business case is strong enough as all additional digitisation activity expands and enriches our digital holdings to some extent. Although the selection for digitisation projects will in some measure be driven by what funding is available both internally and externally there appear to be two primary reasons for digitising our existing collections are:

- 1. Broader access to the iconic or heritage items at the Library that could not be achieved by access to the physical items due to preservation considerations. The target of this approach is both the researcher and the public. In this category the digitisation process is attempting to create a 'virtual' version of the original physical item. The chief considerations in this type of project are around the accuracy and quality of the digitisation process, as well as, the authenticity of the 'look and feel' of the application that is used to display the items. This is where products such as the award winning 'turning the pages' help create the virtual appearance of the original.
- 2. The creation of a critical mass of content for education and research purposes. In this category the content is more important that the exact look and feel of the original. Although it may be important to show the content in the context of the page (newspaper article in context of the complete newspaper) the exact font used to render the text or type of paper in the original is of secondary importance to the content itself. The quality of the digitisation in this category should be primarily driven by the needs of the OCR process, rather than, the look of the final image. In this category searching, cross-linkage between items and other valued added navigation is critical to achieving the full value of creating a large digital collection. Projects should ensure sufficient investment in this licensing or creation of tools to ensure much more than a collection of digital images is created.

A project in this class could include the digitisation of the Shakespeare Quartos and all the surrounding literature (works published at the same time, source material for the Quartos, Folios & etc.). The value that the BL can add to this type of collection is the comprehensive nature of the sources, the linkage and navigation that can be constructed

between the various items and the additional presentation and analysis that would only be practical if the items were digitised, (for example an analysis of the language and words used in the documents.). A collection of this scope would create new opportunities for scholarship due to the dramatically enhanced access to a large collection of material from the same era.

Over time the Library has been shifting resources from the single iconic or heritage items (category 1) to the creation of a critical mass of digital content (category 2). Although there will always be a case for digitising more of our heritage items it is likely that much more value will ultimately be gained by creating a larger collection of digital material and creating the value-added features discussed. This strategy therefore proposes that the primary focus of fund raising and partnership for digitisation be shifted to the creation of a critical mass of digital content. This investment must include the creation of analysis and presentation tools, as well as, the digitisation of the content itself.

This approach does not preclude small highly focused project, funded if necessary by GIA (Grant in Aid), that focus on the single iconic or heritage items (category 1). Small projects of this type fulfil our mission of making our collections more accessible to a much larger audience.

#### Content

Some analysis has been done to identify areas of our collections where large scale digitisation could be performed at the request of Google. (Ann Clark - November 5, 2004). This analysis showed that although various groups of content could be selected various issues such as prior digitisation by commercial partner, in copyright material, likely to be digitised by others and limited utility were identified. Google has undertaken to digitise large collections of public domain material particularly books, periodicals and government publications where items are of a uniform size and free of copyright constraints.

In considering areas of the Library's collections that are unlikely to receive attention from Google, have real value for the research community and potentially help our cost and business objectives the most obviously content class to digitise is newspapers. Not only is there a huge collection of at risk material that could be digitised, but also, the time-based nature of newspapers makes them ideal to 'map' and link to other historical collections and items. The biggest difficulty with newspaper digitisation is the variety of sizes and relative fragility of the material. In addition due to copyright considerations only 19<sup>th</sup> century and earlier newspapers can be considered for digitisation.

Another unique area of the Library's collections that also meets many of our digitisation objectives is audio. Audio resources are generally harder to manage (size, format, and media) and the resultant resource is more difficult to search as searching an audio database currently relies on metadata. However for speech based material much progress has been made in speech recognition technology and automated full-text creation of audio content may soon be possible making searching of speech based items much easier.

#### Ownership

Although asserting a copyright claim over the results of digitisation may prove difficult, because the digitisation process per se does not add intellectual content, a database of such content with associated metadata could be licensed to third parties and hence generate additional revenue for the Library. This strategy therefore proposes that the results of the digitisation process are not 'given away' but that a set of computer defined interfaces to the metadata and digital images are published so that interested third parties can create

applications which use this information. This approach may generate additional interest in this content and hence increase its value.

### Large Collections

As larger digital collections are inherently more valuable than smaller collections partnership arrangements with other libraries and archives should be entered into to create larger 'virtual' collections. However as the 'value added' tools that create links, sophisticated navigation, and content analysis are more effective if the content is brought together into a single location these partnership arrangements should be focused on creating a single joint collection rather than multiple distributed collections that are connected by metadata. The value added tools should be created where possible by commercial or research based efforts rather than being directly built by the BL.

Larger digital collections will contain both digitised and born-digital items. Greater interoperability between digitised and born-digital collections can be achieved by using the same structural and metadata standards for both types of digital collection. As born-digital items will tend to predominate over time, standards should be drawn from these classes of content rather than the results of digitisation.

#### Access

As digitisation projects create digital surrogates of physical collection items that are primarily accessed by an electronic catalogue considerable value would be added to the catalogue by the inclusion of the digitised items where available. This has the potential to reduce the cost of basement retrieval as researchers may be satisfied by access to a digital surrogate rather than the physical item itself.

#### Recommendations

The following recommendations emerge from this strategy:

- 1. Cost justified digitisation projects supported by a complete business case should not be hindered by this digitisation strategy but reuse of this type of content should always be explored.
- 2. Large-scale fund raising for digitisation should focus on building large digital collections.
- 3. Fund raising for small digitisation projects to broaden the access to our collections or to assist with cultural restitution issues should proceed on a case by case basis.
- 4. Considerable attention should be paid to the OCR, metadata extraction, content analysis tools and navigation applications in any digitisation project.
- 5. A large collection of digitised newspapers should be a focus of our digitisation efforts.
- 6. The technical challenges associated with creating a large collection of publicly accessible audio material and providing easy access to this material should be investigated.
- 7. The development of software applications (eLearning) that use the BL collections should be encouraged by publishing a computer accessible interface (web service) to our digitised collections.
- 8. Partnerships with other organisations should be formed to create larger virtual collections. The implementation of such an arrangement could result in a set of links from one collection to another. For example 19<sup>th</sup> century newspapers at the BL with 19<sup>th</sup> century parliamentary papers at Southampton University.
- 9. The license terms for commercially funded digitisation projects should ensure that any restrictions on use of the resultant digital collection should last for a limited period of time.

# 5.2 Norwegian National Library

# Digitization of books in the National Library – methodology and lessons learned September 2007

#### Author: Svein Arne Solbakk

#### The digital national library

The vision of the National Library of Norway is to be a living memory bank, by being a "Multimedia Centre of Knowledge" with focus not only on preservation but also on mediation.

To succeed with this ambition, one of our main goals is to be a digital national library, as the core of a Norwegian digital library. A digital National Library is just another way of being a National Library. It is then important to have as much digital material as possible, not only historical material but also the modern part of the cultural heritage, to give access to as much material as possible to as many as possible whenever required.

The Norwegian National Library has therefore started a systematic digitisation of the entire collection. Based on a modern Legal Deposit Act we receive everything that is produced and of interest to the public, be it books, newspapers, periodicals, photos, films, music or broadcasting. All broadcasters with a license in Norway may be asked to deliver copies of their programmes for preservation, and we have an extensive collaboration with the national broadcaster NRK about preservation and dissemination. We are also to preserve those digital signals that are never converted into anything else before they reach the user. Starting in 2005 we have harvested large parts of the Norwegian web domain .no.

Some of this material is delivered to us in digital formats, and we are about to have a relatively large digital collection when it comes to the audio-visual, but little when it comes to printed material.

The issue now is how we shall take advantage of having established a digital repository for preservation in order also to give access to this material both to scholars, students and the public. This is a challenge in many aspects, but mainly with regard to copyright, technical quality and dissemination pedagogic.

Our aim to establish a unified Digital Library for all our different media, with our Digital Long Term Repository as a basis, will require the following strategies:

- As soon as possible (depending on funding, of course) to digitize all our collections.
  We do this systematically
- To digitize on demand
- To negotiate with the publishers to get as much material as possible deposited in digital formats
- To be a digital archive for other institutions, e.g. publishers and newspapers
- To find strategic partners to cooperate with, financially and on know-how
- To be a trusted repository for digital material in the Norwegian society
- To give access to as much as possible of our cultural heritage, e.g. through search engines
- To negotiate with the right-holders to give access to material that is not yet in the free, so that also modern books, films, and music can be searched

To be able to do this, it is also important to take part in discussions and developing the pedagogic of the net. Libraries' knowledge about the users' needs is a good experience basis on which to develop search methodology in the crossing between metadata and the methodology of search engines. Our aim is to give access to information, knowledge and experience on a given topic across media types.

#### Digitizing the Norwegian cultural heritage

#### Scope

For more than 10 years the National Library has been digitizing a wide range of media. The main focus for this digitization has been photographs, sound recordings and microfilm (newspapers). As a result, the digital collection today contains more than 150,000 hours of radio, more than 300,000 photographs and more than 1,000,000 newspaper pages. In addition, we have digitized more than 25,000 books over the last year.

Still, these are modest numbers compared to what we have now set out to accomplish. If we include the estimated growth of the collections through analogue legal deposit until digital legal deposit is up and running, we expect the amount of material to be digitized in the National Library digitization programme to be:

450,000 books 2,000,000 periodicals 4,700,000 newspapers (more than 60,000,000 pages) 1,300,000 pictures (photographs and postcards) 60,000 posters 200,000 maps 4,000,000 manuscripts 200,000 units of sheet music 1,900,000 leaflets 1,000,000 hours of radio 80,000 hours of music 250,000 hours of film and television

Gifts, purchases or deposited material during the program period will add to these numbers.

In order to accomplish such an ambitious program of digitization during a foreseeable period of time, the digitization activity has been greatly increased. As much as possible, we are now streamlining the process from when an object is selected for digitization until it is placed digitally in our digital long term repository, simultaneously offering web access for authorized users. With the multimedia collection of the National Library this poses special challenges, as we need to establish separate production lines for the digitization of different types of material. In addition, there are usually several variants within each type of material, which again demand different adjustments.

Putting these things in place is a challenge, both in terms of technology, logistics, organization, manpower and financing.

#### Legal Rights

The National Library has a legal right to digitize the collection for preservation. However, to make the collection available in the digital library, it is necessary to make agreements with the copyright holders whenever the material is still under copyright protection.

The old material with no copyright restrictions will be made available for everyone in the digital library.

The Legal Deposit Act states that every object subject to legal deposit may be made available for research and documentation. This is also valid for the digital material subject to legal deposit. However, in the digital domain there are still several unresolved challenges, e.g. privacy questions related to the interconnection of several extensive data sources (e.g. Norwegian internet pages), and the risk and consequences related to misuse through illegal copying of digital content, which are much more extensive than for paper based documents. These are the reasons why we still do not have good solutions for access to digital legal deposited material via the Internet.

When the documents are published in a traditional way but deposited in digital formats, we have to make agreements with the copyright holders to be able to give access to the digital documents. In this case we try to get agreements giving us at least the same rights that we have for material subject to legal deposit.

The National Library also works to get agreements with the copyright holders making it possible to give a broader access to the modern part of the cultural heritage. An example is the <u>agreement</u> between the National Library and several copyright holders organisations on giving access to books and journal articles related to "The High North". This agreements gives the National Library rights to make available approximately 1 400 complete works in the digital library. These works will be used to evaluate user behaviour and the frequency of use of the digital material. The result of the evaluation will form basis to negotiate on more permanent agreements with the copyright holders.

#### Financing

Preliminary estimates suggest a total cost of around 1 billion NOK for the National Library's digitization programme. Around 60% of the cost is for the digital storage, the purchase of digitization equipment and software, the development and integration of systems that will be part of the process of digitization and post-processing, in addition to wages for carrying out the digitization and money for some external commissions. The remaining 40% of the cost will be needed for the indexing of the material in order to establish the metadata required for retrieval, and for fetching the material from the collections, for the necessary conservation, and for the return of the material to the collections.

Through a re-channelling of activities towards the strategic initiatives, the National Library has prioritized 12 M NOK a year for the digitization programme. In addition, we have received a grant from the Ministry of 3 M NOK for digitization in 2007. Finally, on top of this, we have a budget of 13 M NOK for the purchase of digital storage in 2007, so the total budget for the initiative is 28 M NOK this year.

Preliminary estimates show that the whole digitization programme can be carried out in approximately 20 years. Naturally, there is considerable uncertainty associated with an extensive and long-term effort such as this. We expect developments in technology both in terms of digital storage, digitization equipment and of tools that will enable retrieval in a way that reduces the need for manual indexing. However, it is difficult to accurately predict the consequences of such a development, in terms of both cost and increased efficiency.

#### Status

Large parts of 2006 were spent on carrying out the process of inviting tenders for the purchase of digitization equipment, software, digital storage and other ICT solutions required. In addition, a project manager and some system developers were hired late 2006.

In the new digitization programme, books were given priority. A closer scrutiny of this work can be found in a separate chapter.

Following an invitation for tenders we have outsourced the digitization of microfilm to a German company (CCS). So far 400,000 pages from the newspapers Aftenposten and Adresseavisen have been digitized. However, this number will be increased in the fall of 2007.

In addition, the digitization of photos has been made more efficient through the purchase of software for automatic post processing of digitized material, and through an upgrade of the digital cameras used for this work. We have also bought efficient equipment for the digitization of 35mm roll film negatives.

We are also working on an upgrade of equipment for the digitization of music, radio and moving images. An important part of this is a more automated connection between the digitized material and metadata that describes the material.

Other types of material will be digitized on demand, but so far we have not established adequate production lines that will enable speedy transfer of digitized materials into our long term digital repository, and immediate availability in our digital library.

The systematic digitization is the foundation of the digitizing process. In addition, we will carry out on-demand digitization which can be both user initiated and based on the strategic priorities of the National Library. On-demand digitization will be given priority over the systematic work. This applies to all parts of the digitization process. However, at this time it is only true for those types of material for which production lines have already been established (books, photography and sound).

#### Future plans

Of most importance in the short term is to make operative all the production lines for the digitization of books. What remains to be done is mainly to establish a regime for quality assurance of the digitized books, and an adequate handling of exceptions through all of the production line. This includes all stages of the process starting with ordering material out from the stacks and ending with the digital versions of the material residing in the digital long term repository and becoming accessible via the NB digital library.

We have also started work on establishing production lines to preserve both digitized newspaper pages and digitally deposited newspapers in the digital long term repository, and to make them available in our digital library. In addition to the newspapers that have been digitized from microfilm, we have established a pilot of daily legal deposit from two leading national newspapers in preservation quality PDF format. When this pilot gets into regular operations, we will work towards extending it to more newspaper titles.

We also plan to start the work of establishing a production line for the digitization of manuscripts. This is a type of material in heavy demand from our users and which will be of relevance in connection with upcoming authors' anniversaries in the years ahead. For this material, in addition to digitization, we must establish the necessary metadata in order to achieve satisfactory search and retrieval in the digital library.

Later again, we will establish production lines for the digitization of periodicals and posters. In addition, we plan to automate parts of the existing production line for the digitization of photography, in order to increase efficiency.

Another very important activity will be to initiate legal deposit of new publications in digital formats of preservation quality. Besides the pilot setup with daily legal deposit from two newspapers, today we have an operative solution for the digital deposit of radio in preservation quality. Also, audio books produced in digital format are delivered from the Norwegian Library of Talking Books and Braille.

We also have an ongoing dialogue with publishers and TV broadcasters, planning to start digital deposit of books, periodicals and television during the coming year.

#### Establishing a production line for the digitization of books

In establishing a brand new process for the digitization of books, it was important to achieve a well integrated production line capable of covering all the steps through which the book would pass before it was preserved in the National Library's digital long term repository, and also becoming available to authorized users in the digital library. We wanted to automate as much as possible of the data flow and processing of the digital book, at the same time leaving enough flexibility in order to accommodate new process stages in the production line if the need should arise.

The processes we wanted to include were: selection for digitization and ordering of the books in question, fetching of material from the stacks, transport, extraction of metadata from the catalogue, digitization, OCR treatment and structural analysis, format conversion, the generation of preservation objects, ingest to the digital long term repository, notifying the catalogue of the digital object, and indexing OCR text and metadata in our search engine.

We also saw that different types of scanner technology gave very different efficiency in the digitization. If books could be dismounted, the scanning would become at least 10 times faster.

#### Fundamental decisions

#### Preservation - quality and format decisions

The digitization programme is a part of the National Library's strategy for preservation and dissemination. Digitization will make preservation of the collection more efficient and less vulnerable to physical deterioration. This means that the digitization must be performed in a quality so high that, after digital preservation, it must be possible to satisfactorily recreate the properties of the original at a later time. At the same time, the digitization and the chosen formats must fulfil the requirements for dissemination of the material.

We have chosen to digitize books at a resolution of 400 dpi and a colour depth of 24 bits. Our preservation format is JPEG2000 with lossless compression. The preserved image is not processed or reduced in any way after the scanning process.

By choosing losslessly compressed JPEG2000 instead of uncompressed TIFF as a preservation format, we will reduce the need for digital storage by about 50%. For the whole digitization programme this means savings on the order of NOK 70 M. Through practical tests we have demonstrated that we are able to convert from the JPEG2000 format back to uncompressed TIFF with absolutely no loss of information.

An argument against using JPEG2000 is that one bit error in a JPEG2000 image file will be able to destroy the whole image, whereas a bit error in an uncompressed TIFF image file will not affect more than one pixel. With the storage policy in our digital long term repository, we find the risk of bit errors to be negligible.

The quality requirements for preservation are far stricter than those normally used for dissemination. Also, the formats used for dissemination have a shorter lifespan, partially because new formats are developed with more advanced compression algorithms that offer better quality with less data, and partially because there are new versions developed of the existing formats, with better algorithms and better quality. Therefore, we have chosen to generate the dissemination format from the preservation file at the moment a user asks for the image. Using this strategy we will easily be able to switch dissemination formats by replacing the algorithm that generates the dissemination format.

In today's solution we generate a JPEG file of the desired quality for viewing (typically around 200 Kbytes) from the JPEG2000 file located in the digital long term repository (typically around 20 Mbytes).

# Perform our own digitization or hire others?

Today, there are several organizations offering cultural institutions to digitize their book collections for free or at a very low price, in return asking for the right to store the digital copies and to offer search and display of the books. Examples of this are Google and the Internet Archive. In the National Library's book collection there are relatively few books which are no longer protected by copyright, and to which free access accordingly can be offered. Out of the 450,000 titles that will be digitized under the digitization programme, at present only around 5,000 titles are no longer copyrighted. It is important to the National Library's digital long term repository. Access to such books will only be given in accordance with agreements made with the right holders. Also, it has been a fundamental principle for the National Library that other service providers must be given equal opportunity to offer services based on our collections. The sum of these considerations has meant that we have not deemed it natural to cooperate with this type of organizations for digitization. Still, we will invite them to disseminate what we have digitized, on par with other actors.

It has also been part of the picture that the National Library has been able to reassign existing staff to tasks associated with the digitization production line. This has meant that the total cost of in-house digitization of books is lower than it would have been for outsourcing the digitization.

For other types of material we have chosen to outsource digitization. For instance, this applies to digitization, OCR and structural analysis of newspapers on microfilm.

#### **Dismounting books**

In order to achieve efficient digitization, we have chosen to dismount books for digitization if we have at least three copies in our repository library. The dismounted copy is then thrown out after digitization.

When we have fewer copies, the books are scanned manually, with operators opening the books and scanning two pages at a time. The most vulnerable books are scanned under the supervision of a conservator, and any necessary conservation measures are performed before or in connection with digitization.

The process of preparing books for dismounted scanning is more labour intensive than preparing books for manual scanning. Special operators are required for deconstructing the books (separating the binding from the rest of the book, removing glue with a hydraulic cutter), and the scanning of the binding is a separate process. Because of this, 4 operators are needed in order to feed one scanner of dismounted books. Still, the total picture means lower cost and higher production than if the same resources had been used for manual scanning.

As of today, around a quarter of the National Library's book collection can be dismounted for digitization. In order to improve this ratio, the National Library plans to invite Norwegian libraries to contribute copies of books we have too few of. In this way, book digitization can be carried out faster and more efficiently in the National Library, and the libraries will be able to free up space in their stacks.

We have not yet tried out scanners that automatically turn the pages of whole books that have not been dismounted. Such scanners are developing rapidly, and this technology is becoming very interesting. For books that can not be dismounted this technology may offer considerably higher production per operator, both because the scanners are fast and treat the material gently, and because one operator can serve several such scanners at the same time. However, investment costs for this type of scanner are still high.

# OCR and structural analysis

In order to allow for full text search, all digitized books undergo a process of optical character recognition (OCR). In the regular production this process is fully automated, and there is no manual quality control or correction phase. The text that results from the OCR treatment is indexed in our search engine together with the metadata. If a search gives results from the text, the page of the book where the text was found will be displayed, and the user can browse the book from that page.

Also, an automated structural analysis is performed, during which any table of contents is annotated, and page numbers in the book are verified so that the user interface relates to the actual pagination of the book. This is also an automatic process. The software allows for very advanced structural analysis, but at this stage it is not feasible to increase complexity without also applying extensive manual quality control and post control. For selected parts of the collection we will perform more advanced structural analysis including annotation of several parts of the documents, again allowing for more advanced navigation of the books in the user interface.

At present, about 2,000 - 3,000 books are digitized every month in the National Library. With this volume it is not possible in practice to perform manual post control of the OCR and structural treatment.

Both OCR and structural analysis are performed using software called docWorks.

# The National Library's digital long term repository

The digital long term repository is an infrastructure for the long term preservation of digital objects.

Everything that is digitized as part of the National Library's digitization programme is to be preserved as digital objects in the National Library's digital long term repository.

The digital long term repository separates the use of digital content from the technology which is employed for the storage. This allows for easy migration to new generations of storage technology without affecting the systems for retrieval of the digital content. This is very important in a 1000 year perspective.

All digital content is stored in three copies on two separate storage media in the digital long term repository. At present one copy is stored on disk while two are on tape.

#### Search engine

In order to realize search across large aggregations of data, the National Library has chosen to employ search engine technology rather than traditional data base solutions. Both metadata and full text are indexed by the search engine, and searches are performed without regard to types of material. We have also implemented a so-called drill-down search in the metadata. Metadata for objects satisfying the search criteria is analyzed in real time during searches, and alternate paths of navigation and different ways of narrowing the search results are built and displayed to the user.

The search engine used is delivered by FAST.

#### Authentication and authorization – access control

The National Library has chosen to employ role based access control. We have chosen to cooperate with FEIDE, which is the national infrastructure for authentication and authorization of users at Norwegian universities and other institutions of higher education. This means that we can open for access to defined groups of scientists at the universities and institutions, without knowing every individual. The universities are responsible for the authentication of those who satisfy the requirements for the different roles which have been defined within the system.

If we had picked an access control based on user names and passwords with associated rights for individuals, the National Library would have needed to spend considerable resources on the administration of users and the maintenance of access rights.

At present we have not implemented this access control solution for digital books. Therefore, we have only indexed and made available books that are either out of copyright, or covered by agreements with the right-holders on giving open access.

The software used for user authorization and authentication was supplied by SUN Microsystems.

#### Production lines Prioritization

The basis for the digitization is the systematic selection. We have chosen to start with the oldest material in order to quickly get the material which is out of copyright into our digital library. In addition to the systematic selection we give priority to material on the basis of external requests and strategic priorities in the National Library. Especially prioritized material is placed at the head of the queue, in front of the systematic selection.

A special case is the agreement between the National Library and several rights organizations regarding books and articles related to the "High North". These works have been given special priority in the digitization process.

#### Ordering and extraction from the stacks

In order to achieve an efficient extraction of material for digitization, a special function has been implemented in BIBSYS, our book cataloguing system. Here we can order a given number of titles for dismounting, chosen automatically by the system among titles that there are a sufficient number of copies of, starting with the oldest. In addition, we can order single titles to be given special priority (both on extraction from the stacks and through the whole production line). Adaptations have also been made to the software that runs our automatic storage system for books, so that the operators can give top priority to remote loans, and then extract books for digitization. This system is integrated with the catalogue, so that books ordered for digitization automatically appear in the interface for the operators of the automatic storage system.

Already we have spent more than one man-year on system adaptations of the catalogue system and the software for the automatic storage system.

#### Digitization

For the books to be dismounted we have two hydraulic cutters, three binding scanners (i2s Copibook) and two auto-feed scanners (Agfa S655). For the page-turner scanning we use i2s Digibook Suprascan. Five of these are A2 scanners for normal page-turner scanning and one is an A0 scanner for special material. The A0 scanner is operated by conservators.

Before the bindings are scanned, all metadata for the book is retrieved from the catalogue (BIBSYS) by way of a bar code assigned to every book in BIBSYS. A digital ID for the book is generated and inserted into an XML file together with the metadata obtained from the catalogue.

In the case of autoscan, a sheet of paper with the bar code of the digital ID is printed out after the scanning of the binding. This sheet is put on top of the stack of the dismounted book. When the bar code later is sent through the autoscanner, it is identified, thus making an automatic connection between the metadata file and the scanned binding.

In the case of page-turner scanning, the binding and the content of the book are scanned on the same equipment. This process too fetches metadata from the catalogue and generates an XML file with metadata that accompanies the book through the rest of the process.

# OCR/DSA

After digitization the digital book and its accompanying metadata will be placed on temporary storage ready for further processing. The books must be manually imported into the docWorks software, but from there on the processing of most books is fully automated. Manual operators are only employed for handling of exceptions when the software calls attention to errors in processing (i.e. when processing failed to stay within the defined tolerance limits).

In addition, operators are used for quality control of special parts of the collection that we want to process further.

Books of high priority are placed in special folders that are imported before the standard systematic digitization.

After OCR treatment and document structural analysis, losslessly compressed JPEG2000 files are generated from all the image files of the book. This is the format used for preservation.

#### Digital preservation

After processing in docWorks, a METS object containing metadata, the digital book, the OCR processed text and structural information is generated. This object is placed in the National Library's digital long term repository for preservation.

Simultaneously the catalogue is updated with the digital ID of the book.

#### Indexing

At regular intervals, an OAI import of catalogue data is performed. If this import finds that a book has been updated with a digital ID, a process is started in order to fetch the metadata

and text of the book from the digital long term repository and index both, making the book available for search in the digital national library.

#### Important lessons learned this far Scope, complexity and implementation

Implementing an integrated production line for the digitization of books with a high degree of automation turned out to be far more extensive and complicated than anticipated.

In order not to waste time, as soon as the decision to establish a production line for the digitization of books had been made, we launched activities aimed at realizing the first part of the chain of production (ordering, extraction, transportation and the digitization itself). In order to realize the efficient extraction of material, adaptations had to be made both to BIBSYS, which is the National Library's book cataloguing system, and to the software that runs the automated storage system from which the books are retrieved. This made us dependent on two external suppliers, which had consequences for the rate of development.

Carrying out an invitation for tenders for the purchase of scanners is also a long and time consuming process. We could not develop the method of digitization until it was clear what kind of equipment we would be using, and then we had to get the first scanners installed before we were able to terminate the implementation and start testing.

After the first part of the production line was in place, we began test production. Having aimed at a high production rate, we soon found ourselves with large amounts of data in temporary storage. While waiting for the rest of the production line to be put in place, we had to establish temporary routines for the safeguarding of the digital content.

In order to set up the rest of the production line and the functionality of our digital library that would facilitate the search and display of books, a multitude of development activities had to be launched. Some examples: Installing and putting into operation the software for OCR and structural analysis of documents, and the integration of this system into the production line, the generation of preservation objects based on the METS standard, the process of ingesting the METS objects into the digital long term repository, a setup for updating the catalogue system with the digital ID, OAI harvesting of metadata from the catalogue system, the process which retrieves text and metadata from the digital long term repository for books which have a digital ID, the indexing of these, and the development of necessary functionality for the search and retrieval of books in the digital library. At the same time there arose great pressure to quickly get to see the results of the digitization already under way, which led to the pushing of deadlines and, during one period, a very high level of stress among the section responsible for development.

After the functionality for the display of books in the digital library was in place, it soon became clear that this would become a very interesting service which would provide a"lift" to our digital library. Also, we had received considerable media coverage of the digitization, and expectations of seeing results were great, both from external users and from inside the National Library. We therefore decided to launch the service, even though the production line still was under development, meaning that many manual operations were needed to speed a book through the full production line. The service has functioned well, but the expectations of quickly reaching a large volume of digital books in the service were not fulfilled. This was mainly due to the fact that the production line was not fully implemented, and accordingly not put into regular service.

With hindsight it is easy to see that we ought to have had a greater focus from the beginning on the unity of both the whole production line and the necessary functionality in the digital

library, and that from the start we should have better assigned ownership to the timeline in this development work in the whole organization.

#### Actual efficiency

With basis in the specifications of the digitizing equipment, we assigned production targets already from the start. These took into consideration that this was a ramp-up phase. Still, it turned out that some factors we had not been aware of reduced the overall efficiency. This became most apparent when looking at the automated scanners.

The book pages were on the whole thicker than the reference paper used to measure the scanners' specifications. This led to a decrease in paper feed speed, which made a great impact on the daily production numbers.

Since we started with the oldest books, we had an issue with them being very dusty. This meant that the scanners needed considerably more cleaning and maintenance than expected by the supplier. This in turn meant reduced production time per day, and therefore reduced daily production compared to expectations.

Our plan called for the scanners to work as continuously as possible through the whole working day. This was to be realized through operators relieving each other at the scanning stations, taking their breaks at different times. This was a new and unfamiliar way of working, creating some resistance among the operators. In practice we have not been able to realize this well enough, and also this has contributed to reduced production time per day on the scanners compared to our prognoses.

Actual production has been between 60% and 80% of our stated production goals.

#### Quality

During initial testing, all pages of the books digitized were subjected to quality control. Since then we have not had routines for quality assurance of the digitization work. After the functionality of the digital library became capable of displaying the books, there have been random quality checks of the quality of what is available through the service.

Books that were displayed in the digital library revealed that the compression of the digital display copies had been performed using inadequate settings. The visual quality of the images was lower than expected. After adjustments had been made, the results became much better.

Obviously, the quality of the digital books is closely related to the quality of the originals, Our algorithm for the automatic selection of books for digitization does not take this into account, and we accordingly run the risk of extracting inferior specimens for digitization even if there actually are very good copies in the collection.

The scanners that digitize the dismounted books scan both sides of every sheet in one operation. This means that two different digitizing units are processing the two sides of a sheet. It has proved quite difficult to calibrate these two units identically, resulting in colour variations between the pages. This has improved immensely since the test phase, but the problem has not yet been fully solved. Experiments have been made with scanning a reference sheet at the start of every book, to facilitate subsequent automatic colour adjustment. So far these experiments have not yielded the desired results, but we will strive towards a solution.

When the production line enters a phase of regular production, we plan to establish random quality control of digitized books.

# OCR/DSA

From the outset, we had planned for fully automated use of OCR and document analysis tools. These kinds of tasks had never before been performed at such a large scale, and we had no prior experience using tools of such advanced nature.

The first challenge was to establish a large scale production setup with eight instances of the software on eight blade servers. This was necessary in order to achieve sufficient processing capacity, but it turned out to be harder than expected to make this stable and operative.

The next challenge arose when it turned out to be impossible to run the system fully automated. This created an unforeseen need for resources, and manning this task caused us some headaches. We spent some time figuring this out, and had to postpone the planned training on the system. This in turn gave us a short-term competence problem since it is a complex system with very advanced functionality. The answer was in part found through close contact with the supplier. This challenge has now been overcome, and training has taken place.

Our initial expectations of precision in the fully automated structural analysis have so far not been met. Advanced structural analysis can be done, but with such a degree of uncertainty that manual quality control is absolutely necessary. The more advanced the analysis you want to employ, the more manual quality control you will need. Accordingly, we will use this only as an exception, in special dissemination projects. A simple calculation shows that a post control that takes on average 15 seconds per page will in total require an effort equal to 18 man days per day at the present production level. For this we don't have resources. So we have been forced to stay at an absolute minimum level by only requiring correct pagination in the digital library service and that the table of contents must be directly linkable whenever a table of content is present.

So far we have focused on OCR and DSA of publications in Latin letters. Here we have acceptable precision in the letter recognition. For Gothic letters the results are worse, but even there we have a degree of recognition that opens up interesting possibilities for free text search. We are running separate configurations of the system for Latin and Gothic letters. Books are categorized when the bindings are scanned, and then they are routed to the correct configuration. We expect there to be a potential for improvement by further training of the software and more advanced configuration of the system.

# The digital long term repository - scaling and performance

So far we have operated a separate instance of the DSM (digital long term repository) for the digitization of books. The use so far does not show performance problems for the DSM, but the usage is expected to rise considerably compared to the present traffic, as the volume of digital books offered in the service increases.

Since we do not have in place an access solution (user authorization and authentication) for books, we have so far chosen to only place in the DSM books that we are allowed to give full access to in the digital library. Some of the logic today is that when a book is placed in the DSM its digital ID is entered into the catalogue. This in turn lets the book be fetched automatically from the DSM for indexing to the search index of our digital library.

This strategy means that most of the digitized books are still in temporary storage, albeit subject to the same storage policy as those in the DSM.

When a user asks for a certain book page at a certain quality (at present the interface gives a choice of three quality levels), a JPEG file of the desired quality is generated from the JPEG2000 file in the DSM. So far we have not implemented intelligent pre-caching or placing pages in a buffer outside the DSM to increase performance as perceived by the user, but this may be an interesting future development.

So far the rate of progress in technology has given us enough flexibility that we can operate a one machine solution for the digital long term repository (DSM). But this makes it vulnerable when errors arise. We are therefore continually evaluating the possibility for more robust solutions.

#### Statistical tools – production supervision tools

So far we have been using simple UNIX tools to generate the statistics necessary in order to supervise production. We are now considering the development of more advanced general production supervision tools that will at any time give us updated information of where a given object is in the process. We should also be able to use them to generate production statistics from all stages of production.

#### **Exception handling**

With a few exceptions everything in the production line that can be automated, has been automated. Still, some times there can arise deviations in all stages of the production line, and these deviations must be handled and followed up manually. This has been one of the greatest challenges in the production so far. We are now working on establishing routines and assigning precise responsibility for such follow-up in the line organization.

#### Summary

The implementation of a production line for books has had its fair share of problems. However, we have learned from our mistakes, and today we have in place an advanced production line for books. By the end of the year the production line will come into regular operation, and the last pieces of assigning responsibility and exception handling will fall into place.

In spite of the challenges we have met on the way, we have had a considerable production during our test year. Close to 26,000 books with an average of more than 200 pages per book have now been digitized, and most of these have also undergone OCR and structural analysis. A little more than 1,500 books are freely accessible in their entirety in our digital library, where they are also searchable in full text.

The challenge we now face is to establish production lines for all the types of material that are to be digitized, so that we truly will be able to establish the multimedia digital national library.

# 5.3 Netherlands

#### Digitisation at the Koninklijke Bibliotheek

#### Author: Astrid Verheusen

Some twelve years ago only highlights of collections were digitised and they were mainly used to build attractive expositions on the Internet. The digitisation projects were small and they were focused on images of visually attractive source materials. There was a lot of emphasis on techniques: how to create a good image and how to make images available and searchable on the web. Libraries started to think about new possibilities and they also began to co-operate in digitisation projects on a small scale. In those early years the Koninklijke Bibliotheek build several web expositions with highlights of the special collections, like maps, manuscripts and old sheet music<sup>1</sup>.

At the turning of the century there was a shift in emphasis in digitisation activities. The Koninklijke Bibliotheek moved from digitising highlights to digitising complete collections. Digitisation projects became bigger and due to that project management became more profound. Developments in methods and techniques became more stable and there was a growing awareness about the problem of digital preservation of the digital files. Instead of visual attractive materials the Koninklijke Bibliotheek started digitising text materials and audio and video collections. New possibilities to use the digitised collections were discovered such as applications for specific target groups like scientists and education.

In this period the Memory of the Netherlands was developed. The Memory of the Netherlands is the national digitisation program of the Koninklijke Bibliotheek. In the program digitisation activities of about 50 cultural heritage institutions in the Netherlands are coordinated. The collections have been made available on the website which now contains about 350.000 objects. Most of these objects are images, there are hardly any text sources available on the website<sup>2</sup>.

In the present strategy plan of the Koninklijke Bibliotheek the "Development of a national program for the mass digitisation of sources for research in the humanities" is foreseen. The program should facilitate new research by scientists but it also aims at the public at large. Besides digitisation, the Koninklijke Bibliotheek is working on the development of services and standards and has a special department involved in research and development in the area of digital preservation.

There is a big turn around in the type of material that is being digitised. The library almost completely turned to text digitisation instead of making images of visual materials. This causes new challenges in the development of techniques for navigation, OCR, digitisation from microfilm and search and retrieval. Microfilm as a medium for conservation purposes has been abandoned. Instead guidelines have been developed for preservation imaging for our national preservation program Metamorfoze.

However, the main difference between current projects and those executed ten years ago lies in the size of the projects. Table 1 contains an overview of current digitisation projects at the Koninklijke Bibliotheek. Although the figures look very impressive these projects are not really about mass digitisation. Only a selection of materials are being digitized and special collections like old books before 1800 and old newspapers are included. But the Koninklijke Bibliotheek is certainly digitising on a very large scale and ideas for even more projects are underway. In the next four to five years, the Koninklijke Bibliotheek is going to digitise at least 40 million pages while planning even bigger projects. However, there are some problems getting there.

<sup>&</sup>lt;sup>1</sup> See for example 'Hundred hightlights', http://www.kb.nl/galerie/100hoogtepunten/index-en.html.

<sup>&</sup>lt;sup>2</sup> See http://www.memoryofthenetherlands.nl.

Project	Number of pages
Dutch parliamentary papers 1814-1995	2.300.000
Dutch daily newspapers 1618-1995	8.000.000
Special collections - books before 1800	1.300.000
Radio news bulletins	1.500.000
Metamorfoze - preservation imaging	28.000.0001
Books from Aceh	200.000
Memory of the Netherlands	350.000
Total	41.650.000

# Table 1. Digitisation projects at the Koninklijke Bibliotheek, 2007-2011

# Obstacles for mass digitisation

First of all there are the costs of digitisation. In the projects of the Koninklijke Bibliotheek the average costs for digitising one page amount to  $\notin$  1,30. This sum includes everything: digitisation, the adding of metadata, OCR, person months for selection, preparation of materials and quality assurance, overhead costs and hard and software. Even though the costs for digitisation have declined in the last ten years, digitisation is still rather expensive. At the same time exploitation costs for the maintenance of websites and storage after a project has ended are increasing fast. Especially the costs for long term preservation storage of master files are rising as the amount of files to be preserved increases with millions a year. It is therefore necessary to reduce costs for digitisation to be able to digitise more pages with the available budget.

The technical infrastructure also causes problems. For the storage of the digital files of 40 million digitised pages a capacity of more than a Petabyte is needed. It is not easy to get everything in place for this purpose. From this year on about 2 million files a month have to be processed. These files need to go through a quality assurance system and subsequently be indexed and stored in different storage systems. In the end they should be made available on the web. At the moment the infrastructure of the Koninklijke Bibliotheek is not capable of dealing with such a large amount of files. Furthermore, the search and retrieval system is not ready to give the end user what he needs to find. All in all, the digitisation process is very slow, digitisation should move faster to be able to digitise on a large scale.

At the Koninklijke Bibliotheek digitisation activities have been evaluated in 2007. To do so all steps in the workflow for digitisation had to be taken into account. Since then things slowly began to change. This change is still in its first phase but for many aspects of the workflow it is already possible to see the effect of new approaches. Below, for some of the most important aspects of the workflow a comparison is given between old and new approaches.

# Selection and preparation of materials

The first step in each digitisation project usually consists of selection and subsequently the preparation of the source material for digitisation. At the Koninklijke Bibliotheek much effort is spend on both of these activities. For several digitisation projects advisory commissions have been created consisting of specialists regarding the subject of the digitisation project to advise on the selection of materials. In most of the digitisation projects much time is also

<sup>&</sup>lt;sup>1</sup> This figure for the number of objects to be digitised in Metamorfoze are based on a rough estimation.

spend on completing collections. For the digitisation of the parliamentary papers, which contain about 2.3 million pages, 20 different sets were acquired to build up one complete set for microfilming, digitisation and conservation purposes. Project members turned all pages one by one looking for missing pages that were completed with pages from other sets. A large team worked on this for several years and it really slowed down the project.

Not only are collections completed, pages that are in a bad shape are being replaced or repaired if possible. Due to the high quality standards for selection and preparation of materials selection and preparation take up about 10 to 15 percent of the available budget of digitisation projects.

To lower the costs and increase the pace of digitisation it is necessary to limit the time spend on selection and preparation of materials. Because of limited budgets selection is necessary but it is more a matter of priority - what to digitise first - considering that in the end all important collections will be digitized anyway. At least one should try to digitise complete collections and not to select objects within a collection.

The time spend on preparation of materials can also be limited. This will have consequences for the final quality of the digital files but a comparative assessment has to be made to balance between reasonable quality and perfect solutions. In the newspaper project of the Koninklijke Bibliotheek a decision was therefore made to put limited effort in completion and restoration of the source materials.

# Digitisation

When the Koninklijke Bibliotheek started its digitisation activities a lot of effort was put in creating the perfect image. At the same time scanning had to be done in such a way that the original was not damaged which really slowed down the digitisation process for the more vulnerable library materials. Like most libraries did always two or three different copies of one image were delivered; a master file for long term preservation and reuse in the future and access copies for presentation on the internet. For master files mostly the TIFF format was used.

At the beginning the Koninklijke Bibliotheek started to digitise in house, bought own scanners, but it was soon discovered that scanning is not the core business of the library. From then on digitisation was outsourced. Because of the high standards, only a few companies in the Netherlands could meet the demands. The high standards led to very high quality images, but also to very high costs for scanning and storage of the large master files.

Because of the growing scale of digitisation projects some of these old basic assumptions have to be reconsidered. There is now a preference for digitising from microfilm which mostly delivers less quality images but is cheaper than digitising from originals. Besides, research is being done into alternative file formats for TIFF like jPeg2000 which decreases the amount of necessary storage capacity. For the same reason the use of only one format for both access and preservation is being considered. While outsourcing all scanning activities, expertise about imaging techniques is still being held within in library.

It is necessary that companies can meet the desired standards because of the capacity needed. This means quality standards for scanning have to be lowered to a level that companies can handle.

#### Quality assurance

Because of the high standards for digitisation quality assurance of the digital files automatically was on a high level too. From the beginning quality control managers checked on everything that was digitised. In the first small scale projects sometimes every file was checked and had to be scanned again if quality was not perfect.

For the first large text digitisation project, the parliamentary papers, a quality procedure was set up to check on everything: the quality of the images and the OCR, the

integrity of the XML files and the way in which al different files were connected to each other. For this purpose a dedicated tool was developed for the project. The costs for development of the tool amounted to more than  $\notin$  200.000 and about  $\notin$  80.000 per year for maintenance. In this project quality assurance takes up about 10 percent of the available budget.

Such a high level of quality assurance is not feasible for all large scale digitisation projects. It is not realistic to check the quality off all files and even if we only check a sample, it has to be done as automatically as possible. The Koninklijke Bibliotheek is therefore developing tools to assist in automatic quality assurance.

For a while most effort was spend on quality assurance of OCR. For every batch of 50.000 scanned pages a sample of one percent was taken en checked character for character by project members. Unfortunately, when quality standards for OCR were not met, there was little a supplier could do because at this moment the quality of OCR software for historical documents is just not good enough and manual correction of the text is not feasible. For these reasons quality assurance of OCR is abandoned for the time being.

#### Storage and long term preservation

At the start of digitisation activities at the Koninklijke Bibliotheek the master files were stores on CD Roms and DVD's. Thousands of them were loaded into the presentation system and the CD's were saved as a backup. By now everyone will agree that this kind of storage is not a good idea. Examples of corrupt CD's are well known.

A few years ago it was decided to store the master files of digitisation projects in the e-Depot, the digital archiving environment of the Koninklijke Bibliotheek that ensures longterm access to the national electronic deposit collection. However, to store the master files of all digitisation projects about 1 Petabyte of storage capacity is needed and storage in the e-Depot is very expensive. Until recently there was a policy to store all master files of all digitisation projects in the e-Depot. Because of the high costs the storage strategy had to be changed so to balance between costs, access and preservation. At the same time research is being done into alternative file formats to minimize the need for storage and the Koninklijke Bibliotheek is considering to use only one file both as master and access file.

#### Costs

In earlier projects nobody really knew what the precise costs for several steps in the digitisation process amounted to. It was therefore hard to make detailed budgets for new projects. Now, all costs are made visible, including all kinds of overhead costs. By now it is clear which aspects of the workflow absorb most of the budget which makes it easier to handle those aspects more efficiently. At the Koninklijke Bibliotheek digitisation takes up only 50 percent of the available budget. A few years ago this percentage was even lower: only 30 to 40 percent was available for digitisation while a high percentage of the budget was needed for analyses of the material and quality assurance. It is desirable that the available percentage for digitisation rises, up to 70 to 80 percent of the budget. As mentioned earlier exploitation costs are becoming 'dramatic' and this needs to be solved. At the Koninklijke Bibliotheek costs are shared between different projects by joining similar activities. And - although no funds are received from commercial partners - new business models are explored, for example, digitisation on demand.

#### Organisation and project management

At the Koninklijke Bibliotheek all digitisation activities take place in the R&D department while other departments of the library are not really involved. Because digitisation activities gradually become a regular task of the library it is important to integrate these activities in the other departments of the library.

All digitisation activities are project based. Now that the scale of digitisation projects is growing, each project has become a department on its own, with different project managers for the same activities like selection and digitisation. At the moment 7 very large project
teams are in place and work has to be re-arranged to prevent each project re-inventing the wheel.

#### 5.4 Austria

#### Digital Library Strategy (2008)

#### Author: Bettina Kann

#### Introduction

The purpose of this document is to provide an overview of the main strategies of the Austrian National Library in the digital library environment. It describes current activities and strategies for the next years concerning digital collection building, digitisation, long-term preservation and access to digital resources. In addition organisational and legal aspects, research projects and the European Digital Library initiative are taken into account.

#### Mission and objectives

As part of its Digital Library Strategy the Austrian National Library aims to

Develop and manage digital collections comprising digitised and born-digital resources.

Follow international standards for digital collection development and management. Provide access to these digital collections for scholars and the general public via appropriate user interfaces, considering user needs and taking into account legal requirements.

Preserve these digital collections for the long term, considering them to be a significant part of the Austrian cultural heritage.

Promote the library's digital services in order to foster the Austrian information society, to provide democratic and equal access to information, and to encourage lifelong learning.

Conduct research and development in a national and international context in order to permanently enhance digitisation workflows, optimise the digital preservation strategy and to improve digital services for scholars and the general public

#### Collection Development

#### Digitisation strategy

The Austrian National Library has a strategic plan to digitise its valuable collections of printed publications, as well as other valuable resources like manuscripts, photographs, incunabula, and papyri.

The highest priority is given to publications and objects

which are published or created in Austria;

which contents focus on Austria, the Austrian people or Austrian cultural, political, social, economical past or present;

which are part of historical focal points of the library's collections;

which are in great demand by users, like newspapers, law gazettes etc.;

which contents are at high risk due to deterioration of physical carriers, like sound carriers etc.

The Austrian National Library's digitisation programme is currently based on two pillars:

- 1. Mass digitisation,
- 2. Digitisation of special collections.

A third focus will be the development of digitisation-on-demand services. Following user demands the Austrian National Library will replace analogue reproduction services like production of microfilms by digitisation services in 2008. Also in 2008 the equipment of the Digital Services Department will be complemented by new special scanners for fragile and valuable materials. This has the advantage that valuable objects like manuscripts can be digitised in-house.

In the medium term the Austrian National Library plans to further expand its digitisation capacities by establishing a new digitisation centre in connection with the new storage building. At the moment planning is under development and negotiations concerning funding are still ongoing.

#### Mass digitisation

In 2003 the Austrian National Library started its first text mass digitisation initiative with a focus on the digitisation of Austrian historical newspapers from 1700 until app. 1938. Currently more than 4,5 Million newspaper pages are available online, with an increase of approx. one million pages per year. The Austrian National Library's internet portal for digitised newspapers, ANNO (Austrian Newspapers Online, <u>http://anno.onb.ac.at/</u>), is accessed by app. 800 users per day.

The Austrian National Library's second text mass digitisation initiative ALEX concentrates on historical and current laws, parliamentary material and law gazettes from Austria and the Habsburg monarchy. Currently about one million pages are online and more than 300 readers access them daily via the ALEX portal (<u>http://alex.onb.ac.at/</u>).

Currently access to these materials is in image mode only. It is the objective of the library to make all the underlying texts of the documents (full text) searchable in the future. Due to the high costs of full-text conversion and the currently insufficient results of OCR of historical printed material this will be a step-by-step process. Currently the Austrian National Library is participating in a research project co-funded by the European Commission which aims to reduce costs and at the same time improve results of text recognition of historical printed material.

The third current mass digitisation project of the library focuses on the digitisation of photographic collections and historical portraits. 100.000 photographs and 186.000 portraits of the Austrian National Library's picture archive are now digitally available in a special image database (http://www.bildarchiv.at/). Additional objective of this project is the scientific documentation of the portraits.

Scanning for all three mass digitisation projects is carried out by an external service provider. Pages are scanned with a resolution of 300 dpi in most cases (black/white, greyscale or colour) and are stored as TIFF files with lossless compression (CCITT-4 or LZW). Conversion in internet access formats like PNG or JPEG is done on the fly. The application that provides end user access has been developed in-house.

#### Digitisation of special collections

#### Digitisation of audio material

22.000 analogue sound documents carrying about 30.000 hours of recordings have to be protected against deterioration. The largest part of this endangered collection consists of tapes and music cassettes which contain recordings of in-house concerts, dialogues and speeches of writers as well as field recordings made by folk music researchers. Additionally the collection contains a small number of cylinders, self-cut-discs, as well as shellacs and vinyl-discs. Many of these recordings are unique copies of sound happenings with high rated value. They are classified as important examples of Austria's cultural heritage.

Since 2007 the Austrian National Library has been carrying out a 10-year programme to digitise the most endangered audio documents. The objective of this project is the long-term preservation of audio content. Due to the <u>deterioration</u> of the analogue carriers and the obsolescence of the analogue playback devices preservation can only be achieved by digitisation. The digitisation programme has been jointly planned by the Phonogrammarchiv of the Austrian Academy of Sciences, the Austrian Media Centre (Österreichische Mediathek) and the Austrian National Library. The cylinders have been digitised by the Phonogrammarchiv, the tapes have been digitised by the Austrian Media Centre. The Austrian National Library's digital format for long-term preservation of digitised audio material is Broadcast Wave (wav) 24-bit/96kHz without compression. For access and plavback via internet the audio material is provided as MPEG-1 Audio Layer 3 (MP3) 128kBit/s, 48kH compressed files, which are generated from the Broadcast Wave files. The digitised objects are delivered with descriptive, technical and history metadata, that describe the steps of the digitisation process. All files and metadata are stored in the Austrian National Library's Digital Repository, whereas descriptive metadata are additionally stored in the library's integrated library system (ALEPH catalogue). Resource discovery is provided via the Aleph OPAC and via the Digital Repository, whereas due to legal restrictions access to the MP3 files and the Broadcast Wave files is limited to one concurrent user on-site (at the Austrian National Library).

#### Other collections

In addition the Austrian National Library is digitising holdings of the library's special collections like incunabula, papyri, posters, broadsheets etc.

143 incunabula bibles (170.000 pages) of the library's Collection of Old and Rare Books will be digitised and presented via a special interface (books of the bible). The digitisation process is handled by an external service provider who delivers the digitised content structured according to the METS standard (Metadata Encoding and Transmission Standard). METS files are loaded into the library's Digital Repository. Additionally descriptive metadata is kept in the library's integrated library system (ALEPH-500).

Until 2010 8.000 already published papyri of the library's Papyri Collection will be digitised and made available via the library's integrated library system (ALEPH OPAC).

Smaller digitisation projects of the Austrian National Library include the digitisation of single pages of books in Hebrew from the Collection of Old and Rare Books, the digitisation of a collection of broadsheets from the revolution of 1848, the digitisation of books about the women's movement, as well as the digitisation of first editions of publications by Austrian writers.

#### Catalogue enrichment

The IVSCAN-Project is part of catalogue enrichment for bibliographic details as well as for contents. The project's aim is to make hidden bibliographic information accessible, such as author, title, language and page-number of contributions to collections and compilations of different kinds (conference proceedings, essay collections etc.) as well as other texts not published independently (except articles in journals and year-books). Therefore the table of contents of books containing individual contributions is scanned. The relevant information for bibliographic categories is then transformed into descriptive metadata and can be searched within the Aleph catalogue/OPAC (also allowing keyword search). The scans are accessible as pdf-documents. Beyond permanent acquisitions also literature since 1992 and (according to thematic selection) of earlier years (since around 1970 for the moment) is covered by the project.

#### Figures:

Within the next five years the Austrian National Library plans to digitise:

Newspapers:	3,5 Million pages
Books:	2 Million pages
Images:	300.000 pages
Papyri:	10.000 pieces
Audio files:	4.000 hours

#### Acquisition of born-digital material

#### Legal deposit and selection criteria

The current Austrian Media Law (BGBl. II, 65/2001, i.d.F. BGBl. I Nr. 49/2005) regulates the legal deposit of printed material and off-line media (such as CDROMs, DVDs etc.) but there is currently no such regulation for on-line publications in Austria. The Austrian National Library has encouraged an amendment of the Austrian Media Law to include provisions for the deposit of online publications. It is to be expected that this amendment of the Austrian Media Law will be enacted in 2008. With the revision of the Media Law the Austrian government will comply to a request of the European Commission's recommendation of 24 August 2006 "on the digitisation and online accessibility of cultural material and digital preservation" in the context of the i2010 digital libraries initiative.

The revision of the Media Law will provide the Austrian National Library with a legal framework for domain harvesting as well as for the selective collection of online publications.

In accordance to the collection policy for printed publications the Austrian National Library has developed a collection policy for online publications that includes the following main selection criteria:

An online publication must be digital only, i.e. no offline or printed manifestation exists.

An online publication has to be an "Austriacum", which means, that either the content producer is located in Austria or/and the server which hosts the publication is located below the top level domain ".at" or/and the content of the online publication is related to Austria.

#### Web archiving

The Austrian National Library is convinced of its responsibility to collect and to preserve Austrian web publications a part of the Austrian digital heritage. The Austrian National Library therefore aims to collect, archive and give access to those websites published as part of the domain ".at" as well as those websites that meet the criteria of "Austriaca" (see above). Those websites will normally be harvested and archived once a year. For special events (e.g. elections) selected websites will be collected in more frequent intervals. In addition to domain and event harvesting there will be a selective harvesting strategy for special websites like newspapers etc. that need to be harvested very frequently.

Concerning metadata it is not planned to add bibliographic metadata to those websites which are collected as part of the domain harvesting.

The Austrian National Library will start web archiving in 2008 with an 18 month pilot phase and which will be transferred in permanent service in 2009.

#### Digital Long-term Preservation

#### Long-term preservation of digital surrogates

In order to qualify a digital surrogate of an analogue object for long-term preservation it has to meet certain criteria which have been established by the Austrian National Library. Whether a digital surrogate has to be preserved for the long-term depends on the physical condition of the analogue original and if it qualifies as an "Austriacum".

If digitisation serves to safeguard information because the analogue original cannot be preserved any longer due to deterioration of the carrier or due to other technical reasons the digital surrogate will replace the original and has therefore to be preserved in the long-term. If on the other hand an analogue original is in a good preservation state (and could therefore be digitised once again without being damaged) and is digitised in order to improve access, the corresponding digital surrogate will not be considered for long-term preservation.

4.2. Long-term preservation of born-digital material

All born-digital objects collected by the Austrian National Library have to be preserved in the long-term. Although the Austrian National Library will accept all digital formats delivered by external producers, we aim to limit the formats we have to preserve in the long-term. At present favourite formats are XML, PDF/A, TIFF (either uncompressed or with lossless compression), JPEG2000 and Broadcast Wave. The Austrian National Library stores information about objects and metadata in a digital repository that is informed by the OAIS standard. Digital objects are stored on a scalable NAS mass storage system whereas metadata are managed in a relational database.

For describing and administrating digital objects the Austrian National Library uses the following metadata schemas:

Descriptive Metadata: Dublin Core qualified Technical Metadata: Audio Technical Metadata (AUDIO-MD), NISO Z39.87 Technical Metadata for Digital Still Images (MIX), TEXT-MD (METS-Extension), Video Technical Metadata (VMD) History Metadata: "Event" entity of the PREMIS metadata schema Preservation metadata: "Object" entity of the PREMIS metadata schema Access rights: local XML schema

The library's backup strategies consist of incremental backups on a second mass storage system, of tape-backups and of redundant storage in a high-security data centre in 300 km distance.

Bit level preservation is currently carried out for all digital objects including offline media (like CDROMs and DVDs). For preservation planning the Austrian National Library concentrates on migration strategies. The library is partner in the EC co-funded research project PLANETS to investigate the best preservation strategy for its collections.

#### Access to digital content

The Austrian National Library provides free and open access to all materials which are digitised as part of the library's mass digitisation initiative. Copyright requires that newspapers and journals have to be older than 70 years to be digitised and made accessible without restrictions. As law gazettes and parliamentary documents are public material no copyright restrictions govern their use.

Access to material that will be collected according to the regulation of the new Media Law will be restricted to one concurrent user on-site (i.e. in the facilities of the Austrian National Library). Individual contracts with several institutions are negotiated to grant wider access to selected online-publications.

Digital off-line and on-line publications are either accessible via the bibliographic records in the OPACs (ALEPH) of the Austrian National Library or via the user interface of the library's digital repository (Digitool), which also allows for full-text search. The Austrian National Library uses the ALEPH library system for managing and providing access to the library's metadata records. In order to improve access to its metadata records the Austrian National Library currently carries out a project to integrate the two main catalogues from 1500-1929 and 1930-1988. This project will be finished in 2009.

#### Future strategic objectives: Search engine technology and Web Portal

#### Search engine technology:

Full-text conversion (OCR) of digitised printed publications requires implementation of a full-text search engine. In 2008 the library will conduct a research on existing full-text search solutions. Future search engine technology implementation will on the one hand focus on search solutions for large amounts of structured or unstructured full text and on the other hand on solutions for structured metadata records. It will be decided if it is feasible to implement one single solution for full-text and metadata search or if two separate solutions will deliver better results. Within the framework of the EC co-funded project TELplus the Austrian National Library is currently investigating existing search engine solutions.

#### Web Portal:

The Austrian National Library will implement a web portal which will provide a single search interface for the digital collections and the library's different OPACs. It will provide users with a single solution for discovery and delivery of both analogue and digital information resources, regardless of format and location and will include faceted browsing technologies as well as Web 2.0 features like social tagging and RSS feeds.

The web archive will be built up as a stand alone archive. The contents will not be indexed together with the other collections because of the too large amount of data. If search technology is developed that is able to manage indexing and search of several hundred Terabyte of data it may be possible to merge the collections at a later stage.

#### Research projects and European Digital Library Initiative

The focus in research and development corresponds with the strategic planning of the Austrian National Library. This ensures that solutions which are developed and implemented

prototypically in research projects can at a later stage be transferred into standard services of the library. Currently research and development focuses on the following areas:

Implementation of OCR for historical documents in the context of mass digitisation Evaluation and prototypical implementation of search engine technologies Development of digital preservation strategies and implementation of a testbed for digital preservation

Development of web services and interfaces for exchange of metadata in digital libraries

Knowledge that is generated within research and development will be fed into the plans of the Austrian National Library in the fields of full-text conversion (OCR), search engine technologies, user interface development and long-term preservation. Also in the future we will only engage in research projects which promise to contribute to the library's strategic aims or demands in development. In the near future the following areas will require consideration in research and development:

Providing access to web archives Long-term preservation of web archives Web services, service oriented architectures, Web 2.0 technologies Text mining technologies, automatic content analysis and classification Search engine technologies Providing access to digital audio material

The Austrian National Library endorses the i2010 Digital Libraries initiative of the European Commission. Since 2005 the library has been a full member of The European Library, the internet gateway of the European national libraries. The European Library will be one of the building blocks for the European Digital Library. The Austrian National Library intends to make all electronic catalogues as well as the digital collections fully interoperable with The European Library and the future European Digital Library. The Austrian National Library regards The European Library as an excellent means to enhance visibility of our analogue holdings and to broaden access to our digital collections.

The library's preferred protocol for metadata interoperability is OAI-PMH (Open Archive Initiative Protocol for Metadata Harvesting). We are currently working on making all library catalogues fully OAI-PMH compatible.

#### Organisational structure

In 2008 the main department "Digital Library" was established which now pools all departments dealing with digital material and digital services that beforehand were split up in several organizational units of the library. The main department "Digital Library" consists of the four departments:

#### **Digital Services**

The digital services department is responsible for the implementation of the digitisation programme of the Austrian National Library, including mass digitisation projects and digital reproduction services on demand. The department is also responsible for planning and implementation of a new digitisation competence centre at the Austrian National Library that will be situated in the new storage building.

#### Long-term Preservation

The long-term preservation department focuses on developing a collection of born-digital resources. These resources are either voluntarily deposited online-material or material deposited under the legal deposit regulations. The department is further responsible for managing the digital repository and for planning preservation strategies like migration scenarios. In addition the implementation of web archiving will be carried out by this department.

#### Research and Development

The research and development department carries out basic research, generates know-how and implements prototypical solutions which can at a later stage be integrated in the library's services. The research and development department is partner in several projects co-funded by the European Commission which carry out basic and applied research in the field of digital libraries.

#### Library Systems

The library systems department is responsible for managing the library systems Aleph 500 and Digitool and for setting up the various OPACs of the Austrian National Library. The department is also responsible for training and supporting staff members in the usage of the library system.

### 6 Recommendations

European national libraries have been digitising parts of their collections for the last ten years.

Many Libraries have created national memory collections and have built up a digital library infrastructure.

The last ten years of digitisation at the European national libraries have created a valuable and enduring resource for scholars and the public alike.

#### However:

If we look at the general outcomes from the survey on digitisation (chapters 4 & 5) it becomes clear, that although National Libraries are well prepared to meet the challenges of digitisation on a strategic level, the production of digital content is still on a very low level. Only 1-2 % of national libraries' collections will be digitally available in the near future. In terms of material types digital collections reflect the structure of analogue collections, with one basic and most critical exception: book digitisation and digitisation of printed research collections is far below expectations for the future European Digital Library. This will not change unless all stakeholders at EU, national and institutional level change their policies. Due to successful private projects and initiatives, access to digital books and printed collections are what users regard as the essential basis of a digital library. Currently a large part of digitised printed heritage is accessible in image mode only. Although the issue of full-text conversion (OCR) of digitised historical printed material is now being tackled by the European projects TELplus (eContent*plus*) and IMPACT (FP7) much more needs to be done to create a critical mass of full text for the European Digital Library.

Due to this situation there is a high risk that users will not understand the concept of the European Digital Library based on a digital library infrastructure which guarantees sustainability, long term preservation, interoperability and access to multilingual collections, but without reaching a critical mass of digital content.

Given the fact that mass digitisation of printed materials (with 100.000 volumes per year) (case studies of France and Norway) is not only still very cost- intensive but also timeconsuming and takes at least a preparation period of half a year it becomes clear that for the time being only very few libraries (Norway, France, UK, Netherlands) will be capable of doing large scale digitisation. On the other hand the example of Norway indicates that for the first time complete digitisation plans for national libraries can be based on a realistic background and that cost reduction in storage technologies and mass digitisation make it for the first time imaginable to "duplicate" the total collections of libraries in electronic format, to preserve them and to give virtual access to the complete content.

#### Recommendations:

The first results of EDL Report on Digitisation were presented and discussed at the CENL annual meeting in Helsinki in October 2007. The members of CENL identified the following as the most critical issues: large scale digitisation of books and printed material; understanding of user needs; innovation and technological change; smarter marketing of the European Digital Library. A need for co-ordinated digitisation of content and reporting on digitisation on a regular basis (including statistical data) may be added to this list as a result of this EDL Report on Digitisation.

The following recommendations emerge from this discussion and from the conclusions of this EDL Report on Digitisation:

#### Large scale digitisation of books and printed material

Recommendation: Given the massive size of the digitisation challenge which they face and the resource issues alluded to by many of them, national libraries should consider carefully whether and how public private sector partnerships might contribute to solving the problem.

- all efforts have to be undertaken to enlarge the scale of book digitisation and to give access to printed material, including 20<sup>th</sup> century documents
- all stakeholders should be involved, including appropriate public-private partnerships
- special attention should be paid to the work on orphan works and out of print works being undertaken by the High Level Expert Group

#### Understanding of user needs

Material held in national libraries has been traditionally organised in a ways that help the trained librarian find information on the shelves. Content needs to be organised for a web audience. This means granularity of data to allow full text searching, the use of protocols allowing data to be crawled and harvested and the organisation of data to ensure maximum discovery.

For EDL to be a true success all the domains involved need to produce digitised, searchable, interoperable data

Recommendation:

• The best way to meet user needs is to give direct access to digital content. CENL should support The European Library in developing a technological platform with central harvesting to create searchable and interoperable data

#### Innovation and technological change

Recommendation:

• Considerable attention should be paid to the OCRing of historical material, metadata extraction, content analysis tools and navigation applications

#### *Co-ordinated digitisation of content*

- National libraries that have not yet done so should build up at institutional level a digital library infrastructure based on approved strategies and development plans
- National digitisation strategies should be based on a territorial approach and should focus on building large digital collections
- Partnerships with other organisations should be formed to create larger virtual collections.

#### Report on digitisation an a regular basis (including statistical data)

Recommendation: CENL should continue to maintain close links with the Numeric study team and the CENL Content Working Group should continue to support the Commission's efforts to develop standardised reporting in the field of digitisation and raise the profile of the related issues with governments and other stakeholders.

• National Libraries should report on an annual basis their progress in digitisation in a way compatible with emerging European standards for such reports. (Numeric)

Improve awareness of The European Library and its role in building Europeana

There is not enough public awareness of The European Library. More effort is required to market this resource in each participating country. The crucial role that the national libraries are playing in the creation of Europeana should also be brought to public attention.

• CENL members should support the TEL-office in the marketing of The European Library as a resource and together with the European Commission make content providers from other domains aware of the importance of contributing material to Europeana

#### Annex

The following part gives an overview on statistical data collected from EDL questionnaires on programmes and strategies and on individual projects in European national libraries.

In most libraries statistical data on collections are derived from bibliographic entities. When asking for digital objects EDL questionnaire was asking for the number of physical items digitised and for the number of pages digitised for printed material. Numbers are based on figures reported by national libraries. Digital pages include books, serials, other printed materials and newspaper pages.

The following basic material types have been asked for and counted separate: books, newspapers, (medieval) manuscripts, maps, engravings, photographs, posters, drawings, audio, video, dissertations (theses). Statistical data on archival materials have been categorized as "other materials". Archival materials comprise e.g. industrial price lists (Finland), obituary notifications, leaflets (Hungary), proceedings of parliament (Netherlands) and other archival documents hold in national libraries.

If not indicated otherwise by national libraries the following average numbers have been used for completing statistical tables

1 book = 300 pages 1 medieval manuscript = 300 Seiten 1 dissertation = 200 pages 1 newspaper issue = 15 pages

All data have been reported to numeric project (<u>www.numeric.ws</u>), which aims at collecting statistical data on digital objects on a regular basis. The following definitions are given by the numeric team for collecting statistical data on digitisation of analogue material:

#### digital objects

is a general term used for the body of digital material that is subject to digital preservation, both digital publications and digital records; both online and offline; both born-digital and digitised.

Digital objects in EDL questionnaire are restricted to digitised analogue materials, excluding born digital material and digital publications.

#### digitisation

The process of creating digital files by scanning or otherwise converting analogue materials. The resulting digital copy, or digital surrogate, would then be classed as digital material and then subject to the same broad challenges involved in preserving access to it, as "born digital" materials.

#### retrodigitisation

See also "digitisation". Retrodigitisation or retrospective digitisation means the digitisation of analogue publications (e.g. print media, films, tapes). Retrodigitisation is above all used in libraries and archives for the purpose of preserving documents in danger of decay and of giving access to broad usage.

#### archival unit (ISO 5127)

single document or set of documents in an archive, treated as an entity.

**article** (ISO 5127) independent text forming a part of a publication.

#### audiovisual document (ISO 2789)

document in which sound and/or pictures are prominent, and which requires the use of special equipment to be seen and/or heard.

NOTE 1 Adapted from ISO 5127:2001.

NOTE 2 This includes audio documents such as records, tapes, cassettes, audio compact discs, DVD's, files of digital audio recordings; visual documents such as slides, transparencies, and combined audiovisual documents, such as motion pictures, video recordings, etc. Microforms are excluded.

**book** (ISO 2789) non-serial printed document in codex form.

#### cartographic document (ISO 2789)

conventional representation, on a reduced scale, of concrete or abstract phenomena which can be localized in space and time.

NOTE 1 Adapted from map in ISO 5127:2001.

NOTE 2 This includes documents such as two- and three-dimensional maps, globes, plans, topographic models, tactile maps and aerial representations, but excludes atlases and any other cartographic documents in codex, micro, audiovisual and electronic form.

collection (ISO 5127)

gathering of documents assembled on the basis of some common characteristic, without regard to their provenance.

#### drawing (ISO 5127)

picture made with a solid mineral substance or a pointed tool.

#### engraving (ISO 5127)

print made from any kind of intaglio plate, whether engraved with hand-tools or a machine, or etched with acid, so that the printing areas are lower than the non-printing areas.

#### government document (ISO 2789)

document published at government expense or as required by law or by an international agency (e.g. United Nations, European Union and UNESCO).

#### graphic document (ISO 2789)

print document in which pictorial representation is the most prominent feature.

NOTE This is pictorial rather than linguistic, musical or cartographic in form. It includes art prints, art originals, art reproductions, photographs, posters, study prints, technical drawings, etc., but excludes graphic items in codex form or in microform, audiovisual and electronic form.

#### incunabulum (ISO 5127)

volume printed in Europe from movable type and dating from before 1501-01-01.

**leaf** (ISO 5127)

sheet of paper or similar thin material on which information has been or may be recorded.

manuscript (ISO 2789)

original document that is handwritten or in typescript.

NOTE Bound volumes and other units (fragments, rolls, autographs, etc.) may be counted separately.

**map** (ISO 5127) conventional representation, on a reduced scale and usually flat, of phenomena which can be localized in space and time.

**microform** (ISO 2789) photographic document requiring magnification when used.

NOTE 1 Microfiche and microfilm are included.

#### monograph (ISO 5127)

publication in print or non-print form, either complete in one volume or complete, or intended to be completed, in a finite number of volumes.

#### newspaper (ISO 2789)

serial, which contains news on current events of special or general interest, the individual parts of which are listed chronologically or numerically and usually appear at least once a week.

**page** (ISO 5127) one side of a leaf.

**pamphlet** (ISO 5127) monograph having not more than 48 pages.

#### patent (ISO 2789)

government document granting an inventor the sole right to use or license an invention together with associated documentation.

#### periodical (ISO 2789)

serial under the same title published at regular or irregular intervals, over an indefinite period, individual issues in the series being numbered consecutively or each issue being dated NOTE Series of reports, transactions of institutions, series of regular conference proceedings and annuals are included, while newspapers and monographic series are excluded.

#### personal papers (ODLIS)

In archives, the private documents and related materials accumulated by an individual in the course of a lifetime (letters, diaries, journals, legal documents, etc.). In contrast to official papers, which may be subject to the disposition of an employer or government, personal papers are subject to the owner's disposition.

**picture** (ISO 5127) primarily two-dimensional presentation of one or more objects or shapes.

**piece** (ISO 5127) smallest physically indivisible archival unit.

postcard (ISO 5127)

card for conveyance by post, often with a picture on one side.

poster(1) (ISO 5127)

singlefaced print for display on a flat surface.

#### poster(2) (ODLIS)

A large single sheet of heavy paper or cardboard, usually printed on one side only, with or without illustration, to advertise a product/service or publicize a forthcoming event (meeting, concert, dramatic performance, etc.), intended for display on a bulletin board, kiosk, wall, or other suitable surface.

#### printed music document (ODLIS)

document, the essential content of which is a representation of music, normally by means of notes

NOTE May be in sheet or codex form.

#### records (ODLIS)

Documents in any form, created or received by an agency or person, accumulated in the normal conduct of business or affairs, and retained as evidence of such activity, permanently or for a limited period of time, usually arranged according to a discernible system of recordkeeping.

#### serial (ISO 2789)

document in print or in non-print form, issued in successive parts, usually having numerical or chronological designations, and intended to be continued indefinitely, whatever its periodicity.

#### NOTE 1 Adapted from ISO 5127:2001.

NOTE 2 Monographic series are excluded and should be counted as books.

#### set (ISO 5127)

assembly of objects or concepts considered as a wholesheet see leaf.

#### sheet music (ISO 5127)

printed music issued without covers, whether actually printed on single sheets or not.

**volume** (ISO 2789) physical unit for a printed document assembling a certain number of leaves under one cover to form a whole or part of a set.

#### 6.1 Statistics

2006	Austria	Belgium	Croatia	Cyprus	Czech Rep.	Denmark	Estonia
holdings physical (items)	7.719.094	12.000.000	2.526.000	84.000	6.281.000	30.277.871	3.310.000
obects digital 2006	440.000	30.000	8.781	0	15.000	187.752	4.700
pages digital 2006	4.500.000	975.500	59.698	0	3.200.000	800.000	620.000
book(object)	2.600	19.000	79		5.200	7.000	700
book(page)	800.000	950.000	15.720		2.500.000	730.000	80.000
manu(object)		152	47		1.800	350	
manu(page)		25.500	2.328		600.000	70.893	
newsp(title)	80		3				94
newsp(issues)	250.000		2.460				2.760
newsp(page)	3.700.000		37.722				533.920
serials							
fotos	60.000	3.010	1.170				
engravings	120.000	3.805	300			4.433	
maps		1.934	406		300	142.335	38
posters			167				1.183
audio	508						
drawings			600			1.635	
video							
other	1.500	2.503	1.285				
diss./theses							
budget	100.000				350.000	90.000	215.000
2012	Austria	Belgium	Croatia	Cyprus	Czech Rep.	Denmark	Estonia
holdings	7.719.094	12.000.000	2.526.000	84.000	6.281.000	30.277.871	3.310.000
physical (items)							
digital	620.000	150.000	200.000	0	28.000	240.000	31.000
pages planned digital	10.000.000	4.075.000	4.000.000	0	11.000.000		2.000.000
book(object)	8.000	21.300	11.179		25.000	8.600	3.700
book(page)	2.400.000	1.020.000	2.170.720		7.500.000	780.000	1.110.000
manu(object)		275	1.047		500	475	
manu(page)		55.000	302.328		300.000	105.000	
newsp(title)	160	30	296				100
newsp(issues)	200.000	100.000	100.296				27.000
newsp(page)	7.500.000	3.000.000	1.563.000				400.000
serials							
fotos	200.000	3.010	1.170				
engravings	200.000		300				
maps		2.300	5.000		300	167.500	350
posters							
audio	4.000						
drawings							
video							
other	8.000	37.000	72.740				
diss./theses	0.000	5					

#### EDL Report on Digitisation

2006	Finland	France	Germany	Greece	Hungary	Iceland	Italy
holdings physical (items)	6.670.000	30.000.000	23.000.000	1.500.000	8.000.000	980.000	6.274.000
obects digital 2006	188.450	504.000	78	73	347.000	112.000	9.100
pages digital 2006	630.000	36.000.000	128.698	240.406	366.000	1.520.000	170.600
book(object)	550	117.000		68		452	554
book(page)	85.000	35.100.000		20.400		120.000	120.000
manu(object)					610	910	45
manu(page)					366.000	54.600	16.208
newsp(title)			78	5	114	250	1
newsp(issues)	177.900	137.000	7.600	14.000		110.000	8.500
newsp(page)	415.000	685.000	114.000	220.006		1.175.000	50.000
serials							
fotos		232.000			150.000		
engravings		10.000			591		
maps		8.000			3.400	500	
posters					7.000		
audio					18.000		
drawings					1.587		
video					15.000		
other	130.000				150.000		
diss./theses							
budget	600.000	5.000.000				665.000	
2012	Finland	France	Germany	Greece	Hungary	Iceland	Italy
holdings physical (items)	6.670.000	30.000.000	23.000.000	1.500.000	8.000.000	980.000	6.274.000
objects planned digital	400.000	1.500.000	78	73	370.000	223.000	19.100
pages planned digital	5.965.000	141.200.000	128.698	240.406	366.000	3.475.000	220.000
book(object)	2.200	462.000		68			554
book(page)	335.000	138.600.000		20.400		120.000	120.000
manu(object)					610	1.900	45
manu(page)					366.000	1.140.000	16.208
newsp(title)			78	5	114		
newsp(issues)	366.666	400.000	7.600	15.000	435	220.000	12.500
newsp(page)	5.500.000	2.000.000	114.000	220.006	670.000	2.350.000	80.000
serials							
fotos		454.000			160.000		
engravings		10.000			621		
maps		8.000			5.200	500	
posters					9.900		
audio	20.000				30.000		6.000
drawings					1.587		
video							
					30.000		
other	130.000				30.000 150.000		

2006	Latvia	Lithuania	Luxembourg	Macedonia	Malta	Moldova	Netherlands
holdings physical (items)	5.100.000	6.912.266	720.000	3.000.000	500.000	2.510.254	4.000.000
obects digital 2006	11.000	195.000	55.500	58	13		1.100.000
pages digital 2006	303.400	2.800.000	241.300	3.000		6.600	3.000.000
book(object)	1	663	48			1	600
book(page)	495	90.000	10.500			800	80.000
manu(object)	54	3.000		58			
manu(page)	32.400	1.800.000		3.000			5.300
newsp(title)		195	12			1	
newsp(issues)	1.800	60.300	37.500			200	100.000
newsp(page)	271.000	905.000	213.300			3.000	1.500.000
serials		74					
fotos	3.800	1.221					216.245
engravings	217	7.024			10		5.000
maps	5.000				3		48.295
posters	1.207						5.150
audio	3	1.268					100
drawings		4.970					38.205
video	2						
other		116.245	18.000				632.884
diss./theses							
budget	130.000	530.000	350.000				23.000.000
2012	Latvia	Lithuania	Luxembourg	Macedonia	MALTA	Moldova	Netherlands
holdings physical (items)	5.100.000	6.912.266	720.000	3.000.000	500.000	2.510.254	4.000.000
objects planned digital	165.000	1.200.000	68.800	400		30.000	3.100.000
pages planned digital	2.100.000	18.000.000	408.800	20.000		300.000	15.000.000
book(object)	4.000	5.000				900	5.200
book(page)	1.200.000	1.500.000	10.500			270.000	1.560.000
manu(object)		3.000		400		48	
manu(page)	32.400	1.800.000		20.000		28.000	5.300
newsp(title)		3.000					250
newsp(issues)	61.300	970.000	50.800			200	530.000
newsp(page)	920.000	14.563.000	390.800			3.000	8.000.000
serials							
fotos	5.190	5.000				300	216.245
engravings	217	13.000			10		5.000
maps	6.787	5.976			3	40	48.295
posters	7.000	2.527					
audio	70					26.000	
	70						
drawings	70						
drawings video	80.000						
drawings video other	80.000	189.236	18.000				2.300.000

#### EDL Report on Digitisation

2006	Norway	Poland	Russia P	Russia M	San Marino	Slovak Rep.	Switzerland
holdings physical (items)	16.000.000	6.973.614	34.763.500	43.400.000	200.000	5.130.000	6.281.000
obects digital 2006	470.000	5.100	198.888	162.000		6.000	50.000
pages digital 2006	1.000.000	1.050.000	648.186	33.300.000		30.000	0
book(object)	20.200	3.500	431				
book(page)	4.510.000	1.050.000					
manu(object)	2		975	15		500	
manu(page)	1.000		4.765	9.000			
newsp(title)			6.000				
newsp(issues)	200.000		16.576				
newsp(page)	1.000.000		126.208				
serials				30			
totos	320.000		1.163			2.000	3.682
engravings		150	500				
maps		50	6.486	110			440
posters	5.000		050	0.40			45.000
audio	150.000		250	310		60	1.525
drawings	570			184		2.000	4.40
VIDEO	570	4 000	0.450	4 747		0.000	148
otner	10.000	1.200	8.156	1./1/		2.000	
diss./theses	1 000 000	200,000	180.000	160.000		100.000	400.000
Dudget	1.800.000	200.000	85.714	145.000	Con Marina	120.000	133.000
2012 boldings	Norway	Poland	Russia P	Russian M	San Manno	Slovak Rep.	Switzenand
physical (items)	16.000.000	6.973.614	34.763.500	43.400.000	200.000	5.130.000	6.281.000
objects planned digital	3.000.000	160.000		800.000	15.000	1.700.000	260.000
pages planned digital	74.000.000	13.560.000		160.000.000		145.000.000	600.000
book(object)	265.000	43.200				350.000	
book(page)	68.500.000	12.960.000				120.000.000	
manu(object)	3.300	1.005					
manu(page)	2.000.000	603.000		9.000			
newsp(title)						10.000	
newsp(issues)	250.000					1.300.000	4.000
newsp(page)	2.350.000					20.000.000	600.000
serials						1.500	
fotos	500.000	50.000	1.000			2.000	7.000
engravings		50.060					
maps	70.000	5.050	1.200	110			440
posters	30.000						250.000
audio	500.000			1.500			
drawings		10.310	500				
video	1.250.000						
other	100.000						
diss./theses			180000	800000			

	Siovenia	Spain	Sweden	UK
holdings physical (items)	2.495.140	15.000.000	6.845.361	102.605.184
obects digital 2006	364.170	47.000	500	362.000
pages digital 2006	223,462	500.000	24.000	33.000.000
book(object)	400	100	43	100 000
book(page)	120.000	30.000	6.000	30.000.000
manu(object)	542	36	1	
manu(page)	2.264	21.600	612	
newsp(title)		143		1.350
newsp(issues)		39.740	1.160	200.000
newsp(page)		448.383	17.400	3.000.000
serials	8			
fotos	10.572	411		
engravings	0	5.500	353	
maps	68	1.650	50	
posters	0	456		
audio	2.030			12.000
drawings	65		40	
video	0			
other	29.837			50.000
diss./theses				
budget	1.000.000	2.500.000	160.000	8.000.000
2012	Slovenia	Spain	Sweden	UK
holdings physical (items)	2.730.000	15.000.000	6.845.361	102.605.184
objects planned digital	706.600	1.500.000	13000	603.000
pages planned digital	9.072.500	52.500.000	273000	64.000.000
book(object)	21.600	100.000	412	200.000
book(page)	6.480.000	30.000.000	123.600	60.000.000
manu(object)	3.000	300		
manu(page)	12.000	180.000	612	
newsp(title)		2.200		1.800
newsp(issues)		1.400.000	10.000	267.000
newsp(page)		22.500.000	150.000	4.000.000
serials	300			
fotos	11.600	411		
engravings	0	8.000	1.800	
maps	600	1.650	70	
posters	0		40	00.000
audio	10.000		500	36.000
urawings	300		500	
video	570 500			100 000
ourier	579.500			100.000
newsp(page) serials fotos engravings maps posters audio drawings video	300 11.600 0 600 0 10.000 300 0	22.500.000 411 8.000 1.650	150.000 1.800 70 40 500	4.000.000 36.000



### The main types of objects digitized by 37 European National Libraries















#### 6.2 Questionnaire

## CENL / EDLproject Digitisation Survey

#### PART 1: Introduction, Goals, Background

This survey is carried out by **EDLproject** on behalf of the **CENL Content Working Group**.

Digitisation is attracting a great deal of political attention right now and CENL has a major role to play. To do this effectively we need a strategic overview derived from good information on what our members have done and are doing. We hope to use the information we get to influence events and funding. We hope that all CENL members will help us.

We recognise that there is some work involved in completing the questionnaires and for this reason we plan to keep them on file and use them as the basis for reporting on this topic to CENL in future. Once a collection has been reported in this survey, we hope it may not need to be reported again.

Please give high priority to completing these questionnaires.

#### **CENL** Content Working Group

The CENL Content Working Group is investigating how contents of European national libraries can be selected for large scale digitisation for the European Digital Library. Members of the Content WG are Austria, France, Hungary, Norway, Portugal, Spain and the United Kingdom.

#### **EDLproject**

EDLproject (<u>www.edlproject.eu/</u>) is a project co-funded by the European Commission under the eContent*plus* Programme and coordinated by the German National Library. EDLproject, which started in September 2006, is a direct response to the request of Viviane Reding, Commissioner for Information Society and Media, made at the CENL conference in Luxembourg on 29 September 2005 that national libraries should use their influence in the debate on the digitisation of European content for access through the web.

One of the objectives of EDLproject is to leverage the influence and resources of CENL as a key player and stakeholder in the content field to work towards a consensual resolution of certain issues raised by the Commission's *i2010: Digital Libraries* initiative<sup>1</sup>, such as potential availability of digital content from national libraries.

#### Aims of this questionnaire

The aims of this questionnaire are twofold:

1. It will facilitate the overall goal for the coming years to digitise more content more quickly and to make sure that access to a critical mass of content is as complete as possible. The results of the survey will form the baseline of a **roadmap** to be produced by the CENL Content Working Group and EDLproject to describe the issues in digitisation in Europe and possible

<sup>&</sup>lt;sup>1</sup> <u>http://ec.europa.eu/information\_society/activities/digital\_libraries/index\_en.htm</u>

solutions. A draft of this report will be circulated to CENL members prior to the meeting in Helsinki in September 2007. It will be discussed in the CENL plenary meeting and at the TEL management board.

 At the CENL plenary meeting in St. Petersburg it has been decided to integrate a digitisation report in the CENL annual report. The results of this survey will form the input for the CENL report on digitisation in 2007 and will be used to work out on how to monitor digitisation activities of CENL members on a regular basis from 2008 onwards.

The results of the survey should give an overview on institutional digitisation programmes and policies and provide measurable data for content digitisation: What will be digitised when and how in European National Libraries in the next 5 years.

The Digitisation Roadmap which will be based on the survey will describe the current baseline: what is available in what form in the EU national libraries and the other CENL member libraries against the total holdings of these libraries. It will be used to make stakeholder aware of how enormous the potential is.

First collections of the necessary data have been made in 2005 and 2006 in surveys by the TEL-ME-MOR project and The European Library office. However, these surveys did not cover all CENL member libraries and most of the data needs already to be updated.

To make the Digitisation Roadmap a success we need urgently your help and contribution for collecting information on:

- Your institutional digitisation programmes and strategies for the next 3–5 years
- Digitised collections and ongoing digitisation projects in your library

#### How to complete this survey:

This survey consists of 3 parts:

- PART 1: Introduction, goals, background (this part): *DigitisationQuestionnairePart1.doc*
- PART 2: Programmes and Strategies: *DigitisationQuestionnairePart2\_xxx.doc*
- PART 3: Single Collections and Projects: DigitisationQuestionnairePart3\_xxx\_nn.doc

#### What you need to do:

- 1. Each CENL member is asked to complete **one** questionnaire "PART 2: Programmes and Strategies" for your library.
- 2. Each CENL member is asked to complete **one** questionnaire "PART 3: Single Collections and Projects" **for each** digitised collection and **each** digitisation project. Thus it is very likely that you will have to complete **more than one** version of this questionnaire. We have designed the questionnaire to ensure that completion will be very straightforward, but we recognise that it may require some time effort, depending on the number of questionnaires you will have to complete for your institution.

#### Timing:

Please, return the filled-in questionnaire before 1<sup>st</sup> June 2007 by email to:

#### elisabeth.weigl@onb.ac.at

Elisabeth Weigl Picture Archive Austrian National Library Josefsplatz 1 A 1015 Vienna

If you have any questions regarding this questionnaire, kindly contact:

Dr. Hans Petschar Director, Picture Archive Austrian National Library Josefsplatz 1 A 1015 Vienna Tel.: +43-1-53410-330 Email: <u>hans.petschar@onb.ac.at</u>

Thank you for your co-operation!

Dr. Britta Woldering (German National Library, CENL Secretary) Dr. Hans Petschar (Austrian National Library, Chair CENL Content Working Group)

## **CENL / EDLproject Digitisation Survey**

#### PART 2: Questionnaire for programmes and strategies

Please rename this MS Word File by replacing "xxx" by a Short Name of your institution.

Example: "DigitisationSurveyPart2\_xxx.doc" "DigitisationSurveyPart2\_AustrianNationalLibrary.doc"

I Please complete <u>one</u> Part 3: Questionnaire <u>for each</u> digitised collection and <u>each</u> digitisation project mentioned Question 12 of this Part 2: Questionnaire!!

#### Thank you very much in advance for your cooperation.

#### Please confirm:

□ We agree to the use of the data collected through participation in this survey solely for the purposes outlined in Part 1 of the Questionnaire.

1.	Library Name	
2.	Country	
3.	Name of Person who is completing this questionnaire	
4.	Position of this contact person within library	
5.	E-Mail address of this contact person	
<b>6</b> .	Further contact details (Postal Address, Fax, Skype etc.)	
7.	Number of Full Time Equivalents working for	FTEs
	your institution	Comments:
8.	Annual budget of your institution	Euro
		Comments:
9.	Annual budget for digitisation	Euro
		Comments:
10.	Does your library have one or more digitisation programmes?	□ yes □ no

	if yes continue with question 11 if no continue with question 16
<b>11.</b> What is/are the name(s) of the programme(s)?	
12. Which digital collections or projects are part of the programme(s)? Please list all collections and projects.	<i>I! Complete a Questionnaire Part 3 for <u>each</u> of the collections listed here!!</i>
<b>13.</b> Short description of the programme(s)	
<b>13.b.</b> Can information on these programmes be found on the internet (programme description, publications, press material, etc)? Please provide the URLs.	
<b>14.</b> Time scale: Start year and (planned) end year of each programme. List all programmes.	
<b>15.</b> Does your institution have a written digitisation strategy?	□ yes □ no
	if yes continue with question 16 if no continue with question 17
<b>16.</b> Please give a summary of the main aspects of your institutional digitisation strategy.	
<b>16.b.</b> Is this summary published on the internet? Please provide the URLs.	
<b>17.</b> Are your institutional digitisation programme(s) and your institutional strategy part of a national programme?	<ul> <li>yes</li> <li>no</li> <li>if yes continue with question 18</li> </ul>
	if no continue with question 23
<b>18.</b> Name(s) of the national programme(s)	
<b>19.</b> Who is responsible for the national	

	programmes(s) (e.g. Ministry of Culture)?	
20.	Which institutions are involved in the nationa programme(s)?	I
21.	Could you provide a short description all of these national programme(s)?	
22.	Can additional information on these national programme(s) be found on the internet or WWW? Please provide the URLs.	
23.	Which collections and materials will be digitised and made available in your library over the next 5 years? Please write down ho many objects you are going to digitise. Pleas specify whether these objects are books, newspapers, pages, images etc.	w e
24.	What collections need priority in digitisation over the next 5 years?	
25.	What are your selection criteria for digitisation? If several please list in order of priority.	
26.	In your opinion what should the priorities be in digitisation to build the European Digital Library?	
27.	What, in your opinion, are the main obstacles to building the European Digital Library?	
28.	Additional Remarks	

Please complete <u>one</u> Part 3: Questionnaire <u>for each</u> digitised collection and <u>each</u> digitisation project mentioned Question 12 of this Part 2: Questionnaire!!

Thank you very much in advance for your cooperation.

## **CENL / EDLproject Digitisation Survey**

# PART 3: Questionnaire for each, single digitised collection and/or each, single project

Please rename this Word File by replacing "xxx" by a Short Name of your institution and replace "nn" by a running collection / project number.

Example: "DigitisationSurveyPart3\_xxx\_nn.doc" "DigitisationSurveyPart3\_AustrianNationalLibrary\_01.doc"

# I Please complete <u>one</u> questionnaire <u>for each</u> digitised collection and <u>each</u> digitisation project mentioned in Question 12 of Part 2:Questionnaire!!

Thus it is very likely that you will have to complete **more than one** version of this questionnaire. We have designed the questionnaire to ensure that completion is straightforward.

#### Thank you very much in advance for your cooperation.

#### Please confirm:

□ We agree to the use of the data collected through participation in this survey solely for the purposes outlined in Part 1 of the Questionnaire.

1.	Library Name	
2.	Country	
3.	Name of Person who is completing this questionnaire	
4.	Position of this contact person within library	
5.	E-Mail address of this contact person	
6.	Further contact details (Postal Address, Fax, Skype etc.)	
7.	Name of the individual project or the digitised collection	
8.	Is the collection accessible via WWW? Please provide the URL of the entry page.	
9.	Short description of the project / digitised collection	

<b>10.</b> Can information on this collection be found on the description, publications Please provide the URL	project / digitised ne internet (project s, press material, etc)? s.		
<b>11.</b> Time scale: Start year a of project	nd (planned) end year	Begin: End:	
<b>12.</b> What is the total numbe the collection that is digibooks, newspapers, pho	r of <b>physical</b> objects in itised? (Number of otographs etc)		
<b>13.</b> How many objects have	already been digitised /	will be digitised within the	nis collection / project?
Material Type Number of objects already digitised (e.g. books, newspapers, images)		Current number of single digital files. Give number and description (e.g. 150.000 images of book pages).	Planned number of digitised objects by the end of the project
Books			
Newspapers			
Serials			
Other printed material – please specify:			
Manuscripts			
Autographs			
Other non-printed material – please specify:			
Archival material – please specify:			
Maps			
Photographs			
Engravings			
Drawings			
Posters			
Postcards			
Other image material – please specify:			

Three-dimensional objects				
Fabrics and Textiles				
Sound Recordings				
Film and Video Recordings				
Other – please specify:				
<b>14.</b> How many objects are digitised per year within this collection / project?				
<b>15.</b> Actual size of collection in GB (gigabytes)	GB			
<b>16.</b> Total budget allocated to project	Euro			
<b>17.</b> Can you estimate the average digitisation costs per object?	Euro-Cent per object			
<ol> <li>How many staff members are / have been working on this project? (Full Time Equivalents)</li> </ol>	FTEs			
<b>19.</b> How is this project funded?	<ul> <li> % internal funding</li> <li> % public funding – please specify:</li> <li> % public – private partnership (PPP)</li> <li>– please specify:</li> <li> % other – please specify:</li> <li>Comment:</li> </ul>			
<b>20.</b> Within this project do you cooperate with other organisations?	<ul> <li>yes</li> <li>no</li> <li>if yes continue with question 21</li> <li>if no continue with question 23</li> </ul>			
21. Is your cooperation	<ul> <li>□ national</li> <li>□ within European Union</li> <li>□ international</li> </ul>			
<b>22.</b> Do you cooperate with	<ul> <li>other National Libraries</li> <li>other libraries</li> <li>other memory institutions</li> <li>universities or research institutions</li> <li>private companies</li> </ul>			

		□ other – please specify:
		Comment:
23.	Is digitisation for this project carried out in- house or by an external contractor (e.g. a digitisation company that is doing the scanning)?	% in-house % outsourced Comment:
24.	Which of these criteria are relevant for selection for digitisation in this project?	<ul> <li>to enhance Access</li> <li>for Preservation</li> <li>to Provide Services (e.g. Print-on-Demand)</li> <li>Other – please specify:</li> </ul>
25.	Time period covered by the objects in the digitised collection, e.g. 1718 – 1782	
26.	Thematic coverage of the objects in the digitised collection, e.g. Botany, History of French Revolution	
27.	Geographic coverage of the objects in the digitised collection, e.g. Europe or Austria or Paris or Wales	
28.	Have you selected the digital Master Files of this project for digital long term preservation?	□ yes, all files □ yes, a subset of the files — please specify:
		□ no
		Comment:
<b>29.</b> Digitisation of text and images for other materials, continue with question 6.2		
30.	Do you digitise from the originals or from reproductions?	<ul> <li>Originals</li> <li>Microfilm</li> <li>Other – please specify:</li> <li>Comment:</li> </ul>
31.	Resolution	dpi
<b>32.</b> File Format of Master Files	Comment: <ul> <li>TIFF</li> <li>JPEG2000</li> <li>JPEG</li> <li>PNG</li> <li>other – please specify:</li> </ul>	
---	---	
<b>33.</b> Compression method used for Master Files (e.g. LZW Compression, Group-4 Compression)		
<b>34.</b> File Format of Access Copies	<ul> <li>TIFF</li> <li>JPEG2000</li> <li>JPEG</li> <li>PNG</li> <li>other – please specify:</li> </ul>	
35. Digitisation of text		
<b>36.</b> In this project, do you digitise documents containing fonts other than standard Latin fonts?	<ul> <li>yes</li> <li>no</li> <li>if yes continue with question 38</li> <li>if no continue with question 37</li> </ul>	
<b>37.</b> Do you plan to use OCR for this collection in the future?	□ yes □ no continue with question 46	
<b>38.</b> Which non-standard Latin fonts do the documents contain?	<ul> <li> % Gothic ("Fraktur") fonts</li> <li> % pre-1850 fonts</li> <li> % Cyrillic fonts</li> <li> % Greek fonts</li> <li> % other – please specify:</li> <li> % other – specify:</li> <li> % other – specify:</li> </ul>	
<b>39.</b> Do you use Optical Character Recognition (OCR) for digitised texts?	□ yes □ no	

		if yes continue with question 40 if no continue with question 46
40.	What percentage of the digitised texts of this collection is OCRed?	%
41.	What average OCR recognition rate do you obtain within this collection? (e.g. 95 % of the characters are recognised accurately)	%
42.	Do you post-process OCR results? (e.g. manual correction of OCR results)	□ yes – please specify: □ no
43.	For what purpose are the OCR results used:	<ul> <li>automatic indexing for searching the collections</li> <li>computed-assisted reading</li> <li>highlighting of search results within the digital images</li> <li>other – please specify:</li> </ul>
44.	Is OCR in this project carried out in-house or by an external contractor?	% in-house % outsourced
		Comment:
45.	What are the OCR costs per page for this collection?	Euro-Cent per page
<b>46.</b> Digitisation of audio material for other materials, continue with question 51		
47.	Sampling Rate of Master Files	□ 96 kHz □ 48 kHz □ 44,1 kHz □ other – please specify:
48.	Bit Depth of Master Files	<ul> <li>□ 24 bit</li> <li>□ 16 bit</li> <li>□ other – please specify:</li> </ul>
49.	File Format of Master Files	□ WAVE □ Broadcast WAVE

	□ AIFF □ MP3 □ other – please specify:	
<b>50.</b> File Format of Access Copies	<ul> <li>MP3</li> <li>Windows Media</li> <li>QuickTime</li> <li>RealMedia</li> <li>other – please specify:</li> </ul>	
51. Digitisation of video and film for other materials, continue with question 54		
<b>52.</b> File Format of Master Files	<ul> <li>MXF</li> <li>AVI</li> <li>MPEG-4 – please specify:</li> <li>SMIL</li> <li>Motion JPEG 2000</li> <li>other – please specify:</li> </ul>	
<b>53.</b> File Format of Access Copies	<ul> <li>Windows Media</li> <li>QuickTime</li> <li>RealVideo</li> <li>other – please specify:</li> </ul>	
7 Metadata		
<b>54.</b> Do you have descriptive metadata available for this collection?	<ul> <li>yes</li> <li>no</li> <li>if yes continue with question 56</li> <li>if no continue with question 55</li> </ul>	
<b>55.</b> Please specify why you don't have descriptive metadata available for this collection.	continue with question 62	
<b>56.</b> On what level is descriptive metadata available for this collection?	<ul> <li>Collection level metadata</li> <li>Object level metadata</li> </ul>	

	Comment:
<b>57.</b> What standards do you use for creation of descriptive metadata at object level?	<ul> <li>AACR</li> <li>Dublin Core</li> <li>RAK-WB</li> <li>other – please specify:</li> </ul>
<b>58.</b> What standards do you use for descriptive metadata storage and exchange at object level?	<ul> <li>MARC 21</li> <li>Dublin Core</li> <li>Qualified Dublin Core</li> <li>The European Library Application Profile for Objects</li> <li>MARCXML</li> <li>MODS</li> <li>MAB2</li> <li>MABXML</li> <li>other – please specify:</li> </ul>
<b>59.</b> What standards do you use for creation of descriptive metadata at collection level?	<ul> <li>NISO Z39.91 Collection Description Specification</li> <li>The European Library Application Profile for Collection Descriptions</li> <li>RSLP Collection Description Schema</li> <li>Michael Collection Description Schema</li> <li>other – please specify:</li> </ul>
<b>60.</b> How many descriptive metadata records at the object level are already available for the collection?	metadata records
<b>61.</b> How many descriptive metadata records at the object level will be available when the project is finished?	metadata records
62. What other types of metadata are available for this collection (e.g. technical metadata, preservation metadata, structural metadata)? – What standards do you use (e.g. MIX, PREMIS, METS)?	
8 Access and Services	

63.	Is collection available to other online services?	□ yes □ no if yes continue with question 66
		if no continue with question 64
64.	Do you plan to make this collection available to other services in the future?	<ul> <li>yes, The European Library by year:</li> <li>yes, other – please specify:</li> <li>by year:</li> <li>no</li> </ul>
65.	Via which protocols will you make the metadata of the collection available to other services in the future?	<ul> <li>OAI</li> <li>Z39.50</li> <li>SRU</li> <li>FTP</li> <li>Other – please specify:</li> <li>continue with question 69</li> </ul>
66.	To which other online services is your collection available?	<ul> <li>The European Library (TEL)</li> <li>Other – please specify:</li> </ul>
		Comment:
67.	Via which protocols do you make the metadata of the collection available to other online services?	<ul> <li>OAI</li> <li>Z39.50</li> <li>SRU</li> <li>FTP</li> <li>Other – please specify:</li> </ul>
68.	Do you plan to use additional different protocols in the future?	<ul> <li>yes, OAI by year</li> <li>yes, Z39.50 by year</li> <li>yes, SRU by year</li> <li>yes, other – please specify:</li> <li>by year</li> <li>no</li> </ul>
69.	How do you provide access to this collection now?	<ul> <li>On-site within library</li> <li>WWW – please provide URL:</li> <li>via external service – please specify and provide URL:</li> <li>other – please specify and provide URL:</li> </ul>

70. Do you use technical methods for providing controlled access to the collection?	<ul> <li>IP address control (e.g. only within the library's IP range)</li> <li>Authentication system (e.g. usernames / password etc.):         <ul> <li>Vendor provided</li> <li>In-house developed</li> <li>National</li> <li>Other – please specify:</li> </ul> </li> <li>Proxy Server / VPN Access control</li> <li>Smart Card systems (i.e. linked to the library membership card)</li> <li>Digital certificates</li> <li>Other – please specify:</li> </ul>
<b>71.</b> Within this project, is in-copyright material digitised?	<ul> <li>yes</li> <li>no</li> <li>if yes continue with question 72</li> <li>if no continue with question 74</li> </ul>
<b>72.</b> Under what provisos is digitisation of in- copyright material carried out in this project?	<ul> <li>With the owner's agreement</li> <li>By paying the owner a fee</li> <li>Under license</li> <li>Without formalities</li> <li>Other – please specify:</li> </ul>
<b>73.</b> Are there special provisos for giving access to this digitised in-copyright material?	□ yes – please specify: □ no
<b>74.</b> Are you licensing the access to the digital collection?	<ul> <li>yes</li> <li>no</li> <li>if yes continue with question 75</li> <li>if no continue with question 78</li> </ul>
<b>75.</b> For which materials do you use licenses?	<ul> <li>In-copyright material</li> <li>Out-of-copyright material</li> <li>Both</li> <li>Comment:</li> </ul>
<b>76.</b> What types of licenses do you use?	<ul> <li>Creative Commons</li> <li>Attribution</li> <li>Attribution-NoDerivs</li> <li>Attribution-NonCommercial-NoDerivs</li> </ul>

	<ul> <li>Attribution-NonCommercial</li> <li>Attribution-NonCommercial-ShareAlike</li> <li>Attribution-ShareAlike</li> <li>Other – please specify:</li> <li>Comment:</li> </ul>
<b>77.</b> What types of usage are licensed?	<ul> <li>Save a copy</li> <li>Print</li> <li>Reuse material</li> <li>Other – please specify:</li> <li>Comment:</li> </ul>
<b>78.</b> Do you provide free access to the collections or do users have to pay for (parts of) the service?	<ul> <li>Free access to full service</li> <li>Paid access to service or parts of the service, namely:</li> <li>Comment:</li> </ul>
<b>79.</b> How many users / unique visitors are accessing the collection per day?	
80. Additional Remarks	

**!!** Please complete <u>one</u> questionnaire <u>for each</u> digitised collection and <u>each</u> digitisation project mentioned in Question 12 of Part 2:Questionnaire!!

Thank you very much for your cooperation.