

DELIVERABLE

Project Acronym: Europeana1914-1918
Grant Agreement number: 270894
Project Title: Europeana 1914-1918: Remembering the First World War – a digital collection of outstanding sources from European national libraries

D 3.4 Second report on numbers of items digitised in the project

Revision: 1.0
Authors: Ellen Geisriegler, Austrian National Library
 Susanne Tremml, Austrian National Library
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This deliverable contains original unpublished work except where clearly indicated otherwise.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
2.1 Collecting Data and Numbers of Project Partners.....	6
3. Project Progress to Month 24.....	9
3.1. Selection Progress	9
3.2. Digitisation Progress.....	12
3.3. Ingestion Progress.....	15
4. Summary	18
5. Annex I – ‘Digitisation progress monitoring template’	19

1. Executive Summary

This deliverable contains the 'Second report on numbers of items digitised in the project' of Europeana Collections 1914-1918 and shall give an overview of the number of objects selected, digitised, and ingested in Europeana until the end of the second project year.

The information in it was compiled from all partner institutions with the help of 'Digitisation progress monitoring templates' in paper forms in month 6 (October 2011) and in month 12 (April 2012) of the first project year. After project year one information on objects selected, digitised and ingested were gathered with the help of a collaborative workspace.

To ensure that the progress and processes of the project are correctly documented, comprehensible, and transparent, a 'Digitisation progress monitoring template' was created, discussed, and agreed on by all partners at the Work Package 3 Kick-off Workshop in Vienna on 7th and 8th September 2011. Despite the fact that the monitoring of the selection, digitisation and ingestion progress are concerning two workpackages (WP2 and WP3), it was agreed that one unified monitoring template will be used by all project partners. The 'Digitisation progress monitoring template' was slightly adapted after the first internal reporting (MS10 and MS 16: Statistical report on numbers of items selected/digitised to month 6). A modified blank template was delivered as appendix D.3.3 'First report on numbers of items digitised in the project'. To further simplify the documentation and data collection processes for the project partners it was agreed to use the collective workspace platform EditGrid¹ for this purpose.

In project month 24 (April 2013), all partners were asked to update the 'Digitisation progress monitoring template' with their current numbers of items. The information provided by the partners shall be presented on the following pages. Several graphs are included in the deliverable in order to present the documentation in a more descriptive way. All concrete numbers provided by the partners can be found in the 'Digitisation progress monitoring template' attached as an annex to this report.

¹<https://www.editgrid.com>

No.	Short name	Name	Country
1	SBB/SPK	Berlin State Library/ 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 Københavns 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	Instituto Centrale per il Catalogo Unico delle Biblioteche Italiane e per le Informazioni Bibliografiche/ Central Institute for the Union Catalogue of Italian Libraries (ICCU)	Italy

Partner 11, UBER (Humboldt-Universität zu Berlin) does not provide content.

2. Data Collection and Progress Reporting

In the DoW the project consortium has indicated that 425.000 WW1 sources will be made publicly and freely available online via Europeana within the project Europeana Collections 1914-1918. To guarantee and monitor the success of the digitisation project, the partners have agreed on the schedule 'Selection/Digitisation/Ingestion in Europeana' 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

2.1 Collecting Data and Numbers of Project Partners

The partners' digitisation plans have been described in detail in D3.1 'Synthesis of agreed digitisation plans'. For the collection and reporting of the partners' digitisation progress, a unified 'Digitisation progress reporting template' was created and agreed on by all project members in month 5. To further improve the quality of the collected data, to collect more essential information, and to gain an even better overview of the project progress, a few more columns have been added to the monitoring template.

To simplify the process of data collection for all partners, to avoid confusion and to guarantee constant availability of the latest version of the 'Digitisation progress monitoring template', the project consortium has decided to use the collaborative workspace platform EditGrid². All project partners were given access to a shared 'Digitisation progress monitoring template' on EditGrid and were asked to update their data directly online. Additionally, regular backups of the template were made during the reporting period to avoid a loss of data in case of problems with the platform.

The template provides one sheet for each partner institution and one as a summary with the numbers of all content providers.

²<https://www.editgrid.com>

Content and description of the monitoring template:

Column A (Provider) and B (Last date of reporting) contain the abbreviation of the content provider and the latest date of reporting. For D3.3 the last day of the reporting period was April, 30th2013.

As already described in Deliverable D3.1, the project consortium has agreed on seven different source types for the digitisation project:

- Printed books
- Songbooks and sheet music
- Publicistic material
- Maps
- Manuscript material
- Picture material
- Miscellaneous material.

The source types shall be indicated in **Column C (Source type)**.

All partners are asked to fill in the numbers of each source type according to the DoW into **column D (Proposed source type according to DoW)**.

For organisational reasons, most partner institutions split up the different source types into subcollections (**Column E – Collection Description**) and will deliver these subcollections to Europeana. To make it easier for the partners to document their progress in selection, digitisation, and data delivery, various categories for subcollections have been elaborated (following the WP2 ‘Selection criteria workshop’ in June 2011 and the WP3 ‘Kick-off workshop’ in September 2011). The subcollections are optionally extensible by the content providers in case an accurate term can’t be found in the provided list.

Source type 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

Source type 2: Songbooks and sheet music

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

Source type 3: Publicistic material

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

Source type 4: Picture material

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

Source type 5: Manuscript material

Subcollections: War postcards; War journals; Diaries; Letters

Source type 6: Maps

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

Source type 7: Miscellaneous material

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

In **column F (Unit of measurement)** all partners have to indicate the unit of measurement (e.g. volumes, pages, items, etc.) to enable a transparent counting.

In **column G (Numbers of items foreseen according to DoW)**, the content providers are asked to indicate the numbers of items foreseen for each subcollection according to the DoW. The total amount of items from the subcollections within one source type should be the same as the number of items related to the source type.

Column H (Actual numbers of items to be digitised) will be used to indicate the actual number of items (for each subcollection) to be digitised in the project. These numbers will be changing for most content providers within the progress of the project. For many content providers, the figures announced in the DoW for each subcollection are only a quantitative indication. Because the effective selection has to follow the specifications published in Deliverable 2.1 'Specification of the Selection Criteria and Thematic Subcollections applied in the project' and its Annex 1 'Handout', the first estimation will be continuously reevaluated by most content providers. Furthermore, curators may discover new relevant sets of documents which should be added in the selection. For that reason, the effective number of items by subcollection can be re-evaluated every six months to reflect the exact progress of the project.

Columns I to P are meant as **Production Monitoring Summary**, observing the present progress of the project. In the columns I, K, M, and O the partners will continuously update the numbers of items they have selected, digitised, ready for harvest, and ingested in Europeana at the end of each particular reporting period, while the percentage of the respective numbers will be automatically calculated in the columns **J, L, N, and P**.

As the workpackage leaders have discovered a sometimes quite high difference between the numbers of items already digitised and the numbers of items actually ready for harvesting, the project management decided to add a column "ready for harvest" to the progress reporting template. The information provided in this column (metadata and digitised objects available) is intended to improve the planning of the data ingestion processes to Europeana.

The collection of information about controlled vocabulary and metadata was extended with columns **Q to U**.

The next **columns (V to Z)** are provided for indicating '**Format and Quality**' of the master and access copies and for giving information about plans for **OCR**.

The partners will be stating the metadata formats of each subcollection in **column AA (Metadata)**.

As the number of scans might vary from the number of items (in case of newspapers, brochures, books, etc.) **column AB** is provided for specifying the definite number of scans.

3. Project Progress to Month 24

The following chapters will give an overview of the selection, digitisation, and ingestion progress of the project Europeana Collections 1914-1918 after the second year (April 2013). All partners have continued their work started in year one and progressed reasonably in selecting and digitising their objects. This has happened according to the DoW either by subcontracting to external partners or by digitising (see Deliverable D3.1). Most of the partners have continued or were able to start ingesting data in Europeana during the second project year.

3.1. Selection Progress

By April 2013, all partners have successfully carried out their **selection of items** to be digitised within the project. According to the DoW, 200.000 items should have been selected after the second project year (April 2013). **As 2012, the project partners have been more ambitious with their selection procedures than in the initial plan 429.057 items in total have already been chosen by month 24.**

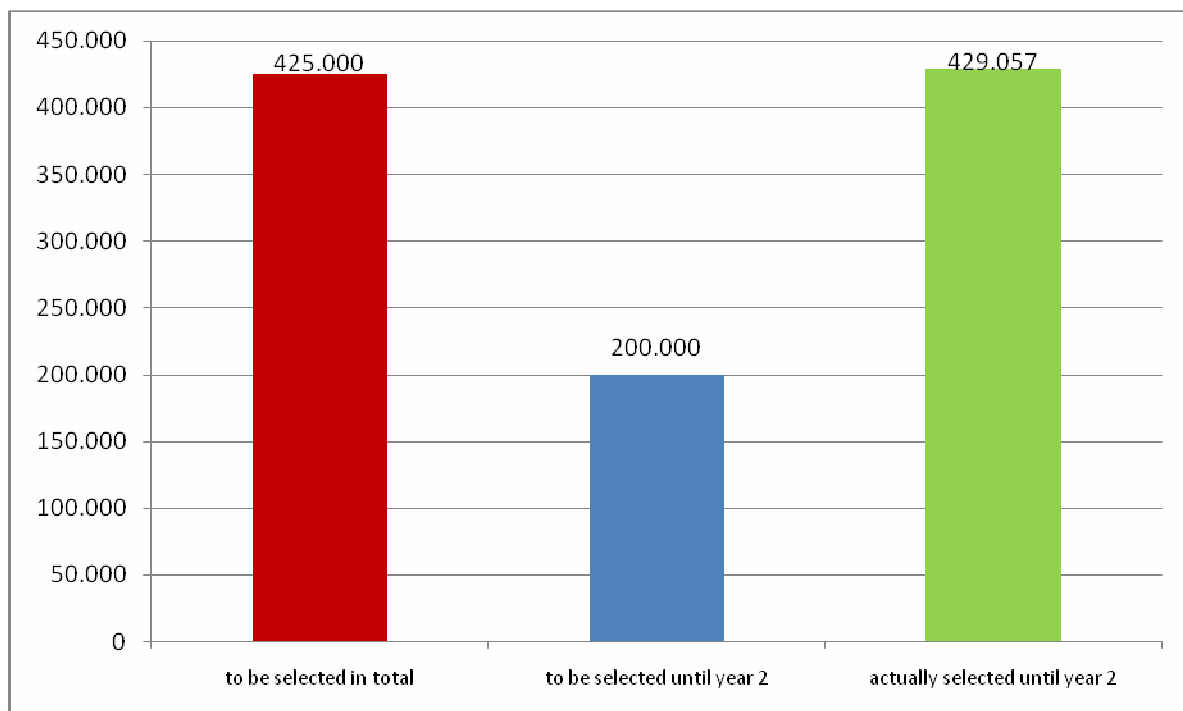


Figure 1: selection plan and progress until year 2

These 429.057 objects represent 99% of the total amount of items to be selected within the entire project and 101% of what the partners agreed on selecting according to the DoW. **As some partners have already indicated that they will select more items than initially planned, it is very likely that the project goal according to the DoW will be exceeded.**

The next diagram shows the selection effort of each project partner. The partners have selected the following amount of objects:

- SBB 6.892items
- BNF 15.037items
- BNU 45.633items
- BNCRM 5.461items
- BNCF 3.258items
- KBR 8.245items
- KB 8.764items
- BL 8.734items
- ONB 77.763items
- NLS 1.171items
- ICCU 248.099items.

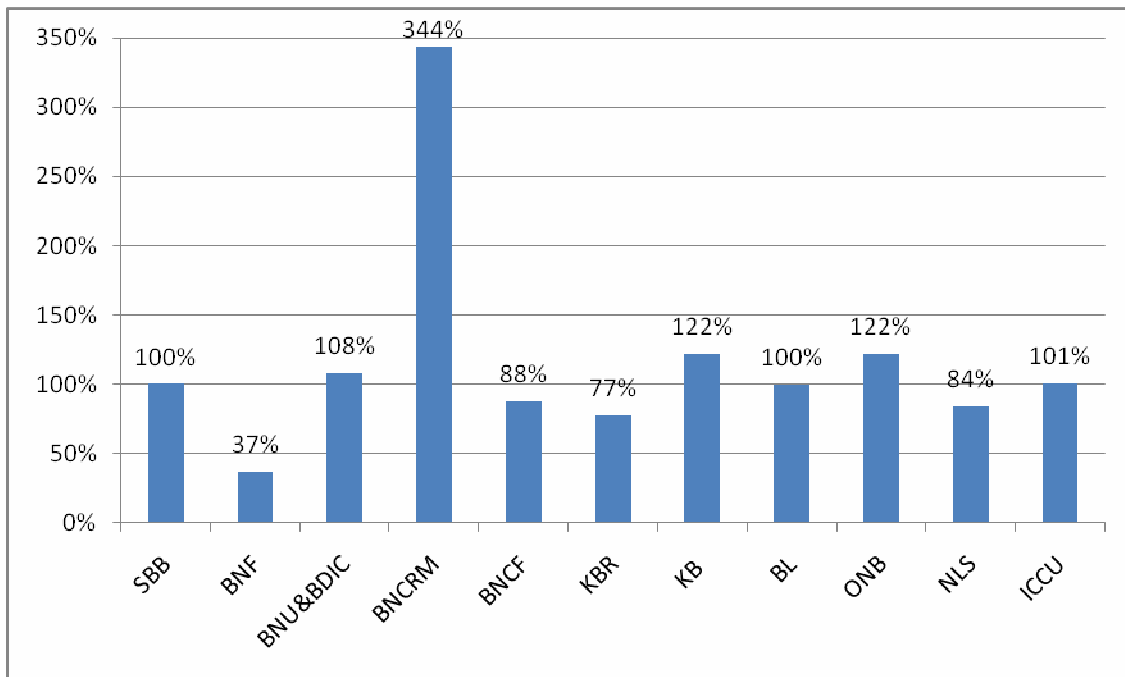


Figure 2: selection effort until year 2

Explanatory notes:

With 344% selected items, the selection number of BNCRM is significantly higher than the average selection numbers of the other project partners. During the selection process BNCRM discovered even more material to contribute to Europeana Collections 1914-1918 than initially indicated in the DoW.

BNF is currently in a process of re-evaluating and re-structuring their indicated content list, for this reason their number on selected material is currently lower than the average numbers of the other project partners.

The following graph gives an overview of the amount of items selected for each object group by all partners in total.

- Printed books: 11.511 items
- Songbooks: 3.495 items
- Publicistic Material 115.710 items
- Manuscript material: 128.508 items
- Picture material: 155.301 items
- Maps: 1.123 items
- Miscellaneous material: 13.409 items.

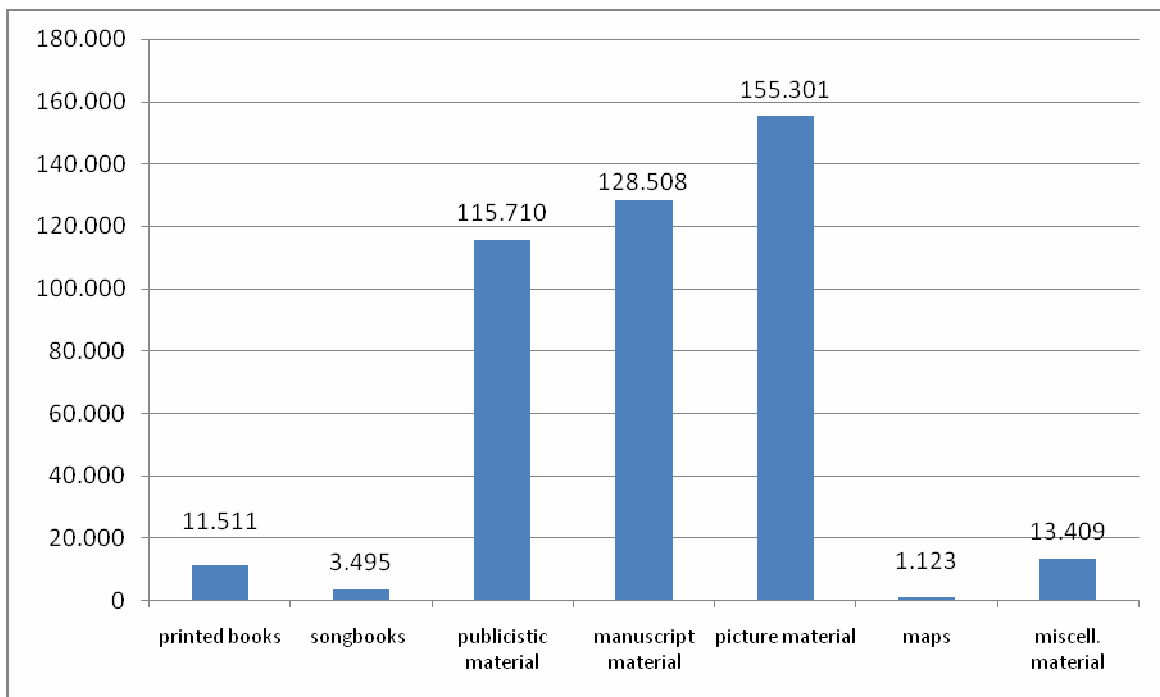


Figure 3: material selected until year 2

3.2. Digitisation Progress

By April 2013, several **partners have successfully digitised further collections of objects**. According to the DoW, 150.000 items should have been digitised after the second project year (April 2013). Again, the project partners have been more ambitious with the digitisation than intended. 267.195 items in total have already been digitised by month 24.

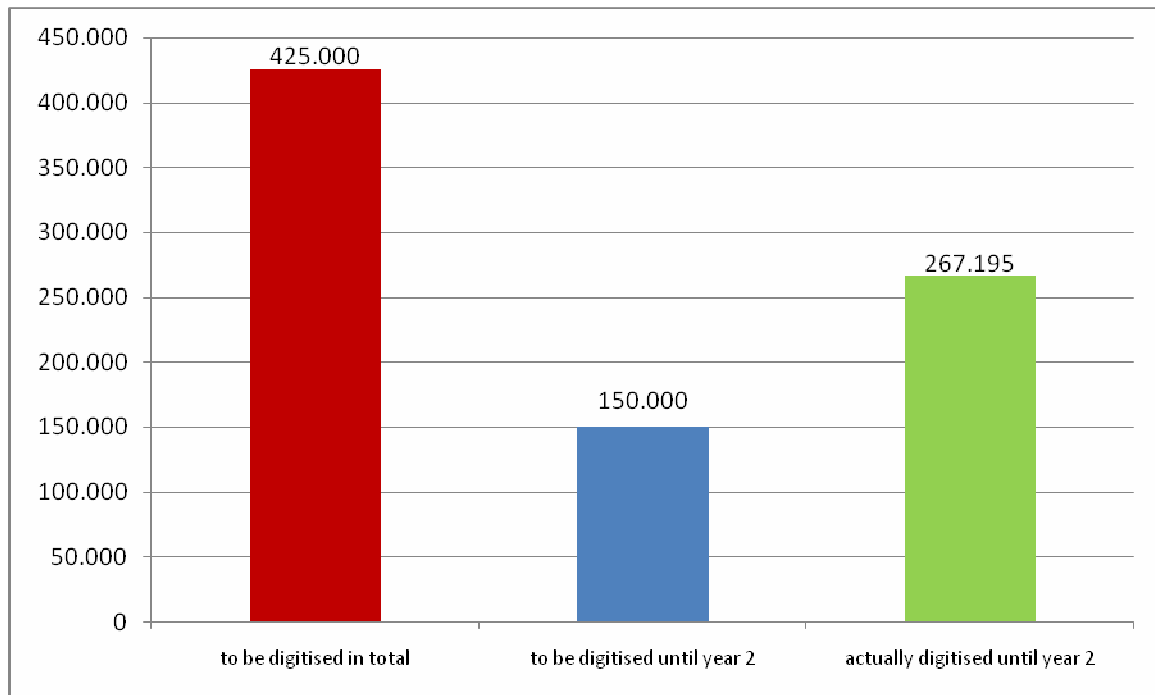


Figure 4: digitisation plan and progress until year 2

These 267.195 objects represent 62% of the total amount of items to be digitised within the project. **As some partners have already indicated that they will digitise more items than initially planned, it is very likely that the project goal according to the DoW will be exceeded.**

The next diagram shows the digitisation effort of each project partner. The partners have digitised the following amount of objects:

- SBB 3.452 items
- BNF 2.686 items
- BNU 18.909 items
- BNCRM 3.287 items
- BNCF 1.978 items
- KBR 3.573 items
- KB 8.764 items
- BL 3.427 items
- ONB 76.646 items
- NLS 677 items
- ICCU 143.796 items.

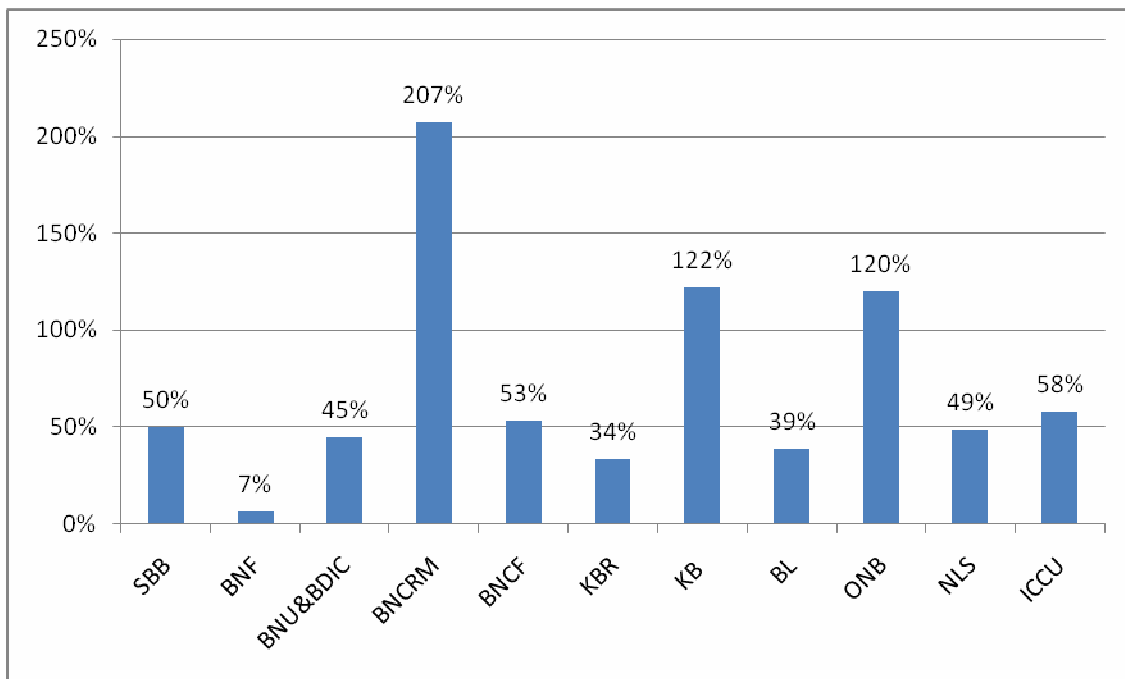


Figure 5: digitisation effort until year 2

Explanatory notes:

BNCRM high digitisation number is a result of their selection process. As they were able to select more items for Europeana Collections 1914-1918, they are able to digitise more material as well.

As indicated above, BNF is currently in a process of re-structuring their content list, so their numbers on digitised material is lower compared to the other project partners' average numbers.

The following graph gives an overview of the amount of items already digitised for each object group by all partners in total:

- Printed books: 8.944 items
- Songbooks: 1.836 items
- Publicistic Material 102.071 items
- Manuscript material: 50.931 items
- Picture material: 101.090 items
- Maps: 585 items
- Miscellaneous material: 1.738 items.

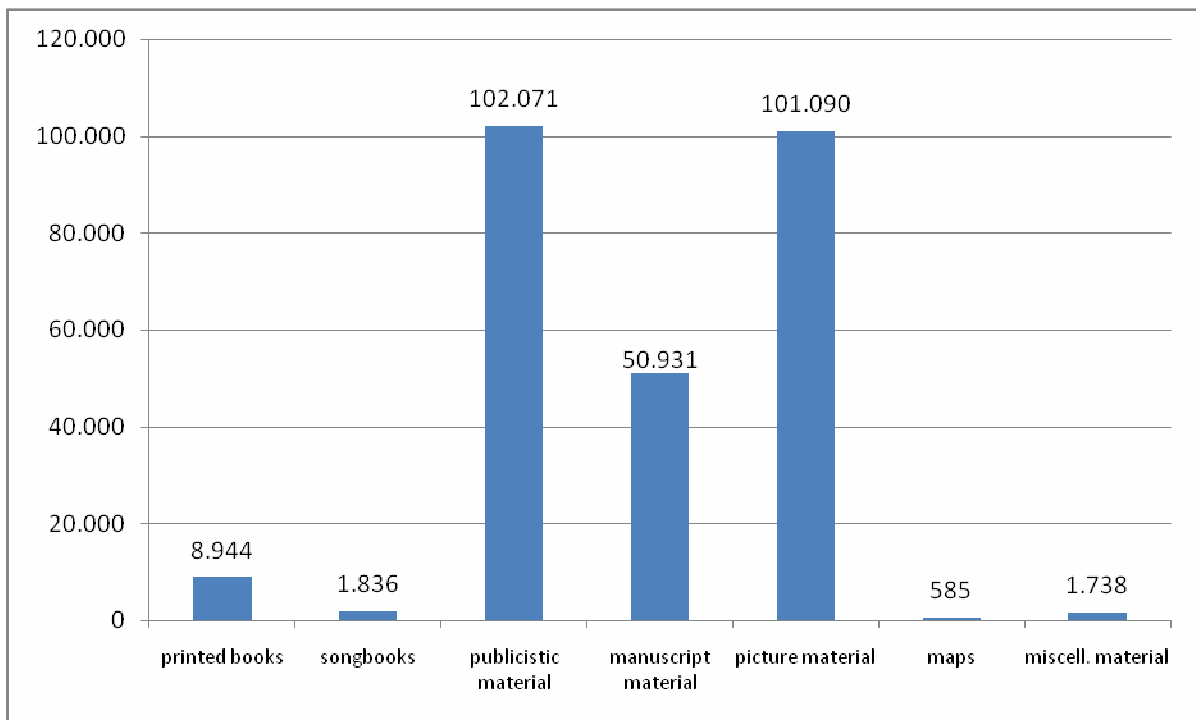


Figure 6: material digitised until year 2

3.3. Ingestion Progress

By April 2013, most of the partners have successfully ingested several collections of objects in Europeana. According to the DoW, 100.000 items should have been ingested after the second project year (April 2013). The project partners have been ambitious with the ingestion here as well. 111.588 items in total have already been ingested by month 24.

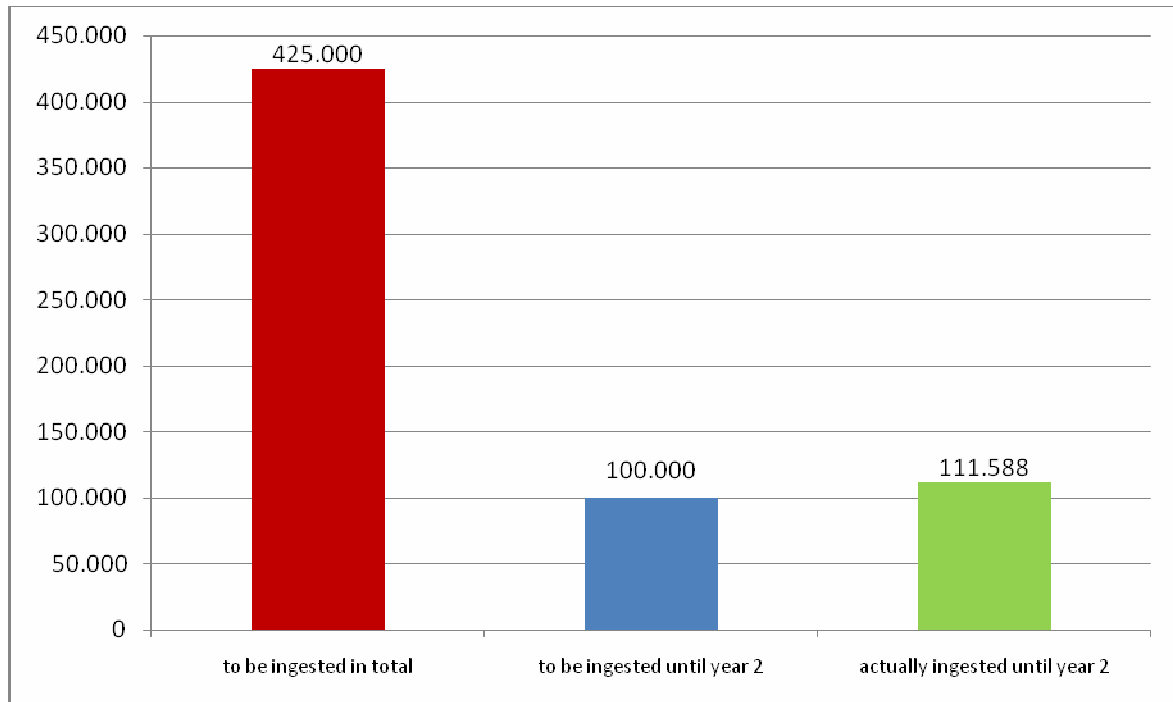


Figure 7: ingestion plan and progress until year 2

These 111.588 objects represent 26% of the total amount of items to be ingested within the project. **As some partners have already indicated that they will digitise more items than initially planned, it is very likely that the project goal according to the DoW will be exceeded.**

The next diagram shows the ingestion effort of the project partners.

- SBB 3.165 items
- BNCRM 3.287 items
- BNCF 1.978 items
- KBR 406 items
- KB 5.629 items
- BL 2.150 items
- ONB 28.282 items
- NLS 378 items
- ICCU 66.313 items.

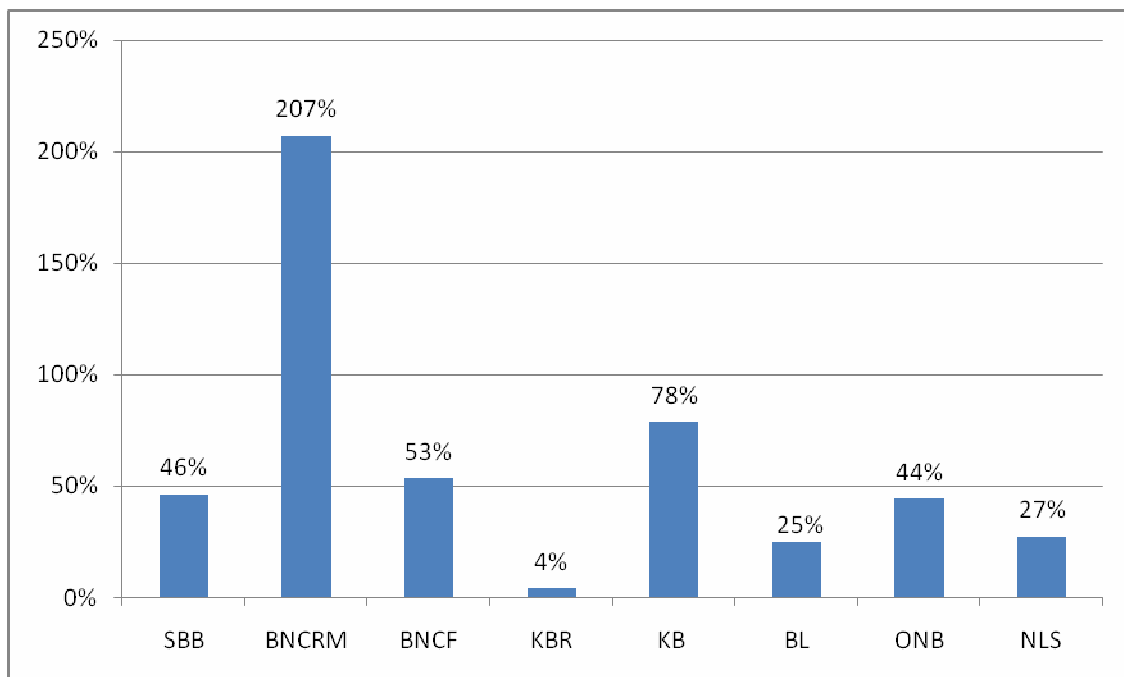


Figure 8: ingestion effort until year 2

Explanatory notes:

Due to content re-structuring processes at BNF, there are no numbers on ingested material of BNF in Europeana, even if one collection has already been ingested to Europeana.

BNU will publish a new digital library portal at the end of 2013. During this process their digitised material for Europeana Collections 1914-1918 will be able to be ingested in Europeana.

As BNCRM is able to provide more material than initially intended, they are able to ingest more items as well.

The following graph gives an overview of the amount of items already ingested for each object group by the concerning partners in total:

- Printed books: 5.525items
- Songbooks: 1.768items
- Publicistic Material 64.455items
- Manuscript material: 650 items
- Picture material: 38.805items
- Maps: 238 items
- Miscellaneous material: 147 items.

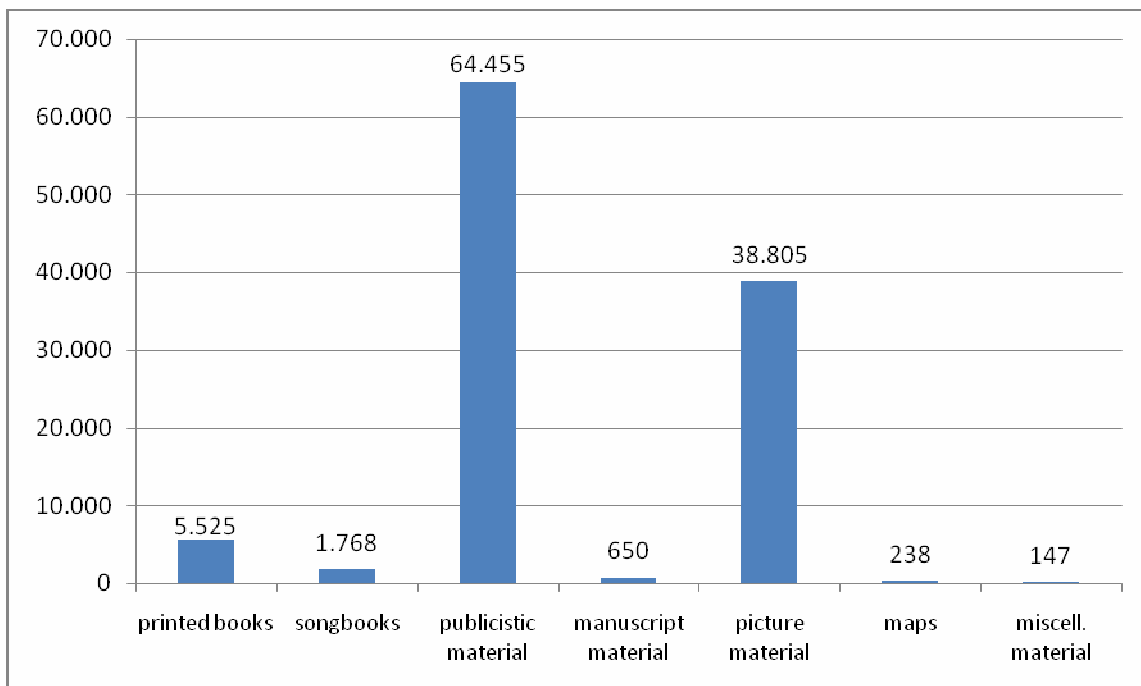


Figure 9: material ingested until year 2

4. Summary

To summarize the selection, digitisation, and ingestion efforts of the project Europeana Collections 1914-1918 after the second project year it can be said that all project goals were successfully overfulfilled (**selection: 429.057 instead of 200.000 items; digitisation 273.615 instead of 150.000 items; ingestion: 111.588 instead of 100.000 items**). From a global point of view, the selection of items has been finished by exceeding the 100% mark; however, few partners still have to select minor amounts of objects to their selections.

The cooperation and communication within the project consortium was very fruitful and supportive. All partners developed digitisation plans for their institutions and provided the workpackage leaders and project management with essential information for the reporting in time. All partner institutions have also developed and initiated their digitisation workflows for the third project year.

It turns out after 24 months of the project that the process from selecting to digitising to metadata generation to ingesting items is more an iterative than a parallel process. While the results show that selection is much ahead of the schedule and digitisation is still reasonably ahead of the schedule, the ingestion progress is more or less exactly where it should be according to the project plan.

The high achievement rates in selection show, that the partners concentrated on the selection of items in the beginning of the project. Digitisation could not start immediately due to reasons that digitisation workflows still had to be defined and subcontractors had to be found. The ingestion (and the generation of metadata), which is a later step in the process chain, went into the focus of activities in year II of the project and is taking up speed now.

This is also represented well in the curves of the diagram below. While the blue curve representing “selection” ascends abruptly from the very beginning of the project, the red and the green curve representing “digitisation” and “ingestion” start to ascend flat-angled and catch up later.

The last graph summarizes the progress of the project Europeana Collections 1914-1918 until the end of the second project year (April 2013).

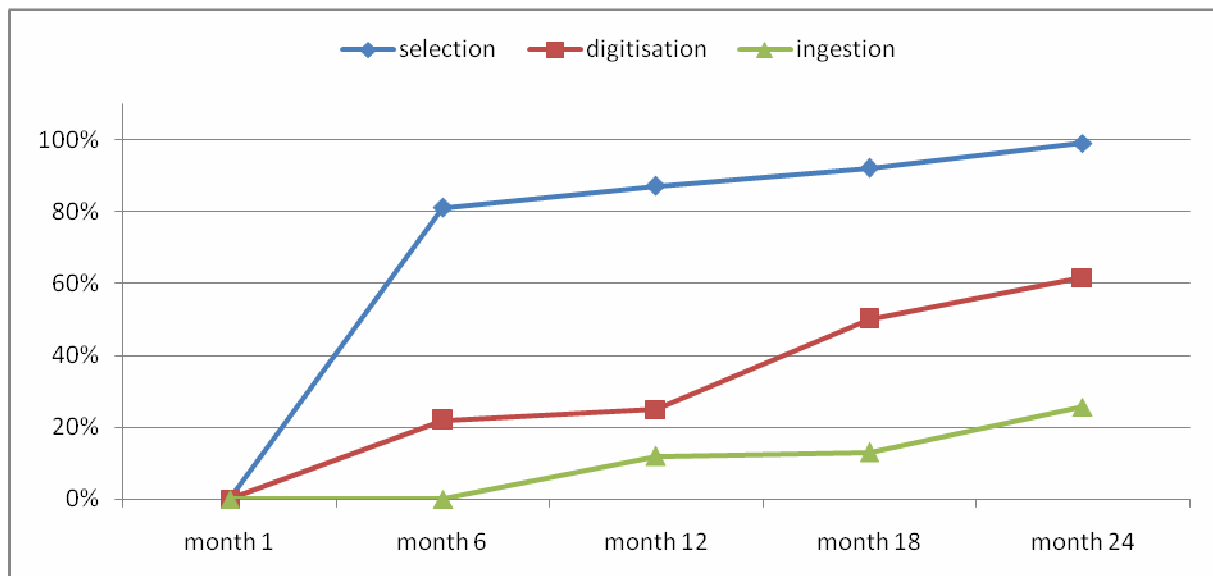


Figure 10: project progress until year 2

5. Annex I – ‘Digitisation progress monitoring template’

Screenshot from the ‘Digitisation progress monitoring template’ – Summary page from April 30th, 2013.

EditGrid Welcome, ec1418
 Preferences | Logout [User Mode]

Spreadsheet / ec1418 / EC1418 Digitisation Progress Template

File Edit View Format Insert Data Share Publish Collaborate Macro Help

Arial 10 pt B I U

	A	B	C	D Selected		E Digitised		F Ready for Harvest		G Ingested in Europeana	
				amount	percentage	amount	percentage	amount	percentage	amount	percentage
1		Sum: Proposed Source Quantity According to Description of Work									
2											
3	SBB	6,880	6,880	6,892	100%	3,452	50%	3,165	46%	3,165	46%
4	BNF	40,540	40,540	15,037	37%	2,686	7%	2,686	7%	0	0%
5	BNU & BDIC	42,363	42,363	45,633	108%	18,909	45%	0	0%	0	0%
6	BNCRM	1,587	1,587	5,461	344%	3,287	207%	3,287	207%	3,287	207%
7	BNCF	3,702	3,702	3,258	88%	1,978	53%	1,978	53%	1,978	53%
8	KBR	10,658	10,658	8,245	77%	3,611	34%	1,903	18%	406	4%
9	KB	7,172	7,172	8,764	122%	8,764	122%	5,629	78%	5,629	78%
10	BL	8,758	8,758	8,734	100%	3,427	39%	2,166	25%	2,150	25%
11	ONB	63,933	63,933	77,763	122%	76,646	120%	28,282	44%	28,282	44%
12	NLS	1,395	1,395	1,171	84%	677	49%	378	27%	378	27%
13	ICCU	246,600	246,600	248,099	101%	143,796	58%	98,515	40%	66,313	27%
14		433,588	433,588	429,057	99%	267,233	62%	142,138	33%	111,588	26%