

DELIVERABLE

Project Acronym: Europeana 1914-1918
Grant Agreement number: 270894
Project Title: Europeana 1914-1918: Remembering the First World War – a digital collection of outstanding sources from European national libraries

D3.1 Synthesis of agreed digitisation plans

Revision: 1.0
Authors: Ellen Geisriegler, ONB
Contributions: all partners

Project co-funded by the European Commission within the ICT Policy Support Programme		
Dissemination Level		
P	Public	x
C	Confidential, only for members of the consortium and the Commission Services	

Revision History

Revision	Date	Author	Organisation	Description

Statement of originality:

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

Table of Contents

1. Executive Summary	4
2. Background Information of Project Partners.....	6
2.1 Experience of Digitisation	6
2.3 Infrastructure and Staffing	8
3. Digitisation Plans for Europeana Collections 1914-1918	10
3.1 Project Digitisation Objectives	10
3.2 Summary of Partners Digitisation Plans.....	11
3.3 Overview of Object Groups for Digitisation	13
3.2 Full-text creation/OCR	17
Appendix I: Digitisation Monitoring Template	

1. Executive Summary

This deliverable contains the 'Synthesis of agreed digitisation plans' to be carried out by the project partners working on the Europeana Collections 1914-1918 project. The information in it was compiled from a detailed questionnaire completed by all partners in spring 2010, following preliminary discussions at a skype 'Selection Criteria Workshop' on the 27th June 2011 and a Work package 3 'Kick-off Workshop' in Vienna on the 7th and 8th September 2011. At the workshop in Vienna all partners were asked to hold presentations about their digitisation plans and various issues such as metadata standards and data delivery to TEL (The European Library)/Europeana were discussed. As one of the results of the workshop a unified digitisation monitoring template for the biannual 'Statistical milestone reports on numbers of items digitised to month x' was produced, discussed, and agreed on. The blank digitisation monitoring template can be found as an appendix to this document.

Digitisation is due to begin in October 2011, and the target is to have 10.000 items digitised and available in Europeana by the end of year one of the project.

List of partners contributing content:

No.	Short name	Name	Country
1	SBB/SPK	Staatsbibliothek zu Berlin / Stiftung Preußischer Kulturbesitz	Germany
2	BnF	National Library of France / Bibliothèque Nationale de France	France
3	BNU	National University Library of Strasbourg/ Bibliothèque Nationale et Universitaire de Strasbourg	France
4	BNCRM	National Central Library of Rome/ Biblioteca Nazionale Centrale di Roma 'Vittorio Emanuele II'	Italy
5	BNCF	National Central Library of Florence/ Biblioteca Nazionale Centrale Firenze	Italy
6	KBR	Bibliothèque royale de Belgique – Koninklijke Bibliotheek van België/ Royal Library of Belgium	Belgium
7	BL	The British Library	United Kingdom
8	KB	Det Kongelige Bibliotek, Nationalbibliotek og Kobenhavns Universitetsbibliotek / Royal Library of Copenhagen	Denmark
9	ONB	Österreichische Nationalbibliothek/ Austrian National Library	Austria
10	NLS	Narodna Biblioteka Srbije/ National Library of Serbia	Serbia
12	ICCU	Istituto Centrale per il Catalogo Unico delle Biblioteche Italiane e per le Informazioni Bibliografiche/ Central Institute for the Union Catalogue of Italian Libraries (ICCU)	Italy

donations (manuscripts, letters, images, drawings), public private partnerships (books), ad hoc bilateral projects (books), and selected minor projects of interest to sponsors (images, manuscripts, sheets of music).

The Austrian National Library (ONB) describes its digitisation programme as based on two pillars:

Large-scale digitisation projects have started in 2003 and are carried out by external service providers (e.g. 'ANNO- Austrian Newspapers Online', 'ALEX', 'Photographs and Historic Portraits', etc.) In September 2010 the ONB started the large-scale digitisation project 'Austrian Books Online'. Within this project ONB will digitise and make available online its complete holdings of historical books from the 16th to the 19th century (about 600.000 public domain works). The project will be carried out in a Public Private Partnership with Google.

Special collections (rare, fragile, valuable objects) are almost always digitised in-house with new special scanners for fragile and valuable materials (e.g. incunabula, papyri, flyers, Hebraica, first editions of Austrian authors, books on women's movement, and others).

The 'Digital Centre' at the **National Library of Serbia (NLS)** was set up in May 2003 with eight employees. The NLS purchased three professional scanners, one in 2005 (BookEye) and two in 2008 (Atiz). A special database was created according to the model of The European Library (TEL) by NLS IT specialists. The NLS 'Digital Library' contains more than 1.5 million items from 100 different collections so far.

The Central Institute for the Union Catalogue of Italian Libraries (ICCU) acts as aggregator of metadata and digital content from the library sector. It has collected over 8 million digital files (jpg, mp3, djvu, pdf) and 650.000 metadata items available on the portal <http://www.internetculturale.it/opencms/opencms/it> (11 content providers through OAI-PMH and 55 partners in digitisation projects).

The ICCU is also coordinator of four content providers of the portal <http://www.14-18.it>

- Istituto per la Storia del Risorgimento Italiano e Museo Centrale del Risorgimento (MCRR)
- Biblioteca Universitaria Alessandrina (BUA)
- Biblioteca di Storia Moderna e Contemporanea (BSMC)
- Biblioteca Nazionale Centrale Roma

2.2 Access Models to Content

Most partners have a mixed economy for presenting their digitised content, but the emphasis is on free and open access wherever possible. The entire content provided by the Europeana Collections 1914-1918 project will be available for free and without any restrictions. In recent years most partners have implemented digital repositories which enable them to manage access according to any restrictions necessary for reasons of IPR or data protection. The project-based history of digitisation means that much content is available through special subject portals.

Some examples for accessing to the digital objects of the partner libraries:

SBB/SPK:

[Digital Library](http://digital.staatsbibliothek-berlin.de) (<http://digital.staatsbibliothek-berlin.de>)

BnF:

[Gallica Bibliothèque Numerique](http://gallica.BnF.fr/) (<http://gallica.BnF.fr/>)

BL:

[Gutenberg Bible](#) (BL's 2 copies of this bible;
<http://www.bl.uk/treasures/gutenberg/homepage.html>)

[Archival Sound Recordings](#) (over 45,000 audio files; <http://sounds.bl.uk/>)

ONB:

[ANNO](#): Austrian Newspapers online (<http://anno.onb.ac.at>)

[ALEX](#): Portal on historical legal texts (<http://alex.onb.ac.at>)

[Picture Archive Austria](#) (<http://bildarchivaustria.at>)

NLS:

[Digital Library National Library of Serbia](#) (<http://digital.nb.rs>)

ICCU:

[Internet Culturale](#) (<http://www.internetculturale.it/opencms/opencms/it>)

[CulturaItalia](#) (<http://www.culturaitalia.it/pico/>)

At the end of the project all the digitised content will be available via Europeana (www.europeana.eu) and be harvested by TEL - The European Library.

2.3 Infrastructure and Staffing

Staffing levels for Europeana Collections 1914-1918 project vary. Most partner libraries have between 1 and 20 staff members working in digitisation, usually in dedicated units, but this might also be supplemented by project staff. The project partners have decided for three different solutions to digitise the contents of the Europeana Collections 1914-1918 project:

1. Some partners have already built up their own in-house digitisation centres and hold scanners and IT infrastructure. For this partners it is the cheapest way to use their existing digitisation infrastructure, e.g. by hiring scan operators. In these cases no subcontracting costs for digitisation but higher personal costs will occur.
2. Other partners will be using combined solutions. They have their own digitisation infrastructure (e.g. to be able to digitise very fragile or valuable material that should not leave the institution for conservation or insurance reasons). However, the digitisation centre might not be adjusted for a larger amount of objects. For these reasons, a part of the material will be treated by subcontractors (inside or outside of the libraries). In these cases proportional subcontracting costs will occur.
3. Finally, some partners have to choose a completely outsourced solution, as they don't have the adequate infrastructure to build up a digitisation centre. In these cases the entire digitisation will be subcontracted.

Four partners report that they carry out all their digitisation in-house. Four other partners will be using a combined solution outsourcing some of their digitisation work and three partners will completely outsource the whole digitisation. In some cases, it is mass digitisation that is

outsourced; other libraries outsource material in formats they do not have the appropriate equipment to do the work themselves.

Regarding the letting-of of subcontracts, the project partners are instructed to follow the standards described in the Grant Agreement and consider their national regulations.

Overview
digitisation

No.	Short name	In-house digitisation/ outsourced digitisation/ combined solution
1	SBB/SPK	in-house digitisation
2	BnF	combined solution
3	BNU	combined solution
4	BNCRM	outsourced digitisation
5	BNCF	outsourced digitisation
6	KBR	in-house digitisation
7	BL	in-house digitisation
8	KB	in-house digitisation
9	ONB	combined solution
10	NLS	combined solution
12	ICCU	outsourced digitisation

on partners'
plans:

3. Digitisation Plans for Europeana Collections 1914-1918

3.1 Project Digitisation Objectives

Europeana Collections 1914-1918 will create by 2014, the centenary of the outbreak of the First World War, a substantial digital collection of First World War items from national libraries and other partners from eight European countries that found themselves on different sides of the historic conflict. This digital collection will comprise books, newspapers, trench journals, maps, music sheets, children's literature, photographs, posters, pamphlets, leaflets, propaganda leaflets, original art, religious works, medals, coins and many more.

The project partners will make 425.000 WWI sources publicly and freely available online via Europeana for the first time, including content that is often rare and highly fragile because of the deteriorating quality of the paper it was produced on, and very often only accessible in reading rooms.

To guarantee and monitor the success of the digitisation project, the project consortium has agreed on the 'Monitoring Selection / Digitisation/ Online in Europeana plan' below.

Objective/ expected result	Indicator name	Expected Progress		
		Year 1	Year 2	Year 3
Selection and Digitisation Process				
1. Number of selected items	SELITM	100.000	200.000	350.000
2. Number of digitised items	DIGITM	50.000	150.000	350.000
3. Number of items online in 'Europeana' (including approx. 75.000 items, that have already been digitised outside the project)	EURITM	10.000	100.000	425.000

At the Work Package 3 'Kick-off workshop' on 7th and 8th September 2011 in Vienna, the project partners have produced, discussed, and agreed on a standardised 'Digitisation monitoring template'. This 'Digitisation monitoring template' will be used by all partners for the biannual statistical reports on the 'Numbers of items selected to month x' (MS 10-14), the 'Numbers of items digitised to month x' (MS 16-20), and the 'Numbers of items ingested to Europeana to month x' (MS 24-26). A blank 'Digitisation monitoring template' can be found as an appendix to this document.

During the proposal phase in spring 2010, all partners were asked to indicate object groups/ collections and approximate numbers of items to be digitised in the project. As these estimations had to be done in a very short time. Therefore it was not possible for all libraries to enumerate all their WWI collections at this stage.

Regarding the object groups, the BNCRM is planning to digitise about 29.950 images of about 3.500 documents (fiction, children's literature, guidebooks and manuals, song books, publicistic material, maps, picture material, manuscript material and many more).

The **BNCF** is planning to outsource its digitisation activities in the project Europeana Collections 1914-1918 to a subcontractor. Most of BNCF's digitisation projects in the past have been carried out by outsourcing and under the scientific supervision of the BNCF. It is planned to start the call for tender by the end of 2011, which will foresee a contract for about two years.

The access to the digitised objects will be gained by the library OPAC of the BNCF. As the BNCF has been a partner of The European Library (TEL) for several years, the harvesting of the collections should be easily performed.

In the project, the BNCF is mainly planning to digitise printed books (fiction, children's literature, religious literature, reference works, etc.), songbooks, publicistic material, picture material, maps, manuscript material, and some other objects.

The **KBR** will carry out the digitisation for the Europeana Collections 1914-1918 project in its own digitisation centre (e.g. with Bookscanner Qidenus RBS TT Pro Scanner ; MicrofilmScanner W&W 8850; for all documents < A3: flat scanner type Epson; for documents > A3: flat Scanner JumboScan or Zeutchel A0). The KBR also holds a Canon camera for some precious documents that cannot be digitised by a flat scanner. Within the project, KBR will primarily digitise printed books (fiction, religious literature, war libraries, reference works, guide books and manuals), song books and sheet music, picture material, publicistic material, and maps. The digitised objects will be made accessible via the library OPAC, the OAI-PMH infrastructure for the data delivery to TEL still has to be built up.

The **BL** is currently working with a team of curatorial 'selectors' from all areas of the BL who are now generating initial item-level information for the selected items. This will be completed by October 2011. Furthermore the BL has recruited a project manager and will soon recruit two metadata editors. Within the planning process, the BL will also define: digitisation workflows (including OCR) and timescales with its in-house digitisation studio (digitisation will start in January 2012), ingest workflows into the BL's Digital Library System (DLS), metadata export possibilities and processes with BL's metadata services department. The clarification of IPR-issues is also in progress.

Objects to be digitised in the project by the BL will come from various collections: e.g. printed books - fiction and reference works (e.g. poems of Ivor Gurney), manuscripts (India Office war records, John Fould's 'A World Requiem', Vaughan Williams's 'Pastoral Symphony'), printed music scores, periodicals (camp magazines, prisoner of war camp theatre/concert/sport programmes), collection of local official recruitment/propaganda posters, photographs (including troop photographs), posters, and many more.

The **KB** has chosen mainly books (in Danish 1914 -1918 and books regarding the reunification of Schleswig with Denmark in 1920), images (field post, prisoners of war in camp), and leaflets (daily life, goods, organizations etc) for the digitisation in the project. The digitisation will be carried out in-house, in the digitisation centre of the KB. More than 3.600 photographs have already been scanned (front/reverse) and the planning of further digitisation is in progress.

The **ONB** will digitise publicistic material (newspapers, special editions of daily newspapers, pamphlets and leaflets dropped by airplanes), posters, picture material (collection war photography during WWI), children's drawings, small form graphics, transcripts of soldier's songs and books and brochures from the Department of Planned Languages.

The majority of the digitisation activities will be accomplished by subcontractors, only fragile objects will be digitised in-house in the Digital Service Department of the ONB. The digitisation of

posters and newspapers has already started in summer 2011 and it is planned to digitise various other object groups until the end of the year. At the beginning of 2012, the ONB will initiate a call for tender regarding the digitisation of the picture material, and the large collection of small form graphics.

The **NLS** has selected printed books (reference works, light fiction, children's literature, religious literature, war libraries, reference works, guidebooks and manuals and travel literature), songbooks and sheet music, publicistic material, pamphlets and leaflets, picture material, maps and manuscript material for the digitisation in the project. Digitisation will be carried out with a combined solution of subcontracting and in-house digitisation. The call for tender has been published by mid-September, according to the national law. The majority of the metadata will be exported from the NLS e-catalogue, for a small number of documents it will be necessary to create metadata from the card catalogue though. The NLS is already an experienced partner of TEL and has set up an OAI-PMH infrastructure. The results of the projects will be accessible in the 'Digital NLS' as one of the special collections – WW1.

The **ICCU** acts as an aggregator of metadata and digital content from the library sector. The ICCU is the co-ordinator of three content providers in the project "Europeana Collections 1914-1918":

- Istituto per la Storia del Risorgimento Italiano e Museo Centrale del Risorgimento (MCRR)
- Biblioteca Universitaria Alessandrina (BUA) -Biblioteca di Storia Moderna e Contemporanea (BSMC)

The collections to be digitized will contain publicistic and manuscript material, picture material, original art, and songbooks and music sheets for a total of 185.000 items.

21.000 newspapers and 40.6000 pictures have already been digitised and are already available to be ingested in Europeana.

The digitisation of all ICCU collections will be carried out by subcontractors. The call for tender will be initiated at the end of October 2011. The digitised objects will be made available on the portal <http://www.14-18.it>

3.3 Overview of different Object Groups for Digitisation

In order to gain an overview of the content that is going to be digitised in the project and for monitoring the progress, the project consortium classified the different item types into object groups. These object groups were further split into subcollections.

As already mentioned previously (see chapter 3.1) most partner libraries stuck to their original digitisation plans except for minor changes. Only a few libraries had to make changes in the selection of their object groups. However, these changes only affected the collections to be digitised and not the number of items. All libraries have indicated that they will be able to digitise the same number of items (or even more items) than they have originally indicated in the Description of work.

The following tables shall give an overview on the object groups and subcollections selected for digitisation to month 6 of the project. Numbers presented in these tables might still increase in the next months as the selection processes are still going on in most partner libraries. Minor alterations of the object groups and subcollections will also be possible.

The tables also contain information about the metadata standards partners are intending to use, beyond the required standards of the Europeana Semantic Elements (ESE) which are already agreed as an essential requirement for project partners.

Table 1. Printed books

Subcollections: Reference works; Fiction; Light fiction Poetry; Theatre; Children's literature; Military literature, guides manuals; War libraries; Daily life; Travel literature; Historical literature; Religious literature; Financial and economic literature; Political, propaganda literature, Medical literature; Welfare literature; Popular regional literature; others

Provider (short name)	Quantity	Metadata	Language
SBB	3.440 volumes	METS/ MODS, DC	German
BnF	5.340 volumes		French
BNU (with BDIC)	320 volumes	MARC, DC and probably EDM	French, German, Alsatian
BNCRM	560 volumes	DC	Italian
BNCF	1.150 volumes	DC	Italian
KBR	6.319 volumes	OPAC, card catalogue	French, Dutch, German
BL	906 *	MARC	English, Italian, Serbian, Croatian, German, Russian, French
ONB	70 volumes	Descriptive metadata to be created	German, Esperanto
NLS	405 volumes	OPAC, card catalogue	Serbian

Table 2. Songbooks and sheet music

Subcollections: Manuscript music book; Manuscript music sheets; Printed music book; Printed music sheets; Music Journals & periodicals; Other printed materials (folders, concert programs, postcards)

Provider (short name)	Quantity	Metadata	Language
SBB	200 items	METS/MODS, DC	German
BnF	400 items		French
BNU	20 items	MARC, DC and probably EDM	French, German, Alsatian
BNCRM	28 items	DC	Italian
BNCF	280 volumes	DC	Italian
KBR	250 items	OPAC, card catalogue	French, Dutch, German
BL	970 *	MARC and EAD in .xls format	English, German
ONB	363 items	Descriptive metadata to be created	German
NLS	100 items	Card catalogue	Serbian
ICCU/ Biblioteca Alessandrina/ Biblioteca	240 items	DC	Italian, European languages

Storia moderna			
----------------	--	--	--

Table 3. Publicistic material

Subcollections: Newspapers; Other periodicals; Pamphlets and leaflets; Other publicistic material (community newsletters, specials announcements of companies and societies, posters, commemorative announcements)

Provider (short name)	Quantity	Metadata	Language
SBB	1.920 items	METS/MODS, DC	German
BnF	27.660 items		French
BNU	100 items	MARC, DC and probably EDM	French, German, Alsatian
BNCRM	2.646 items	DC	Italian
BNCF	2.100 items	DC	Italian
KBR	1.285 items	OPAC, card catalogue	French, Dutch, German
BL	725 *	MARC	English, French, German, Dutch, Bulgarian, Croatian, Serbian
KB	750 items	ESE	Danish
ONB	159.000 pages	Descriptive metadata to be created	German, other
NLS	450 items	Card catalogue, has to be completed and imported to OPAC	Serbian
ICCU/ Biblioteca Alessandrina/ Biblioteca Storia moderna	4.790 items	DC	Italian, European languages
ICCU/ already digitised	21.000 items	DC	Italian

Table 4. Picture material

Subcollections: Drawings, sketchbooks; Paintings; Posters; Photographs; Postcards, Poststamps; Medals, coins; Other Picture Material

Provider (short name)	Quantity	Metadata	Language
SBB	400 items	METS/MODS, DC	
BnF	6.900 items	OPAC	French
BNU	2.138 items	MARC, DC and probably EDM	French, German, Alsatian
BNU/ BDIC	26.300 items	MARC, DC and probably	

		EDM	
BNCRM	285 items	DC	Italian
BNCF	200 items	DC	Italian
KBR	600	OPAC, card catalogue	French, Dutch, German
BL	15 *	MARC	English, other
KB	5.722 items	ESE	
ONB	39.160 items	Descriptive metadata to be created	German, other
NLS	200 items	Card catalogue, has to be completed and imported to OPAC	Serbian
ICCU/ Museo del Risorgimento	44.050 items	DC	Italian
ICCU/ Biblioteca Alessandrina/ Biblioteca Storia moderna	1.460	DC	Italian, European languages
ICCU/ already digitised	40.600	DC	Italian

Table 5. Manuscript material

Subcollections: War postcards; War journals; Diaries; Letters

Provider (short name)	Quantity	Metadata	Language
SBB	500 items	METS/MODS, DC	German
BnF	40 items	OPAC	French
BNU	20 items	MARC, DC and probably EDM	French, German, Alsatian
BNCRM	7 items	DC	Italian
BNCF	2 items	DC	Italian
BL	164 *	EAD in .xls	English, French, German
NLS	100 items	Card catalogue, has to be completed and imported to OPAC	Serbian
ICCU/ Museo del Risorgimento	123.400 items	DC	Italian
ICCU/ Biblioteca Alessandrina/ Biblioteca Storia moderna	10 items	DC	Italian, European languages

Table 6. Maps

Subcollections: Manuscript maps; Printed maps; Atlas; Other geographic material

Provider (short name)	Quantity	Metadata	Language
BnF	290 items	OPAC	French
BNU	290 items	MARC, DC and probably EDM	French, German, Alsatian
BNCRM	67 items	DC	Italian
BNCF	50 items	DC	Italian
KBR	350 items	OPAC, card catalogue	French, Dutch, German
BL	163 *	MARC	English, French, German, Italian
NLS	70 items	OPAC	Serbian

Table 7. Misc.

Subcollections: Printed letters; Biographies; Reports; Journals; Memoirs; Episodes; Other

Provider (short name)	Quantity	Metadata	Language
SBB	420 items	METS/MODS, DC	German
BnF	100 items	OPAC	French
BNCRM	157 items	DC	Italian
BNCF	100 items	DC	Italian
KBR	1.854 items	OPAC, card catalogue	French, Dutch, German
BL	476 *	DC	English
KB	954 items	ESE	Danish
NLS	100 items	Card catalogue, has to be completed and imported to OPAC	Serbian
ICCU/ Museo del Risorgimento	11.000 items	DC	Italian
ICCU/ Biblioteca Alessandrina/ Biblioteca Storia moderna	50 items	DC	Italian, European languages

* The selection of items is still ongoing. The number of selected items will meet at least the number indicated in the Description of Work.

3.2 Full-text creation/OCR

The full-text creation of a large part of the digitised documents is intended by nearly all partners. The OCR will be done separately by each institution. The Consortium will exchange experiences of full-text creation, e.g. regarding the handling of difficult typefaces like the German "Fraktur". The

full-text will be integrated in the already existing or soon to be established full-text indices of the separate digital libraries. The Europeana portal currently does not offer the tools to search in full-text indices.

The following table shows whether and how partners intend to use OCR.

No.	Short name	OCR / OCR standards
1	SBB/SPK	Yes, wherever possible./ All textual object records are marked with a predefined level of ability for OCR (I-V) in regard to typeface and typeset. After the whole digitisation process has been finished, the OCR-able records will be read out automatically. The according images will be sent to an external OCR-Server (OCR-cluster for mass-OCR, based on Abby Recognition Server 2.0 R3 Extendit Edition®) for full-text creation in METS/ALTO format.
2	BnF	Yes, wherever possible.
3	BNU	Yes.
4	BNCRM	Yes./ OCR is planned for a small group of newspapers. The images will be put on the local server in order to test the OCR.
5	BNCF	Yes./ OCR is planned for all material groups.
6	KBR	Yes./ Creation of OCR with Abyy Fine Reader Pdf searchable for consultation.
7	BL	Yes.
8	KB	Yes.
9	ONB	No, not within this project. The ONB is planning to create full-text for the newspapers digitised for Europeana Collections 1914-1918 within another project, though.
10	NLS	Yes./ OCR is planned for a of small amount of printed books, control and corrections will be done in-house
12	ICCU	No