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

_____

D3.2   Updated 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

Table of Contents

1. Executive Summary

This deliverable contains the 'Updated 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 7 - 8 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 Europeana via TEL (The European Library) were discussed. As one of the results of the workshop a unified digitisation monitoring template for the bi-annual 'Statistical milestone reports on numbers of items digitised to month x' was produced, discussed, and agreed on.

All project goals for year 1 were successfully achieved and even exceeded within the first twelve months of the project: 81% of the items selected, 25% of all items digitised, and 12% of all items ingested in Europeana.

For the further planning of project year 2, two Consortium Meetings (Rome 6-7 February 2012; Belgrade (23-24 October 2012) and two Executive Board Meetings (Brussels 28 October 2011; Belgrade 22 October 2012) were held. At the Consortium Meeting in Rome 6th and 7th February 2012 the digitisation monitoring template was extended (see plain template attached to this document). It was further agreed to use the collaborative platform 'Editgrid' to simplify the reporting and to avoid confusion caused by several different versions of the progress reporting template.

At the Consortium Meeting in Belgrade the updated digitisation progress of all partners was presented and the necessary efforts for the second half of the project were discussed.

2. Background Information of Project Partners

2.1 Experience in Digitisation (Status: Beginning of the project, May 2011)

The project Europeana Collections 1914-1918 provides a unique opportunity for some of the leading research libraries from across Europe to share their extensive experience in digitisation for the benefit of each other and of Europeana. Eleven partners (out of twelve) contributing content comprise national and university libraries and other partners from eight different countries.

The **Staatsbibliothek zu Berlin (SBB/SPK)** is highly skilled in digitisation processes and is committed to providing a full range of digital library services. A workflow management for all digitisation processes is well established. Management systems to monitor quotas for digitisation, data input, cataloguing as well as quality control are in place for collaborative digitisation programmes on a national scale and for the running of mass digitisation processes.

The **National Library of France (BnF)** is involved in the field of digitisation since 1992. Since 2009, the BnF is producing 11 million pages, 2.5 million iconographic documents, and 13 million pages from retrospective OCR-isation contracts per year. In the future the treatment of more than 50.000 tables of contents is planned.

The National University Library of Strasbourg (BNU) gained first experiences in digitisation in the 1990s and conducted smaller projects in the early 2000s. In autumn 2011 a 'Digital Library Service' was established, the opening of the BNU 'Digital Library' is planned for January 2012.

The National Central Library of Rome (BNCRM) started its digitisation activities in 2003. The first important project was the founding of the 'Biblioteca Digitale Italiana (BDI)' funded by the MiBAC. The BNCRM participated in the 'Italian Digital Library' with the following projects: The ancient inventories of manuscripts as well as the autographs catalogue: 'Manoscritti musicali' (16.000 images), 'Periodici storici' (210.000 images), 'Immagini della Grande Guerra' (190 photographs), and many more. The BNCRM also conducted several other projects during the same period: 'Fondo Dannunziano' (36.100 images), 'Fondo Siciliani' (120.000 images of printed music rare editions, MP3 files from 5.000 long playing records), 'Biblioteca Nonantolana Virtuale' (16.208 images), etc..

The **National Central Library of Florence (BNCF)** started its first digitisation projects in 1990. The whole digital collections of the BNCF can be estimated at about 3.000.000 images: about 190.000 manuscripts, 442.000 ancient books and 2.350.000 modern books. Additionally, BNCF is currently involved in digitisation projects with ProQuest and Google.

The **British Library (BL)** is looking back on over ten years experience of digitising, including the collections 'Gutenberg Bible' (BL's two copies of this bible), 'Shakespeare in Quarto' (21 pre-1642 copies of his plays), 'Archival Sound Recordings' (over 45.000 audio files), 'British Library Newspapers 1800-1900' (2 million pages of newspapers), digitisation contracts with Microsoft (100.000 19th-century books) and recently Google (250.000 18th- and 19th century books) and many more. The BL has also developed a 'Turning the Pages Software' for Library treasures and is currently building a new 'Digital Item Presentation System' (DIPS) to hold all BL digital content including Europeana Collections images.

The **Royal Library of Copenhagen (KB)** holds about 220.000 digitised items (texts and images) from the ongoing digitisation of selected parts of special collections (manuscripts, photographs, drawings, leaflets, maps, sheets of music, books), special projects financed by funding or

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/ (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
- Instituto 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 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 these 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 of the Europeana Collections 1914-1918 project 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 of partners' digitisation plans:

No.	Short name	
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

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

All success indicators were successfully exceeded in project year 1.

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' is used by all partners for the bi-annual 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.

Within the first half of project year 2, the 'Digitisation monitoring template' was extended with several columns (number of objects ready for harvesting) to gain a better overview of the progress in each institution, and several columns concerning controlled vocabulary/metadata (subject headings, object types, authority files, geo-coordinate/spatial information, dates/times) of the partners.

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.

The precise selection of items has started at all partner institutions in month 2 (June 2011) of the project with a 'Selection criteria workshop' (see D2.1 'Specification of Selection Criteria and Thematic Subcollections applied in the project').
By month 24 (end of project year 2) most partners will have very likely finished their selection process (smaller exceptions are still possible).

In preparation of the Deliverable 'Updated synthesis of agreed digitisation plans', all partners were asked to update their digitisation plans in the 'Digitisation monitoring template' as far as possible to this phase of the project.

The result of this update is very promising. All libraries have indicated that they will be able to digitise the same number of items (or even more items) than originally indicated in the Description of Work. Most partner libraries stuck to their original digitisation plans except for minor changes. Only few libraries noticed minor deviations which are related to internal shifts within the collections without any effect on the amount of items.
Several partner institutions have already ingested material to Europeana (BnF, ONB, NLS, KB, ICCU), the others will follow within the next weeks and months.

3.2 Summary of Partners Digitisation Plans

While all the information on the digitised material in the Europeana Collections 1914-1918 project will be available through the Europeana portal, the partners will make available the digital objects on their presentation platforms.

As the **SPK/SBB** has a high-level equipped digitisation centre available, the digitisation for the Europeana Collections 1914-1918 project is done completely in-house. Three scan operators were hired and one special camera is going to be purchased for the project. OCR will be done by subcontracting though. Objects groups, chosen for digitisation are: printed books, songbooks and sheet music, publicistic material (periodicals, pamphlets and leaflets, others, picture material, maps, and manuscript material).
The SPK/SBB has already started with the digitisation of printed books and made first items available via its Digital Library (http://digital.staatsbibliothek-berlin.de/dms/suche/?DC=krieg.1914.1918). The objects will be further accessible via 'The German Digital Library (Deutsche Digitale Bibliothek – DDB)' and 'The Central Index of Digitised Imprints (Zentrales Verzeichnis digitalisierter Drucke – ZVDD)'.

In the project Europeana Collections 1914-1918 the **BnF** is cooperating with the BNUS (National and Académique Library of Strasbourg) the Défense Département Documentation Centre, and some other French libraries. The BnF is digitising a wide range of documents with a specific attention to iconographic documentation and also considers some rare, sometimes one of a kind, collections. The digitised documents (printed books, song books, publicistic material, picture material, manuscript material, and others) are made accessible via the library portal 'Gallica'. The digitisation is carried out with a combined solution, via subcontracting and in-house digitisation. So far, BnF has digitised printed books, publicistic material, picture material, and maps and has ingested picture material in Europeana.

The **BNU** is cooperating with the BDIC (Bibliothèque de Documentation Internationale Contemporaine) in this digitisation project. A 'light' call for tender (for the digitisation of posters) was carried out in autumn 2011, followed by two calls for the printed material, maps, and the picture material from BDIC in spring 2012. The digitisation should be completed by December 2013. Manuscript material, medals and coins, music scores, photographs, postcards and posters (from the BDIC) are digitised in-house between 2012 and 2013. Accessibility to the digitised

objects will be gained via the Digital Library portal of the BNU, partly via the Digital Library of BDIC and possibly via the library portal 'Gallica'.

In project month 18, BNU and BDIC have already digitised publicistic material and picture material.

The digitisation of the **BNCRM** in the project was outsourced according to the Italian law. A specifications document regarding technical features and metadata standards was provided. 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).

So far, BNCRM has digitised various kinds of publicistic material.

The **BNCF** is outsourcing 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. The call for tender was started in early 2012 and foresees 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.

To month 18, BNCF has started to digitise printed books from various collections.

The **KBR** is carrying 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 Zeutschel 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.

So far, the KBR has digitised printed books, songbooks and sheet music, publicistic material, maps, and misc. material.

The **BL** is still 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 but has nearly finished the selection process. Furthermore the BL has recruited a project manager and metadata editors. Within the planning process, the BL has defined digitisation workflows (including OCR) and timescales with its in-house digitisation studio (digitisation started 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.

So far, the BL has digitised printed books, manuscript material, picture material, and misc. material and has ingested some manuscript material in Europeana.

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.
To project month 18, the KB has digitised publicistic material, picture material, misc. material, and maps and has ingested picture material and misc. material in Europeana.

The **ONB** is digitising 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 are accomplished by subcontractors, only fragile objects will be digitised in-house in the Digital Service Department of the ONB. At the beginning of 2012, the ONB will initiated a callfor tender regarding the digitisation of the picture material, and the large collection of publicistic material.
So far, the ONB has digitised a large amount of objects from publicistic and picture collections and has already ingested 12.600 issues of newspapers into Europeana.

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 is carried out with a combined solution of subcontracting and in-house digitisation. The call for tender has been published by mid-September 2011, 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.
To project month 18, the NLS has digitised printed books, publicistic material, maps, manuscript material, picture material, and misc. material.

The **ICCU** acts as an aggregator of metadata and digital content from the library sector. ICCU is the co-ordinator of four 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)
- Biblioteca Nazionale Centrale Roma

The digitisation of all ICCU collections is carried out by subcontractors. The call for tender was initiated at the end of October 2011. The digitised objects will be made available on the portal
http://www.14-18.it
The collections to be digitised contain publicistic and manuscript material, picture material, original art, and songbooks and music sheets for a total of 185.000 items.
So far, the partners of ICCU have digitised manuscript material, picture material, publicistic material, songbooks and music sheets, manuscripts, and misc. material.

3.3 Overview of different Object Groups for Digitisation

In order to gain an overview of the content that is being 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 DoW.

The following tables shall give an overview on the object groups and subcollections selected for digitisation to month 18 of the project. Numbers presented in these tables might still slightly increase in the next months as the selection processes are still going on in some 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	German
BnF	5.340 volumes	MARC21	French
BNU (with BDIC)	320 volumes	MARC, DC and probably EDM	French, German, Alsatian
BNCRM	569 volumes	DC	Italian
BNCF	1.150 volumes	DC	Italian
KBR	6.319 volumes	MARC 21	French, Dutch, German
BL	1.440 volumes	MARC	English, Italian, Serbian, Croatian, German, Russian, French
ONB	70 volumes	MARC 21/ MAB 2	German, Esperanto
NLS	405 volumes	MARC/DC	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	German
BnF	400 items	MARC21	French
BNU	27 items/volumes	MARC, DC, probably EDM	French, German, Alsatian
BNCRM	30 volumes	DC	Italian
BNCF	387 volumes	DC	Italian
KBR	272 items	MARC21	French, Dutch, German
BL	2.500 items	MARC/XLS	English, German
ONB	363 items	Descriptive Metadata	German
NLS	100 items	MARC/DC	Serbian
ICCU/ Biblioteca Alessandrina / Biblioteca Storia moderna	600 pages	DC	Italian, European languages

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	German
BnF	27.660 items	MARC21	French
BNU	100 items	MARC, DC, probably EDM	French, German, Alsatian
BNCRM	4.047 items	DC	Italian
BNCF	2.045 items	DC	Italian
KBR	1.285 items	MARC21	French, Dutch, German
BL	826 serials/items	MARC	English, French, German, Dutch, Bulgarian, Croatian, Serbian
KB	994 items	ESE	Danish
ONB	30.000 issues/items	ESE, archival descriptions, MARC21/MAB	German, other European languages
NLS	450 volumes/items	MARC/DC	Serbian
ICCU/ Biblioteca Alessandrina/ Biblioteca Storia moderna	5.500 pages	DC	Italian, European languages
ICCU (already digitised 2011)	21.000 pages	DC	Italian

Table 4. Picture material
Subcollections: Drawings, sketchbooks; Paintings; Posters; Photographs; Postcards, Post stamps; Medals, Coins; Other Picture Material

Provider (short name)	Quantity	Metadata	Language
SBB	400 items	METS/MODS	
BnF	6.900 items	MARC21	French
BNU/ BDIC	3.748 items/ 38.700 irems	MARC, DC, probably EDM	
BNCRM	248 items	DC	Italian
BNCF	200 items	DC	Italian
KBR	1.159 items	MARC21	French, Dutch, German
BL	517 items	MARC/XLS	English, other
KB	5.722 items	ESE	
ONB	45.000 items	Descriptive metadata	German, other European languages
NLS	200 items	MARC/DC	Serbian
ICCU/ Museo del Risorgimento	44.050 items	DC	Italian
ICCU/ Biblioteca Alessandrina/ Biblioteca Storia moderna	1.856 items	DC	Italian, European languages
ICCU (already digitised)	40.600 items	DC	Italian

Table 5. Manuscript material

Subcollections: War postcards; War journals; Diaries; Letters

Provider (short name)	Quantity	Metadata	Language
SBB	527 items	METS/MODS	German
BnF	40 items	MARC21	French
BNU	58 items	MARC, DC, probably EDM	French, German Alsatian
BNCRM	8 items	DC	Italian
BNCF	2 items	DC	Italian
BL	217 items	XLS	English, French, German
NLS	100 items	MARC/DC	Serbian
ICCU/ Museo del Risorgimento	123.400 pages	DC	Italian
ICCU/ Biblioteca Alessandrina/ Biblioteca Storia moderna	10 items	DC	Italian

Table 6. Maps

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

Provider (short name)	Quantity	Metadata	Language
BnF	100 items	MARC21	French
BNU	290 items	MARC, DC, and probably EDM	French, German, Alsatian
BNCRM	67 items	DC	Italian
BNCF	50 items	DC	Italian
KBR	595 items	MARC21	French, Dutch, German
BL	162	MARC	English, French, German, Italian
NLS	72 items	MARC/DC	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	MARC21	French
BNCRM	157 items	DC	Italian
BNCF	100 items	DC	Italian
KBR	1.587 items	MARC21	French, Dutch, German
BL	1.854 items	MARC21	English
KB	48 items	ESE	Danish
NLS	100 items	MARC/DC	Serbian
ICCU/ Museo del Risorgimento	11.000 vol./pages	DC	Italian
ICCU/ Biblioteca Alessandrina/ Biblioteca Storia moderna	50 pages	DC	Italian, European languages

3.4 Full-text creation/OCR

The full-text creation of a large part of the digitised documents is intended by several partners and is done separately by each institution. The consortium will however 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 different digital libraries, and in the portal of The European Library (TEL). The Europeana portal currently does not offer the tools to search in full-text.

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

No.	Short name	OCR/ OCR standards
1	SBB/SPK	Yes, wherever possible for printed books/ 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 Abbyy Recognition Server 2.0 R3 Extendit Edition®) for full-text creation in METS/ALTO format.
2	BnF	Yes, whenever possible (for printed books, songbooks and sheetmusic, publicistic material)
3	BNU/BDIC	Yes, for printed books
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.
6	KBR	No.
7	BL	Yes, wherever possible (for printed books, publicistic material, misc. material).
8	KB	Yes, for misc. material.
9	ONB	Not within this project.
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.