DELIVERABLE

Project Acronym: Europeana Collections 1914-1918

Grant Agreement number: 270894

Project Title: Europeana Collections 1914-1918: Remembering the First World War – a digital collection of outstanding sources from European national libraries

_____

D 3.3   First report on numbers of items digitised in the project

_____

Revision: 1.0

Authors: Ellen Geisriegler, Austrian National Library, 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
1	11.06.2011	Thorsten Siegmann	SPK	Correction of a spelling error on page 6

Table of Contents

1. Executive Summary

This Deliverable contains the 'First report on numbers of items digitised in the project' of the project Europeana Collections 1914-1918 and provides an overview of the number of objects selected, digitised, and ingested in Europeana within the first project year.

The information in it was compiled from all partner institutions with the help of the 'Digitisation progress monitoring templates' in month 6 (October 2011) and month 12 (April 2012) of the project year.

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 project internal reporting (*MS10 and MS 16: Statistical report on numbers of items selected/digitised to month 6*). A modified blank template can be found as an appendix to this document. To further simplify the documentation and data collection processes for the project partners it was agreed to use the collective workspace platform EditGrid for this purpose instead of updating Excel sheets.

In project month 12 (April 2012), 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.

No.	Short name	Name	Country
1	SBB/SPK	Staatsbibliothek zu Berlin/ Stiftung Preußischer Kulturbesitz / Berlin State Library – Prussian Cultural Heritage Foundation	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 Belgie/ Royal Library of Belgium	Belgium
7	BL	The British Library	United Kingdom
8	KB	Det Kongelige Bibliotek, Nationalbibliotek og Kobenhavns Universitetbibliotek / 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 does not provide content

2. Data Collection and Progress Reporting

In the DOW the project consortium has indicated that 425.000 sources from World War One 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 process, the partners have agreed on the Monitoring plan `Selection/ Digitisation/ Ingested 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 D.3.1 'Synthesis of agreed digitisation plans'. For the compilation 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's progress, a few more columns have been added to the monitoring template within the last months. These changes shall be described below.

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 EditGrid[1]. 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 (unified) sheet providing an overview on each partner's progress (see tabs at the bottom). The last sheet of the template gives a summary of the numbers of all content providers.

Column A (Provider) and B (Last date of reporting) contain the abbreviation of the content provider and the latest date of reporting. For D 3.3 the last day of the reporting period was 15 April 2012.

---

[1] https://www.editgrid.com

As already described in Deliverable 3.1, 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; and Misc. The source types shall be indicated in **Column C (Source type).**

All partners are asked to fill in the total number of items foreseen for digitisation split up by source type according to the DOW into **column D (Proposed source type according to DOW).**

For organisational reasons, most partner institutions further 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: Misc.
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 as a result of the selection process. 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 to be added to the selection. For that reason, the effective number of items by subcollection varies and will be reevaluated every six months to reflect the exact progress of the project.

(Update!) Columns I to P are meant as <u>**Production Monitoring Summary,**</u> 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 discovered some cases of relevant deviations between the indicated number of items already digitised and the number 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 (ready for harvest means metadata **and** digitised object available) is intended to improve the planning of the data ingestion processes to Europeana.

(Update!) The information collection about controlled vocabulary and metadata has been extended. **In columns Q-U** the partners are asked to indicate, whether they are planning to use any controlled vocabulary to provide information about the object on a certain level. This refers to the support of any kind of normative data in the metadata, but especially to the minimal vocabulary developed in the Europeana Collections 1914-1918 project:
- What? EC1418 subject headings ,- What? EC1418 object types, - Who? (e.g. VIAF, PND or else), - Where? (e.g. geonames), - When? (in general information on time).

The next **columns (V to Z)** are provided for indicating '<u>**Format and Quality'**</u> 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 12

The following chapter provides an overview on the status of the selection, digitisation, and ingestion progress of the project Europeana Collections 1914-1918 after the first year (April 2012). All partners have started selecting objects and several have delivered objects to external partners for digitisation or have started their digitisation workflows in-house (see Deliverable 3.1). Three partners were able to ingest data in Europeana within the first project year. The concrete numbers are presented below.

3.1 Selection Progress

By April 2012, all partners have successfully carried out a first selection of items to be digitised within the project. According to the DOW, 100.000 items should have been selected after the first project year (April 2012). 371.556 items in total have already been selected for digitisation in the project by month 12. The project partners have been more ambitious with their selection procedures than the initial plan though.
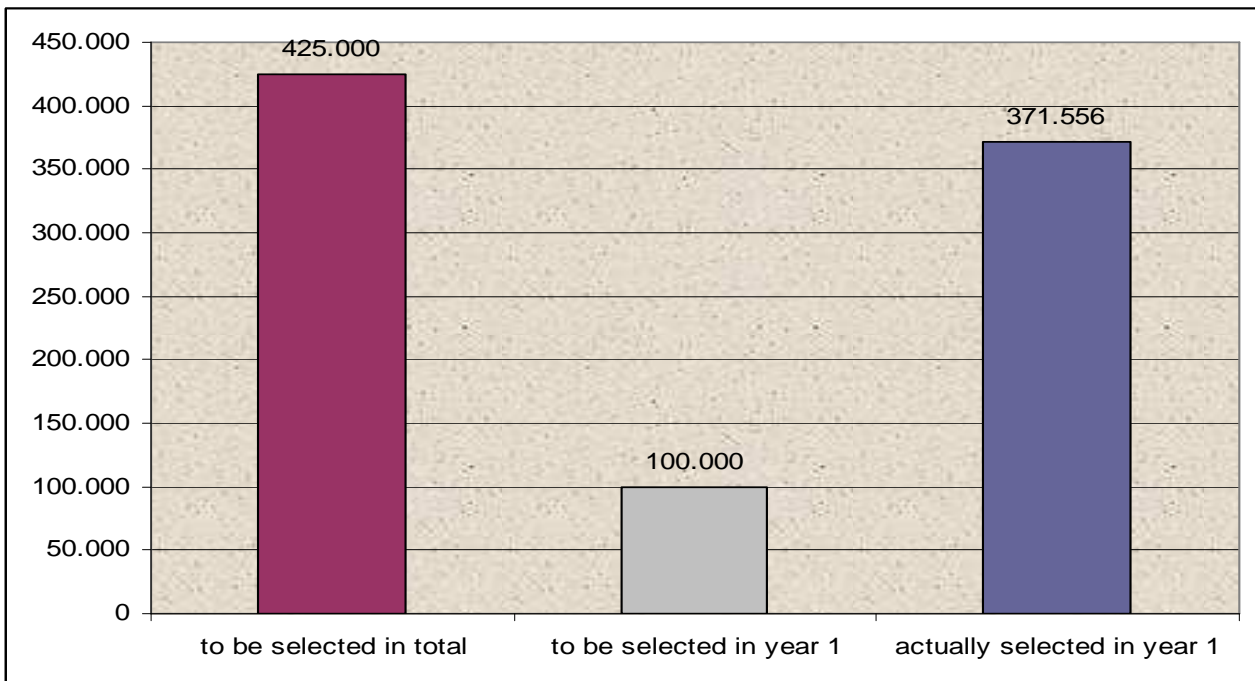
fig.1: Selection plan & progress year 1

These 371.556 objects represent 87% of the total amount of items to be selected within the entire project. 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 progress of each project partner. The partners have selected the following amount of objects until April 2012:
SBB: 2.806 items, BNF: 221 items, BNU: 43.133 items, BNCRM: 3.777 items, BNCF: 3.702 items, KBR: 2.696 items, KB: 6.976 items, BL: 10.409 items, ONB: 52.162 items, NLS: 258 items, ICCU: 246.000 items.
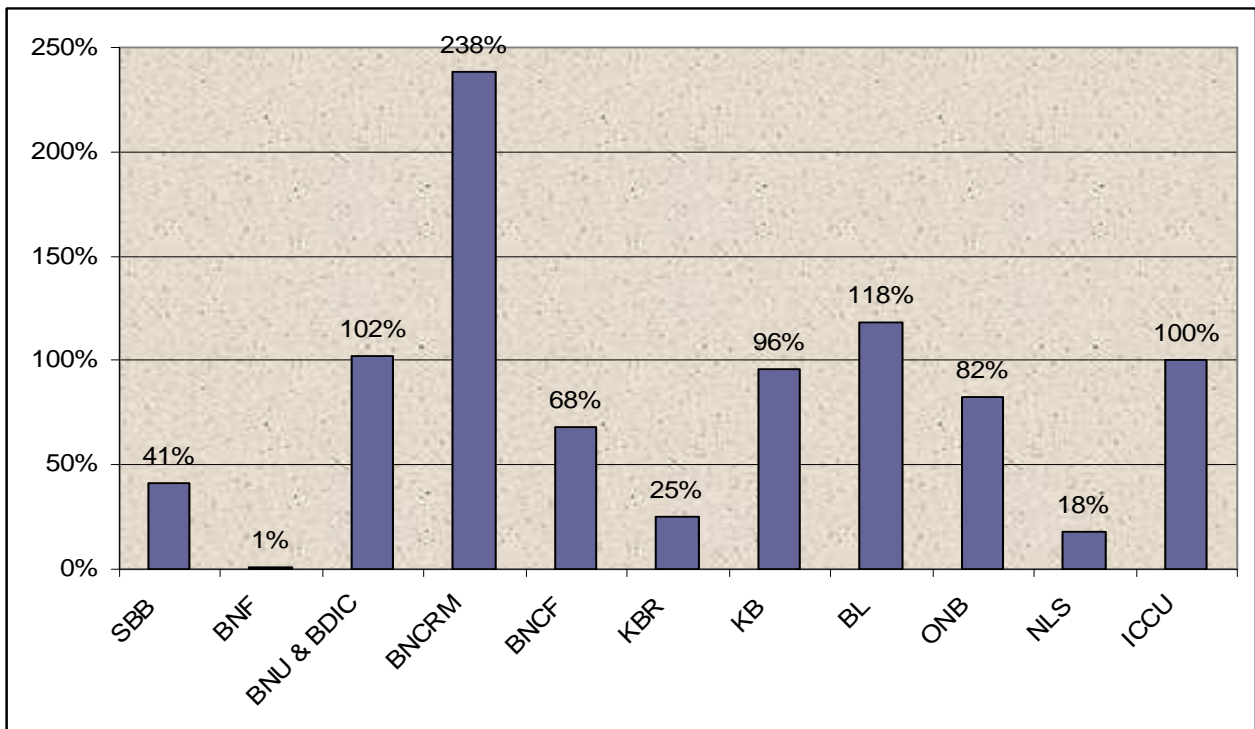

fig 2: Selection effort in year 1

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

Books: 5.316 items; Songbooks: 2.628 items; Publicistic Material 51.383 items; Manuscript material: 126.438 items; Picture material: 171.573 items; Maps: 1.371 items; Misc: 12.847 items.
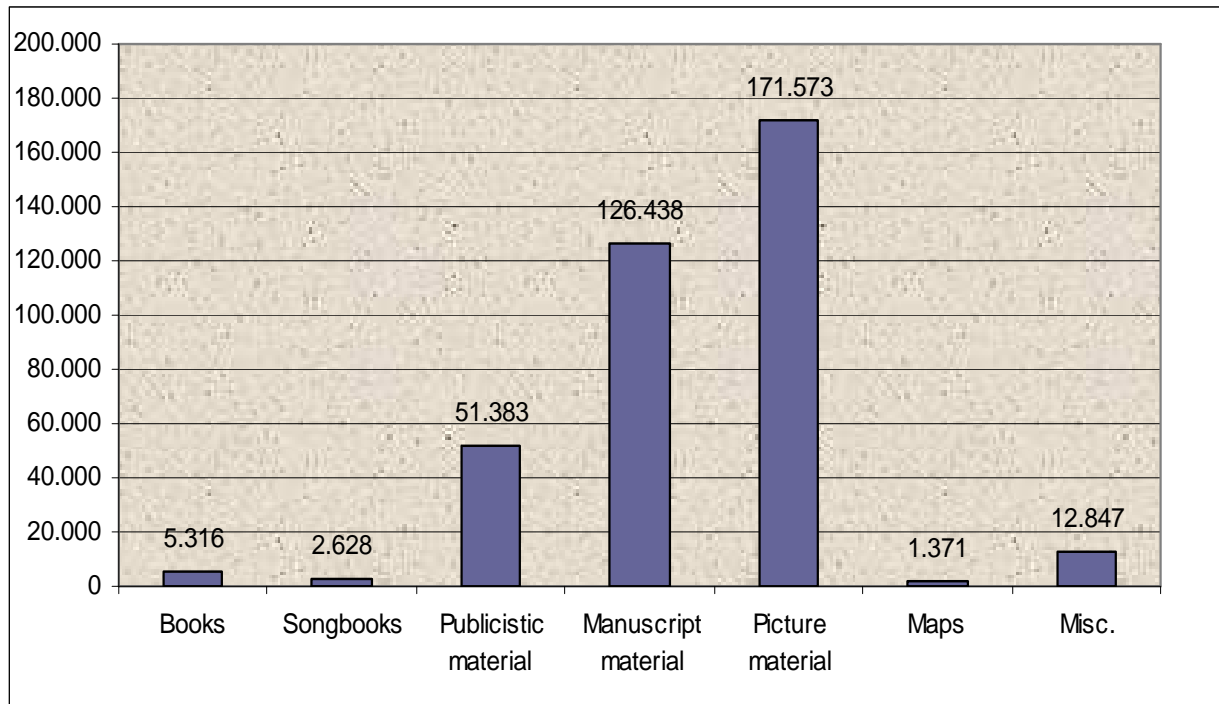
fig.3: Material selected in year 1

3.2 Digitisation Progress

By April 2012, several partners have also successfully digitised first collections of objects. According to the DOW, 50.000 items should have been digitised after the first project year (April 2012). 107.455 items in total have already been digitised by month 12. The project partners have been more ambitious with the digitisation than the initial plan though.
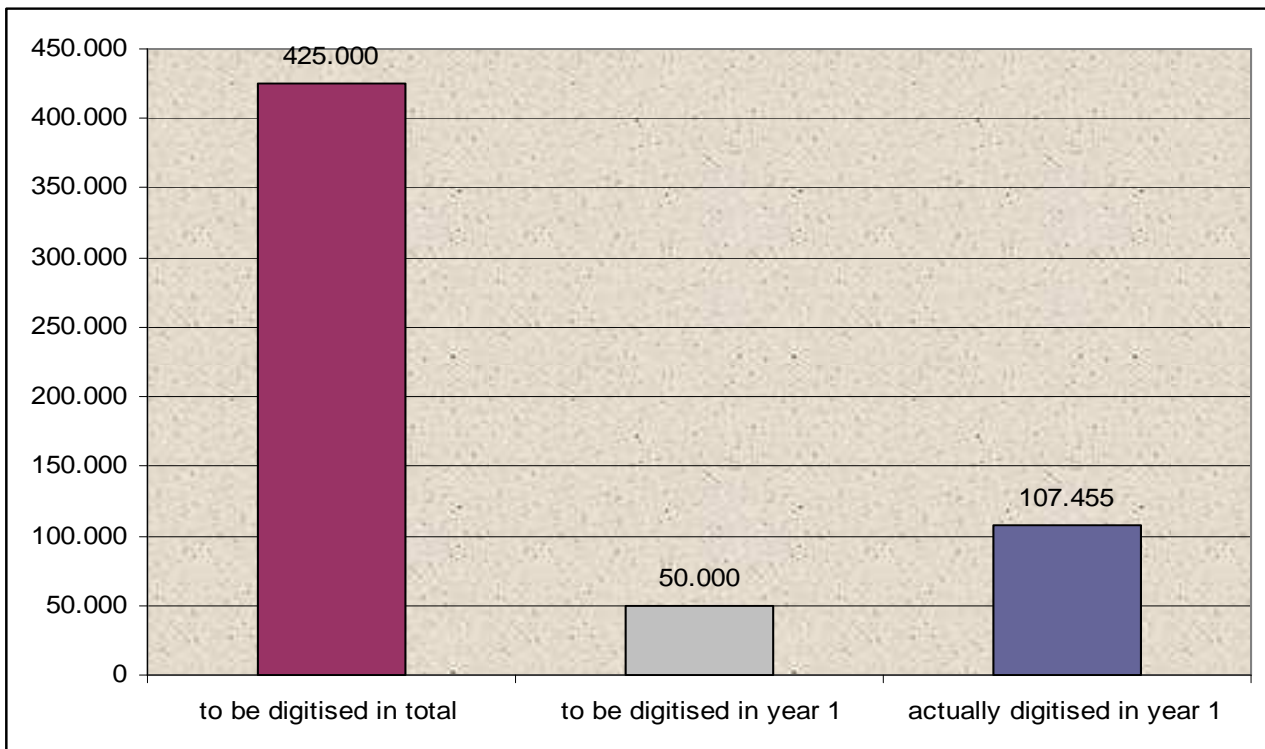
fig. 4: Digitisation plan & progress year 1

These 107.455 objects represent 25% 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 results of each project partner. The partners have digitised the following amount of objects:
SBB: 472 items, BNF: 120 items, BNU & BDIC: 14.494 items, BNCRM: 90 items,
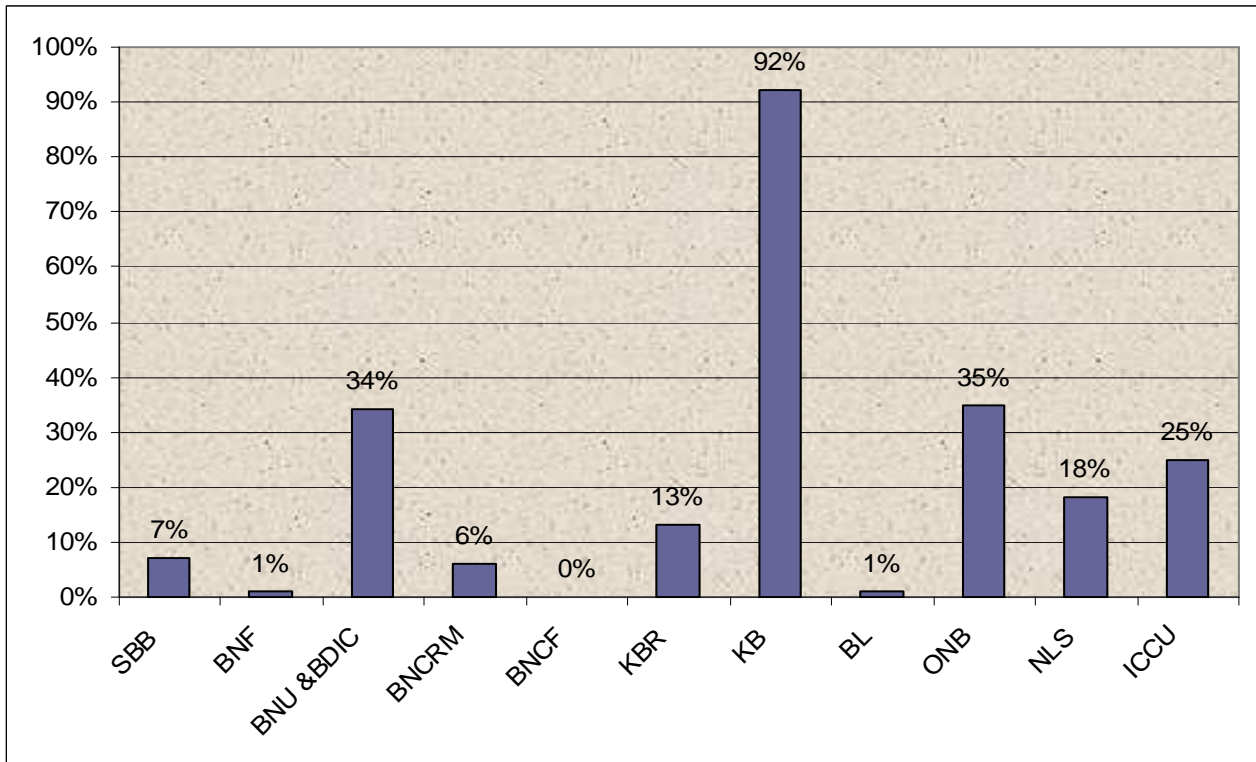KBR: 1.351 items, KB: 6.589 items, BL: 89 items, ONB: 22.401 items, NLS: 258 items, ICCU: 61.600 items.

fig. 5: Digitisation results in year 1

The following graph gives an overview of the amount of items already digitised for each object group by all partners in total: Books: 1.062 items; Songbooks: 242 items; Publicistic Material: 37.491 items; Manuscript material: 38 items; Picture material: 67.666 items; Maps: 389 items; Misc: 567 items.
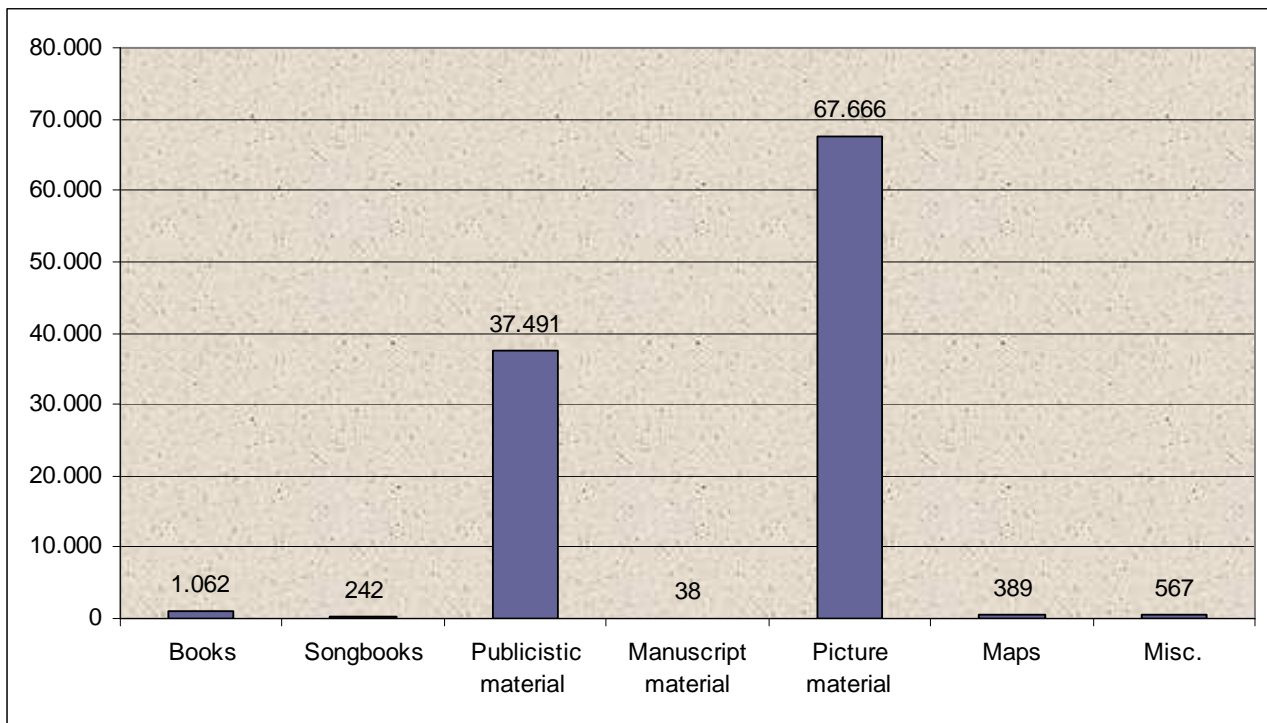

fig. 6: Material digitised in year 1

3.3 Ingestion Progress

By April 2012, three partners have successfully ingested first collections of objects in Europeana. According to the DOW, 10.000 items should have been ingested after the first project year (April 2012). 52.320 items in total have already been ingested by month 12. The project partners have been more ambitious with the ingestion than the initial plan though.
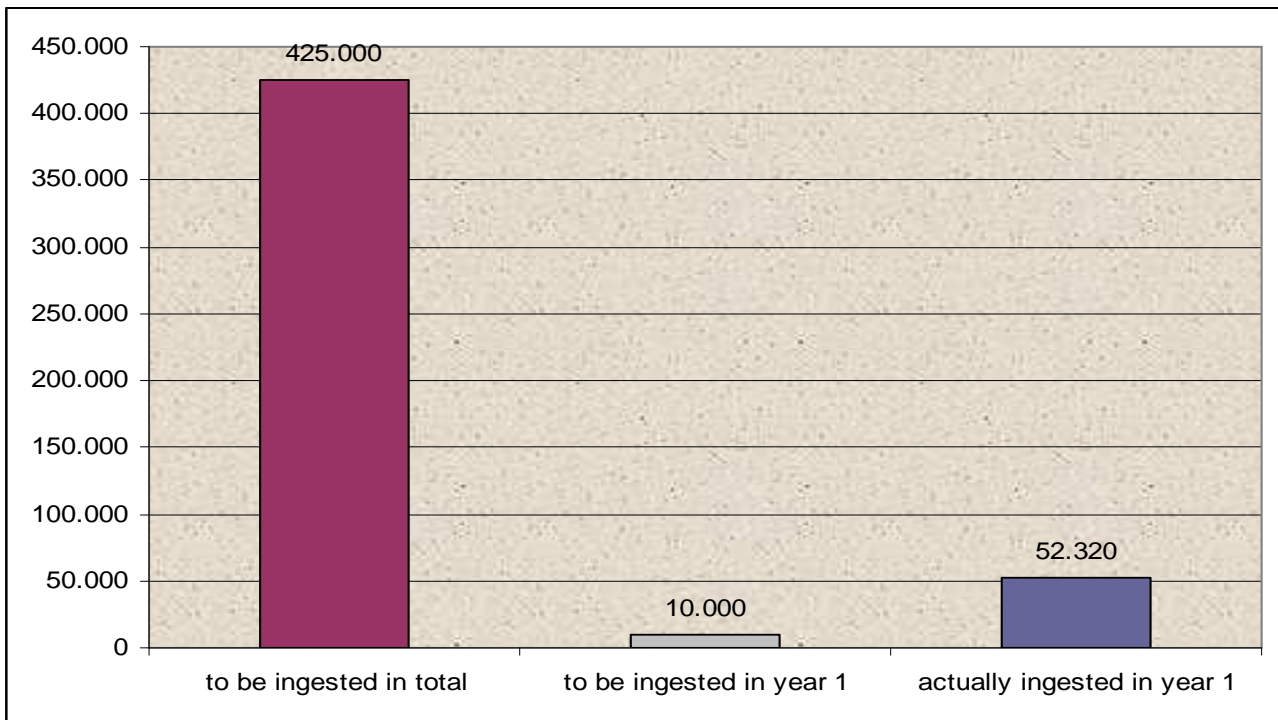

fig. 7: Ingestion plan & progress year 1

These 52.318 objects represent 12% 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 seems to be possible that the initial project goal according to the DOW will be exceeded.

The next diagram shows the ingestion results of each project partner. The partners ONB, NLS, and ICCU have ingested the following amount of objects:
ONB: 13.858 item, NLS: 86 items, ICCU: 38.376 items.
Within the next year, further partner will be harvested by TEL one by one. Currently the partners are updating the ingestion plans to define the ingestion schedule for the next 12 month together with TEL.
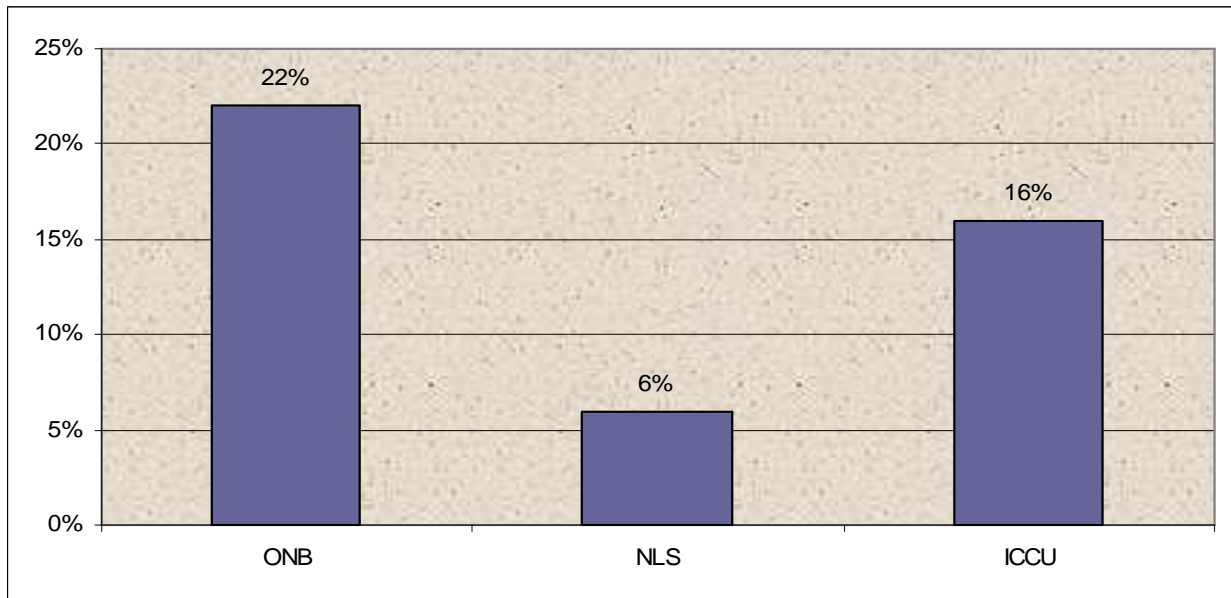

fig. 8: Ingestion effort in year 1

The following graph gives an overview of the amount of items already ingested for several object groups by ONB, NLS, and ICCU in total.

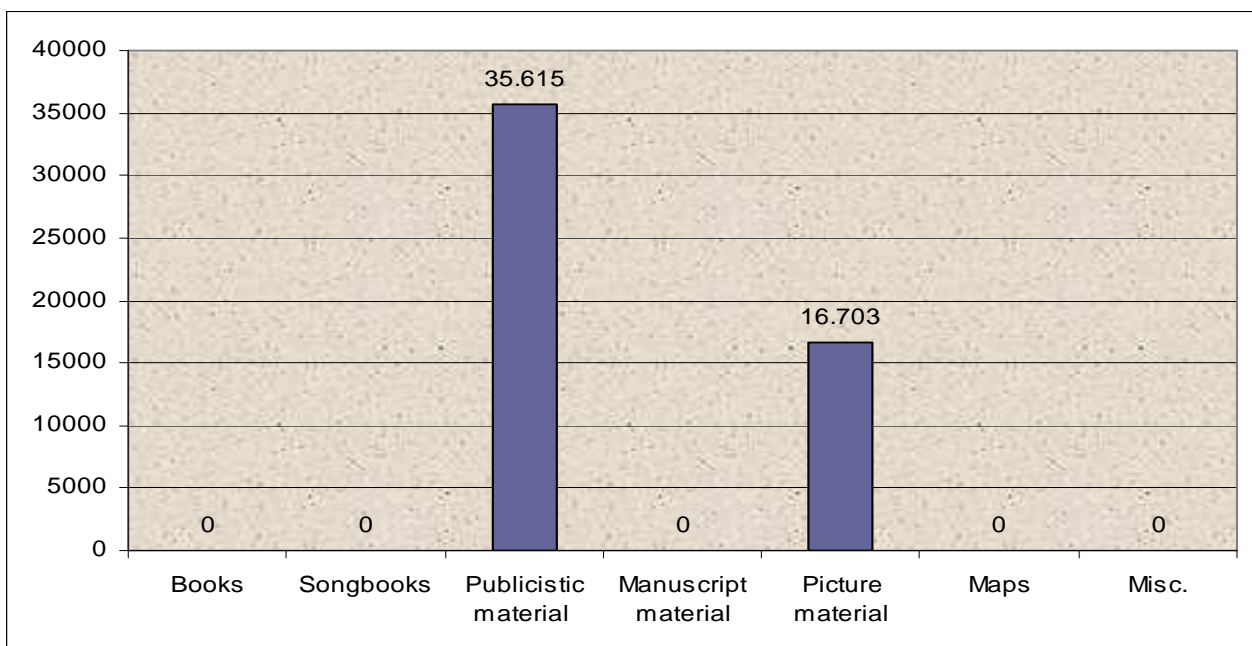Publicistic Material: 35.615; Picture material: 16.703 items

fig. 9: Material ingested in year 1

4. Summary

To summarize the selection, digitisation, and ingestion efforts of the project Europeana Collections 1914-1918 after the first year it can be said that all project goals were successfully exceeded.

The Europeana Collections 1914-1918 partners

- selected 371.556 instead of 100.000 items

- digitised 107.455 instead of 50.000 items

- ingested 52.320 instead of 10.000 items.

The co-operation and communication within the project consortium was very fruitful and supportive. All partners developed digitisation plans for their institutions and provided the WP 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 second project year.

By April 2012 three partners were already able to deliver data to Europeana. Within the next months several partners will follow. The European Library who acts as aggregator for the project is only capable of processing one partner after another. A detailed ingestion plan (MS21 Timetable for metadata ingestions) is currently developed in co-operation with the project management and TEL. The idea of this ingestion plan is to gain an overview of the data delivery plans of all partners and to manage the ingestion procedures with TEL as efficient as possible.

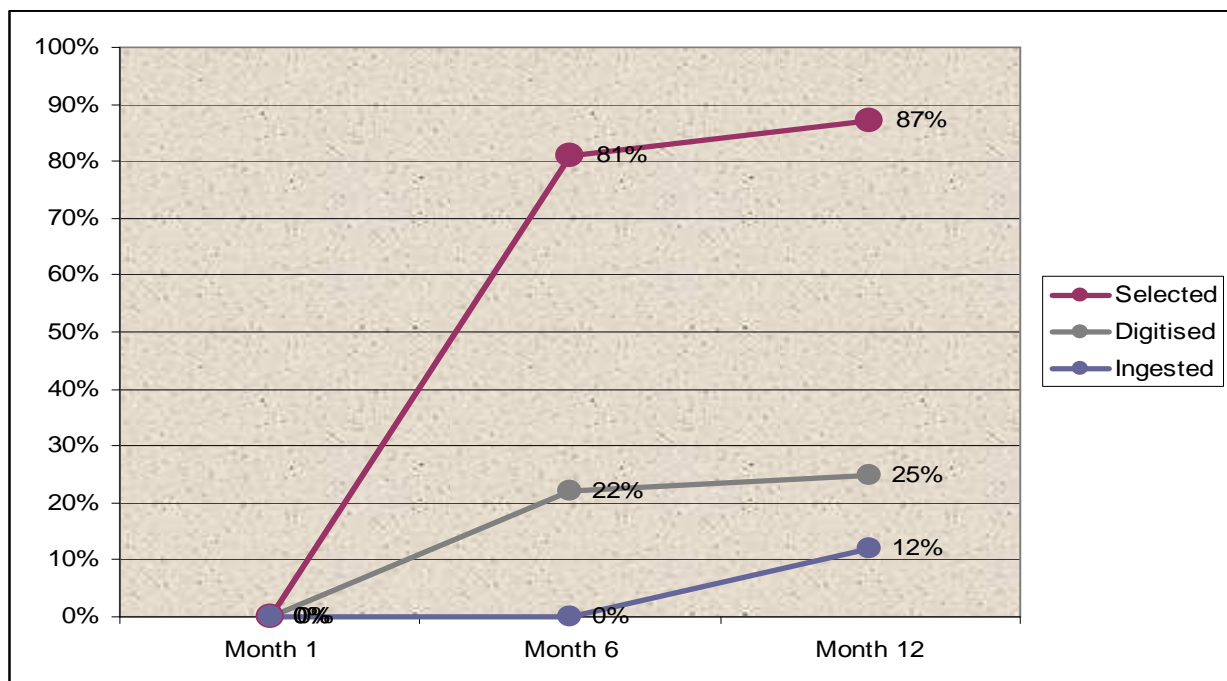The last graph summarizes the progress of the project Europeana Collections 1914-1918 after the first year (April 2012).

fig. 10: Project progress year 1