



DELIVERABLE

Project Acronym: DM2E

Grant Agreement number: ICT-PSP-297274

Project Title: Digitised Manuscripts to Europeana

D1.1 - Requirements Report

Revision: 1.0

Authors:

Ewelina Rockenbauer (ONB); Alois Pichler (UIB); Jorge Urzúa (MPIWG); Klaus Thoden (MPIWG); Dirk Wintergrün (MPIWG); Gerhard Müller (SBB); Evelyn Dröge (UBER); Julia Iwanowa (UBER); Steffen Hennicke (UBER); Christian Morbidoni (Net7); Esther Guggenheim (NLI); Marko Knepper (UBFFM);

Contributors:

Jennifer Waxman (CJH); Laura Leone (CJH); Dov Winer (EACJ); Lena Stanley-Clamp (EAJC); Maria LaCalle (JDC); Paolo D'Iorio (CRNS); Kilian Schmidtner (SBB);

Project co-funded by the European Commission within the ICT Policy Support Programme		
Dissemination Level		
PU	Public	X

Revision history and statement of originality

Revision	Date	Author	Organisation	Description
Draft 0.1	21.05.2012	Ewelina Suchorzebska	ONB	
Draft 0.2	02.07.2012	Alois Pichler	ONB	Input from UIB
Draft 0.3	09.07.2012	Jorge Urzúa; Gerhard Müller	MPIWG; SBB	Input from MPIWG and SBB
Draft 0.4	10.07.2012	Ewelina Suchorzebska	ONB	Input from Evelyn Dröge (UBER); general additions
Draft 0.4.1	20.07.2012	Gerhard Müller	SBB	1 st proofreading and some additions
Final draft	25.07.2012	Ewelina Rockenbauer	ONB	Some additions
Final 1.0	31.07.2012	Ewelina Rockenbauer; Gerhard Müller	ONB; SBB	Some additions; Final revision
Final 1.0	09.08.2012	Stefan Gradmann	UBER	Approval Final 1.0

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This deliverable contains original unpublished work except where clearly indicated otherwise.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.

Contents

1 Executive Summary	6
2 Introduction	7
2.1 Scope and Objective of this document	7
2.2 The structure of this document.....	7
3 Overview of content and metadata standards in DM2E	8
3.1 Underlying content	8
3.2 Main results of the conducted questionnaires.....	10
4 Requirements	13
4.1 ONB – Austrian National Library (Österreichische Nationalbibliothek)	13
4.1.1 Underlying content	13
4.1.2 Requirements	13
4.1.3 Preliminary mapping of mandatory EDM fields	19
4.2 SBB – Berlin State Library (Staatsbibliothek zu Berlin)	20
4.2.1 Underlying content	20
4.2.2 Legacy Data formats.....	21
4.2.3 Requirements	28
4.2.4 Summary	29
4.2.5 Preliminary mapping of mandatory EDM fields	29
4.3 UIB – University of Bergen (Universitetet i Bergen)	31
4.3.1 Underlying content	31
4.3.2 Requirements	31
4.3.3 Preliminary mapping of mandatory EDM fields	34
4.4 UBER – Humboldt University Berlin (Humboldt Universität zu Berlin)	35
4.4.1 Underlying content	35
4.4.2 Requirements	35
4.4.3 Preliminary mapping of mandatory EDM fields	38
4.5 MPIWG – Max Planck Institute for the History of Science (Max-Planck-Institut für Wissenschaftsgeschichte)	39
4.5.1 General information	39
4.5.2 Extensions to EDM.....	39
4.5.3 The MPIWG's rare book collection	40
4.5.3.1 General remarks.....	40
4.5.3.2 Contents of the metadata	40
4.5.3.3 Preliminary mapping of mandatory EDM fields:	42
4.5.4 ISMI Project	43
4.5.4.1 Structure of the metadata.....	43
4.5.4.2 Preliminary mapping of mandatory EDM fields.....	43
4.5.4.3 Challenges with mapping to EDM	45

4.5.4.4	Conclusions and Requirements	47
4.6	NLI – National Library of Israel.....	47
4.6.1	Underlying content	47
4.6.2	Requirements	48
4.6.3	Preliminary mapping of mandatory EDM fields	51
4.7	CJH via EAJC – Center for Jewish History	52
4.7.1	Digital Collections.....	52
4.7.2	Preliminary mapping of mandatory EDM fields	53
4.8	BBAW – Berlin-Brandenburg Academy of Sciences (Berlin-Brandenburgische Akademie der Wissenschaften).....	54
4.8.1	Legacy Data format – Text Encoding Initiative (TEI)	54
4.8.2	Requirements	54
4.8.3	Preliminary mapping of mandatory EDM fields	55
4.9	CRNS – National Center for Scientific Research (Centre de la Recherche National Scientifique).....	56
4.9.1	Underlying content	56
4.9.2	Preliminary mapping of mandatory EDM fields	58
4.10	JDC via EAJC – Joint Distribution Committee	59
4.10.1	Underlying content	59
4.10.2	Preliminary mapping of mandatory EDM fields	64
4.11	UBFFM – University Library JCS Frankfurt am Main (Universitätsbibliothek Johann Christian Senckenberg Frankfurt am Main).....	65
4.11.1	Underlying content	65
4.11.2	Preliminary mapping of mandatory EDM fields	71
5	Requirements gathering WP3.....	73
6	Summary and WP1 Outlook.....	74
7	Appendix.....	75
7.1	Content questionnaire.....	75
7.2	Metadata questionnaire	79

List of Tables

Table 1: Underlying content	10
Table 2: Main results questionnaires Part 1	11
Table 3: Main results questionnaires Part 2.....	11
Table 4: EDM mapping ONB	20
Table 5: EDM mapping SBB.....	30
Table 6: EDM mapping UIB	35
Table 7: EDM mapping UBER	38
Table 8: EDM mapping MPIWG rare books collection	43
Table 9: EDM mapping MPWIG ISMI project.....	45
Table 10: ISMI Object Properties.....	47
Table 11: ISMI Data Properties	47
Table 12: EDM mapping NLI	52
Table 13: EDM mapping BBAW	55
Table 14: EDM mapping CRNS	59
Table 15: EDM mapping JDC	65
Table 16: EDM mapping UBFFM	72

List of Figures

Figure 1: UBER preliminary mapping	37
Figure 2: Diagram ISMI mapping	46
Figure 3: JDC hierarchy	60

1 Executive Summary

The “Requirements report” is the first of two deliverables of WP1 “Content”. It documents the results of Task 1.1 “Collect metadata formats and relational backend structures” and Task 1.2 “Collect requirements”. It was the main objective of both tasks to provide WP2 and WP3 with precise information about metadata and APIs currently available for accessing digitised objects contributed by each content provider.

During the five months after the project’s kick-off meeting, WP1 gathered a complete list of metadata formats. Furthermore, structure and characteristics of backends and formats have been analysed and requirements for the interoperability infrastructure were shared with WP2. This deliverable is the starting point for the crosswalks to EDM and thus for the final integration of content.

This document summarises replies of content providers given in questionnaires delivered after the kick-off meeting. They were asked to provide information concerning content, metadata, as well as detail information about databases and interfaces. They were also asked to provide information about resources available to adjust technical infrastructure, particularly APIs, to enable WP3 to access metadata and digital objects. Both questionnaires are attached in the Appendix (see page 75).

As a result of a close collaboration between WP1 and WP2 it turned out that most content providers need to enrich metadata for finalising crosswalks to comply with mandatorily required elements of the EDM. Most of those enrichments are of general nature and are listed in the Requirements chapter (p. 13 ff.). The Requirements chapter is broken down into three sections for each content provider: (1) underlying content (kind and amount of content), (2) requirements (summarised analysis and derived requirements), and (3) a preliminary crosswalk to EDM. When necessary, a short interim conclusion is added (4).

The outcome of the first questionnaire about the content was encouraging. Most partners confirmed the numbers of pages they agreed upon in the Description of Work (DoW). In addition to that, even new content providers agreed to grant access to digital libraries. Therefore, the objective to win new content providers during the project is within reach. This redounds to WP3 whose attractiveness increases with the amount of digital objects available.

It is to be underscored that this report describes the current state of a work in progress. Crosswalks of content provider’s metadata to EDM may be subject to changes. It might be also the case that not all of the requirements mentioned in this report can be met or additional requirements will emerge during the course of the project. Finally, it is also to be mentioned that few content providers could not employ project staff member in time and thus need to complement analysis of backend databases.

2 Introduction

Already digitised content of contributors often is not accessible outside providers' catalogues, since at least metadata can not be aggregated for they lack modern APIs. These systems are also called deep web/hidden web data applications. It is neither possible to access data via search engines like Google nor to interlink data types like places with web resources like GeoNames. By applying the interoperability infrastructure of WP2, content providers will map their metadata to the Europeana Data Model (EDM) and contextualise their data.

Since content providers make various types of metadata available which serve various purposes subject to the kind of (digital) object, it was crucial for WP1 – "Content" – to analyse source formats according to the joint and divergent needs concerning EDM.

To obtain an overview, content providers answered two questionnaires. The first questionnaire was conducted in early March 2012. It gives an overview about content and technical requirements for the planned content enrichment and recombination platform coded by WP3. The second questionnaire dealt with metadata. It focussed on the crosswalks of metadata to the EDM. It was conducted in May 2012. Both questionnaires showed that a mapping to EDM is only possible if special requirements regarding different metadata are taken into account.

2.1 Scope and Objective of this document

This document gives an overview of the provided content in the DM2E project as well as the current state of requirements needed for a successful integration of even the richest metadata of the content providers. This was a precondition for adjusting tools by WP2 for RDFisation and contextualisation. Furthermore, it is a guideline for WP3, who are exploring the content annotation through creating an advanced annotation platform. It will also give a first impression of common challenges which WP1 partners will face.

While a wide range of different metadata used by content providers had to be dealt with, this document is not designed to provide information about all aspects in full detail. These are available via Redmine, the project's working platform, and can be supplied upon request.

This deliverable contains the requirements gathered in the first 6 months of the project. Due to this, slight changes or adjustments are likely. Also new requirements may be detected during the course of the project.

2.2 The structure of this document

Chapter 3 gives an overview about the content provided by DM2E partners. The amount of images each institution is going to deliver will be established and updated against the amount agreed upon in the Description of Work. A general overview of the answers provided in the content and metadata questionnaires is included in this chapter as well.

In chapter 4 the requirements are listed by content provider.

Chapter 5 deals with the requirements gathering process of digital humanities scholars in WP3. The last chapter gives an outlook for the next steps in WP1.

3 Overview of content and metadata standards in DM2E

The following section provides an overview of the agreed amounts of content that will be ingested into Europeana.

During the first six months of the project it was already possible to encourage new content providers to make content available to the DM2E project. New partners are JDC (Joint Distribution Committee providing content via EAJC) and UBFFM (University Library Johann Christian Senckenberg Frankfurt am Main).

3.1 Underlying content

The table below displays the distribution of total pages provided by content providing partners. It is to be emphasised that the number of provided images increased in comparison with the total amount of images mentioned in the Description of Work (DoW). The majority of content providers confirmed the total amount of images as agreed upon in the project proposal. Moreover, it was possible to encourage two new content providers. They are added to the table. Even though slight changes occurred to the promised amounts by some providers, the overall number of total pages of 19.789.930 stated in the DoW already increased up to 20.006.930:

Institution	Type	Volumes	Pages	Language	Format	Metadata	IPR
BBAW	Books / Historical printings from 1650-1900	1.300	470.000	German	XML	TEI P5	CC BY-NC
CJH via EAJC	Books, posters, photographs, etc.	29.000	400.000	Various	JPG2000 / TIFF	MARCXML	Fair Use clause, Section 107 US Copyright Law: http://www.copyright.gov/title17/92chap1.html#107
	Audio files	1.200		Various	MP3	MARCXML	
	Movies	100		Various	Mpeg-4, Mpeg-2, Mov	MARCXML	
CRNS	Digitale Faksimile Gesamtausgabe (DFGA) – Digital facsimile reproduction of Nietzsche's manuscripts		19.330	German	JPEG	DC/ESE	CC BY-NC
JDC via EAJC (new provider)	Records of the American Jewish Joint Distribution Committee of the years 1914-1918		20.000	Various	PDF	EAD	http://archives.jdc.org/researchers/terms-and-conditions.html

Institution	Type	Volumes	Pages	Language	Format	Metadata	IPR
MPIWG	Rare book collection: books, newspapers	2.500	500.000	Various	JPEG	Own format	Public domain
	Manuscripts – Islamique Scientifique manuscripts	1.000	100.000	Various	JPEG	Own format	Public domain
	Autographs	4	1.200	Various	JPEG	Own format	Public domain
NLI	Manuscripts	1.000	55.000	Hebrew and other Jewish languages	TIFF / JPEG	MARCXML	Public domain, content available free of charge for non-commercial use
	Manuscripts	750-1000	50.000	European languages and Arabic	TIFF / JPEG	MARCXML	
	Books	7500	2.100.000	Hebrew and other Jewish languages	TIFF / JPEG	MARCXML	
	Books	2500	700.000	European languages	TIFF / JPEG	MARCXML	
	Archival material / Portraits		3000	Various / Not relevant	TIFF / JPEG	MARCXML	
ONB	Books - ABO	22.500	6.750.000	German	JPEG2000	MAB2	Content public domain, available free-of-charge for non-commercial use
	Books - ABO	5.500	1.650.000	Latin	JPEG2000	MAB2	
	Books - ABO	4.000	1.200.000	French	JPEG2000	MAB2	
	Books - ABO	3.000	900.000	Italian	JPEG2000	MAB2	
	Books - ABO	2.500	750.000	Czech	JPEG2000	MAB2	
	Books - ABO	2.500	750.000	English	JPEG2000	MAB2	
	Books - ABO	10.000	3.000.000	Various	JPEG2000	MAB2	
	Manuscripts - codices		18.600	Latin	JPEG or TIFF	MAB2	
	Manuscripts - codices		6.200	German	JPEG or TIFF	MAB2	
	Manuscripts - codices		4.800	Greek	JPEG or TIFF	MAB2	
	Manuscripts - codices		12.800	Various	JPEG or TIFF	MAB2	
SBB	Personal papers of Jean Paul		17.000	German	JPEG or TIFF	MAB2	CC BY-NC-SA

Institution	Type	Volumes	Pages	Language	Format	Metadata	IPR
	Personal papers of Theodor Fontane		7.000	German	JPEG or TIFF	MAB2	CC BY-NC-SA
	Personal papers of Arthur Schopenhauer		10.000	German	JPEG or TIFF	MAB2	CC BY-NC-SA
	Collection of Margarete Hauptmann		2.000	German	JPEG or TIFF	MAB2	CC BY-NC-SA
	Personal papers of Adelbert von Chamisso		20.000	German	JPEG or TIFF	MAB2	CC BY-NC-SA
	Personal papers of Gerhart Hauptmann		30.000	German	JPEG or TIFF	MAB2	CC BY-NC-SA
	Personal papers of Dietrich Bonhoeffer		20.000	German	JPEG or TIFF	MAB2	CC BY-NC-SA
UBER	Polytechnisches Journal	346	205.000	German	PNG	TEI P5	Text: CC BY-NC-SA 3.0 Image: CC BY-NC-ND 3.0
UBFFM (new provider)	Manuscripts	600	230.000	Hebrew; Latin; German; Various	JPEG	METS/MODS	Public Domain: CCO
UIB	Wittgenstein Nachlass manuscripts and typescripts	20	5.000	German, English	JPEG	XML; TEI P5	CC BY-NC-SA
		Pages total	20.006.930				

Table 1: Underlying content

Summarised, content providers will provide 20.006.930 images, 1.200 audio files, and another 100 video files. Accordingly, to the Performance Monitoring Table the total number of pages beyond the DoW should be 5000 additional pages within the first year. This number could be already achieved through the two new content providers.

3.2 Main results of the conducted questionnaires

The table below contains the most important results of the conducted content and metadata questionnaires:

	ONB	SBB	UBER	UIB	MPIWG
Metadata available	Yes	Yes (partly)	Yes (partly)	Yes	Yes
Language of metadata	German	German	German / English	English	English / Arabic
Metadata formats	MAB2	MAB2, EAD	TEI XML	TEI P5	Own format
Digitised objects available	Partly	Partly (40%)	Yes	Yes	Yes
Delivery of full-texts	Yes	No	Yes	Yes	Partly yes
Format of full-text	METS, HOCR	-	TEI P5	Plain text	Plain text
Character encoding of full-text	UTF-8	-	UTF-8	UTF-8	UTF-8
Textual mark-up	No	-	Yes (TEI P5)	Yes (TEI P5)	Partly ECHO schema

	ONB	SBB	UBER	UIB	MPIWG
Granularity of objects	Up to word level	-	Up to page level	Up to page level	Up to page level
Persistent Identifiers	Not yet but planned	Not yet	No	Yes	Not yet
Usage of specific thesauri or vocabularies / references to vocabularies	GND ¹	PND ² , GKD ³	PND	No	No

Table 2: Main results questionnaires Part 1

	NLI	BBAW	JDC	CJH	UBFFM	CRNS
Metadata available	Yes	Yes	Yes	Yes	Yes	Yes
Language of metadata	Hebrew / English	German	English	English	German	English / German
Metadata formats	MARCXML	TEI P5	EAD XML	MARCXML	METS/MODS	ESE
Digitised objects available	Mostly yes	Yes	Yes	Yes	Yes, partly	Yes
Delivery of full-texts	Partly yes	Yes	Yes	Partly yes	No	No
Format of full-text	OCR	Transcription	OCR	OCR and transcription	-	-
Character encoding of full-text	UTF-8	UTF-8	UTF-8	UTF-8	-	-
Textual mark-up	No	Yes	No	No	-	-
Granularity of objects	Books/Manuscripts: Up to (double) page level; title level; Portraits: Picture level	Sentence level	Up to page level	Page level	Up to page level	Up to page level
Persistent Identifiers available	Not yet but planned	Not yet	Yes	Yes	Yes, URN	Yes
Usage of specific thesauri or vocabularies	VIAF ⁴	GND	-	AAT ⁵ ; TGM ⁶ ; LCSH ⁷	GND	Simple Scholarship

Table 3: Main results questionnaires Part 2

This list represents the main supported metadata formats used in DM2E:

- MAB2: ONB; SBB
- Encoded Archival Description (EAD XML): JDC; SBB
- MARC21: NLI; CJH
- Text Encoding Initiative (TEI): BBAW; UBER; UIB
- METS/MODS: UBFFM
- ESE / DC: CRNS

¹ GND = Gemeinsame Normdatei (Universal Authority File).

² PND = Personennormdatei (Name Authority File).

³ GKD = Gemeinsame Körperschaftsdatei (Corporate Bodies Authority File).

⁴ VIAF = Virtual International Authority File.

⁵ AAT = Art & Architecture Thesaurus.

⁶ TGM = Thesaurus for Graphic Materials.

⁷ LCSH = Library of Congress Subject Headings.



-
- Provider-specific models: MPIWG

Seven partners deliver digitised cultural heritage objects within full text: NLI, ONB, UBER, UIB, JDC, CJH, and MPIWG. Most of these searchable texts were created through OCR (ONB; UBER; NLI; JDC; CJH). Four partners created full texts through transcription (BBAW; UIB; MPIWG; CJH). All full texts are encoded in UTF-8.

4 Requirements

During the first six months of the project it became apparent that due to the unique character of most providers' collections the herein stated requirements need to be listed by content providers and not by metadata format. Special focus needs to be put on the content to understand the requirements for a successful EDM mapping.

First analyses of the provided sample records showed that there are also requirements from WP2 for the content providers, i.e. a full mapping to EDM is only possible if all mandatory fields are provided. In most cases the concerned fields are very generic and applied for whole collections, e.g. *edm:rights*. In such cases a default value can be added during the transformation process.

Below the requirements are listed by content provider. They contain information needed by WP2 and WP3 for a successful infrastructure development.

4.1 ONB – Austrian National Library (Österreichische Nationalbibliothek)

4.1.1 Underlying content

The Austrian National Library (ONB) will provide the heterogeneous content of 50.000 volumes from the Google Books Public Private Partnership ABO (Austrian Books Online) as well as 170 digitised manuscripts (codices). No deviations occurred to the amounts stated in the Description of Work.

The ABO content will include mainly books, but also newspapers and already digitised manuscripts in different languages. Following statistical analyses of the catalogue about 45 % of the content is in German, 11 % in Latin, 8 % in French, 6 % in Italian and 5 % each in English and Czech. Another 20 % is written in various languages. During the Austrian Books Online project the Austrian National Library digitises and makes available online its complete holdings of historical books from the 16th to the 19th century.

The 170 manuscripts (codices) are in various languages such as Latin, German, Greek, Italian, Arabic, Turkish or French.

Overall, the ONB will contribute about 15.000.000 pages during the DM2E project.

4.1.2 Requirements

The main exchange and internal data format at the Austrian National Library is MAB2. It is the standard exchange format used by most libraries in Austria and Germany. Even though it is possible to export ONB's data in an XML format (MARC21), crucial information gets lost during this export as not all MAB2 fields are represented in MARC21, e.g. during the export to MARC the information in the MAB fields 662a (condition) and 664 (provenance general) gets lost (please see codices examples below). Especially in the case of manuscripts this information might be very valuable for digital humanities scholars. Therefore, the provided XML-based format MARC is not suitable for the DM2E project as the project stresses the importance of extracting as many information as possible from

legacy data. Consequently, the main requirement is the possibility to map directly from MAB2 to EDM. For now, the MINT tool can only map XML based formats. Due to this, a conversion of MAB2 data to XML has to be explored and tested. However, an export to MARC XML is not an option as valuable information gets lost and due to the large amount of provided data not every data set can be analysed individually.

The following examples represent both collections provided by the ONB during DM2E in MAB2, followed by the MARC format export (relevant information highlighted bold):

Codices MAB2

```

000147226 LDR-1 L -----nM2.01200024-----h
000147226 FMT-1 L MH
000147226 001-1 L $$aAL00161817
000147226 002a1 L $$a20110112
000147226 003-1 L $$a20111202
000147226 037b1 L $$alat
000147226 050-1 L |aa|||||||||
000147226 053-1 L 3|
000147226 070a1 L $$aONB-HANNA
000147226 070a1 L $$aONB
000147226 070-1 L $$aONB
000147226 070b1 L $$aONB
000147226 100-1 L $$aLivius, Titus$$9118573624$$b[Verfasser]
000147226 200-1 L $$aReichsabtei <Lorsch>$$9801576-4$$b[Vorbesitzer]
000147226 204a1 L $$aKunsthistorisches Museum <Wien> / Sammlungen
Schloss Ambras$$91050601-9$$b[Vorbesitzer]
000147226 331-1 L $$aHistoriae romanae decas V lib. 1-5
000147226 410u1 L $$aItalien
000147226 425-1 L $$a1. Hälfte 5. Jhdt.
000147226 425a1 L $$a400-450
000147226 433-1 L $$aHandschrift; IV+193+IV* Bll.
000147226 435-1 L $$a250 x 150 mm
000147226 501-1 L $$aVolldigitalisat.
000147226 537f1 L $$aHist. prof. 626
000147226 544-1 L $$aCod. 15$$cHan
000147226 655e1 L
$$uhttp://www.onb.ac.at/sammlungen/hschrift/bibliographie.htm$$zLitera
tur zur Handschrift
000147226 655e1 L
$$uhttp://www.bildarchiv.at/ProfiSzettel.aspx?a=b&wort=Cod%2015&Wien$$
zZettelkatalog des Bildarchivs
000147226 655e1 L $$uhttp://www.manuscripta-
mediaevalia.de/hs/katalogseiten/HSK0751a_b0002_jpg.htm$$zTabulae
codicum 1864-1893
000147226 662a1 L $$aPergament
000147226 662b1 L $$aPappeinband, Wien, 1669
000147226 664a1 L $$aLorsch, Benediktinerkloster (um 764-1232): 1531
von Simon Grynaeus in Lorsch zur Teiledition benutzt [Endlicher, 1836,
Nr. 105]. - Ambras, Schloß bei Innsbruck: Signatur A 287 [Endlicher,
1836, Nr. 105].
000147226 LOW-1 L $$aZNAK
000147226 CAT-1 L $$aONBFISI$$b20$$c20110112$$lACC05$$h1337
000147226 CAT-1 L $$aBATCH-UPD$$b20$$c20110112$$lACC05$$h1337
000147226 CAT-1 L $$aONBFISI$$b20$$c20110408$$lACC05$$h1611
000147226 CAT-1 L $$aONBFISI$$b20$$c20110408$$lACC05$$h1611
000147226 CAT-1 L $$aONBFISI$$b20$$c20111202$$lACC05$$h0950

```

```

000147226 CAT-1 L $$aBATCH-UPD$$b20$$c20120115$$lONB06$$h1408
000147226 RAO-1 L $$uhttp://opac.obvsg.at/acc05-own/acc05-
    znak.html$$zÖsterr. Nationalbibliothek
000147226 TYP L $$aWerk(e)
000147226 PAZ L $$1Livius, Titus$$2[Verfasser]$$3Werk(e)
000147226 KAZ L $$1Reichsabtei <Lorsch>$$2[Vorbesitzer]$$3Werk(e)
000147226 PML L $$aAL00161817
  
```

Codices export in MARC XML

```

<?xml version = "1.0" encoding = "UTF-8"?>
  <collection xmlns="http://www.loc.gov/MARC21/slim"
    xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xsi:schemaLocation="http://www.loc.gov/MARC21/slim
    http://www.loc.gov/standards/marcxml/schema/MARC21slim.xsd">
    <record xmlns="http://www.loc.gov/MARC21/slim"
      xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
      xsi:schemaLocation="http://www.loc.gov/MARC21/slim
      http://www.loc.gov/standards/marcxml/schema/MARC21slim.xsd">
      <leader>^^^^^ntm^^2200000^u^4500</leader>
      <controlfield tag="001">AL00161817</controlfield>
      <controlfield tag="008">110112 |||
lat</controlfield>
      <controlfield tag="005">20111202</controlfield>
      <datafield tag="041" ind1="0" ind2="7">
        <subfield code="a">lat</subfield>
        <subfield code="2">ISO 639-2</subfield>
      </datafield>
      <controlfield tag="007">zz||||||||||||||||||</controlfield>
      <datafield tag="100" ind1="1" ind2=" ">
        <subfield code="a">Livius, Titus</subfield>
        <subfield code="e">Verfasser</subfield>
      </datafield>
      <datafield tag="110" ind1="2" ind2=" ">
        <subfield code="a">Reichsabtei (Lorsch)
      &lt;&lt; [Vorbesitzer] &gt;&gt;</subfield>
      </datafield>
      <datafield tag="710" ind1="2" ind2=" ">
        <subfield code="a">Kunsthistorisches Museum (Wien)</subfield>
        <subfield code="b">Sammlungen Schloss Ambras
      &lt;&lt; [Vorbesitzer] &gt;&gt;</subfield>
      </datafield>
      <datafield tag="245" ind1="0" ind2="0">
        <subfield code="a">Historiae romanae decas V lib. 1-
5</subfield>
      </datafield>
      <datafield tag="260" ind1=" " ind2=" ">
        <subfield code="a">Italien</subfield>
        <subfield code="c">1. Hälfte 5. Jhdt., 400-450</subfield>
      </datafield>
      <datafield tag="300" ind1=" " ind2=" ">
        <subfield code="a">Handschrift; IV+193+IV* Bll.</subfield>
        <subfield code="c">250 x 150 mm</subfield>
      </datafield>
      <datafield tag="500" ind1=" " ind2=" ">
        <subfield code="a">Volldigitalisat.</subfield>
      </datafield>
  
```

```

<datafield tag="090" ind1=" " ind2=" ">
  <subfield code="a">Cod. 15</subfield>
</datafield>
<datafield tag="856" ind1="4" ind2=" ">
  <subfield
code="u">http://www.onb.ac.at/sammlungen/hschrift/bibliographie.htm</s
ubfield>
  <subfield code="z">Literatur zur Handschrift</subfield>
</datafield>
<datafield tag="856" ind1="4" ind2=" ">
  <subfield
code="u">http://www.bildarchiv.at/ProfiSzettel.aspx?a=b&wort=Cod%2
015&Wien</subfield>
  <subfield code="z">Zettelkatalog des Bildarchivs</subfield>
</datafield>
<datafield tag="856" ind1="4" ind2=" ">
  <subfield code="u">http://www.manuscripta-
mediaevalia.de/hs/katalogseiten/HSK0751a_b0002_jpg.htm</subfield>
  <subfield code="z">Tabulae codicum 1864-1893</subfield>
</datafield>
</record>

```

ABO MAB2

```

002792410 LDR-1 L -----nM2.01200024-----h
002792410 FMT-1 L MH
002792410 LBA-1 L $$a0
002792410 001-1 L $$aLZ00958442
002792410 029-1 L $$aAZ00001157
002792410 030-1 L u|luf|||||
002792410 037b1 L $$alat
002792410 050-1 L a|a|||||
002792410 051-1 L |||
002792410 070-1 L $$aONB
002792410 100-1 L $$aAbbot, Robertus
002792410 331-1 L $$aDe suprema potestate regia exercitationes habitae
in Academia Oxoniensi contra Rob. Bellarminum et Franciscum Suarez
002792410 410-1 L $$aLondini$$xLondon
002792410 412-1 L $$aNorton
002792410 425a1 L $$a1619
002792410 501-1 L $$aAus der Bibliothek des Prinzen Eugen von Savoyen.
002792410 507-1 L $$aDe suprema potestate regia exercitationes habitae
in Academia Oxoniensi contra Rob. Bellarminum et Franciscum Suarez. -
Londini, Norton. 1619.
002792410 655o1 L
$$uAZ00001157SZ00094702SZ00414888SZ01137251SZ01137304SZ01158909
002792410 902x1 L $$pBellarmino, Roberto$$94005422-6
002792410 902x1 L $$sPolitische Philosophie$$94076226-9
002792410 902x1 L $$sKritik$$94033229-9
002792410 904x1 L $$aONB-AK
002792410 907x1 L $$sGewaltenteilung$$94071817-7
002792410 909x1 L $$aONB-AK
002792410 912x1 L $$sPolitische Philosophie$$94076226-9
002792410 912x1 L $$pBellarmino, Roberto$$94005422-6
002792410 912x1 L $$sKritik$$94033229-9
002792410 914x1 L $$aONB-AK
002792410 917x1 L $$sPolitische Philosophie$$94076226-9

```



```

002792410 917x1 L $$pSuárez, Francisco$$94058300-4
002792410 917x1 L $$sKritik$$94033229-9
002792410 919x1 L $$aONB-AK
002792410 922x1 L $$pSuárez, Francisco$$94058300-4
002792410 922x1 L $$sPolitische Philosophie$$94076226-9
002792410 922x1 L $$sKritik$$94033229-9
002792410 924x1 L $$aONB-AK
002792410 927x1 L $$sReformationsliteratur$$94201421-9
002792410 929x1 L $$aONB-AK
002792410 SGA-1 L $$aBE.3.L.21$$cZALT$$dPRUNK$$gR$$h02
002792410 CAT-1 L $$aIMP02B$$b20$$c20100727$$lONB01$$h1133
002792410 CAT-1 L $$aSWUPD01$$b20$$c20111203$$lONB01$$h1015
002792410 LDR 9 L 00000nM2.01000024-----1
002792410 FMT 9 L ML$$2ONB50
002792410 012 9 L $$aLZ00958442$$lONB01$$2ONB50
002792410 070 9 L $$aBATCH$$2ONB50
002792410 070b9 L $$aONB$$2ONB50
002792410 711g9 L $$aReformationsschriften$$2ONB50
002792410 711z9 L $$a17. Jh.$$2ONB50
002792410 OWN 9 L $$aONB50
002792410 LOC L
$$mBOOK$$aZALT$$lZALT$$2PRUNK$$cB941629$$bB941629$$dBE.3.L.21$$4BE.3.L
.21$$n02$$k00000000

```

ABO export in MARC XML

```

http://www.loc.gov/standards/marcxml/schema/MARC21slim.xsd"
  <record xmlns="http://www.loc.gov/MARC21/slim"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.loc.gov/MARC21/slim
http://www.loc.gov/standards/marcxml/schema/MARC21slim.xsd"
  <leader>      nam 2200000 u 4500</leader>
  <controlfield tag="001">LZ00958442</controlfield>
  <controlfield tag="008">          s1619          |||
lat</controlfield>
  <datafield tag="041" ind1="0" ind2="7">
  <subfield code="a">lat</subfield>
  <subfield code="2">ISO 639-2</subfield>
</datafield>
  <controlfield tag="007">zz||||||||||||||||||||</controlfield>
  <datafield tag="100" ind1="1" ind2=" ">
  <subfield code="a">Abbot, Robertus</subfield>
</datafield>
  <datafield tag="245" ind1="0" ind2="0">
  <subfield code="a">De suprema potestate regia exercitationes
habitae in Academia Oxoniensi contra Rob. Bellarminum et Franciscum
Suarez</subfield>
</datafield>
  <datafield tag="260" ind1=" " ind2=" ">
  <subfield code="a">Londini</subfield>
  <subfield code="b">Norton</subfield>
  <subfield code="c">1619</subfield>
</datafield>
  <datafield tag="500" ind1=" " ind2=" ">
  <subfield code="a">Aus der Bibliothek des Prinzen Eugen von
Savoyen.</subfield>
</datafield>

```

```

<datafield tag="246" ind1=" " ind2=" ">
  <subfield code="a">De suprema potestate regia exercitationes
habitaee in Academia Oxoniensi contra Rob. Bellarminum et Franciscum
Suarez. - Londini, Norton. 1619.</subfield>
</datafield>
<datafield tag="955" ind1=" " ind2=" ">
  <subfield code="1">ZALT</subfield>
  <subfield code="2">PRUNK</subfield>
  <subfield code="b">B941629</subfield>
  <subfield code="4">BE.3.L.21</subfield>
</datafield>
<datafield tag="600" ind1="1" ind2="4">
  <subfield code="a">Bellarmino, Roberto</subfield>
  <subfield code="x">Politische Philosophie</subfield>
  <subfield code="x">Kritik</subfield>
</datafield>
<datafield tag="650" ind1=" " ind2=" ">
  <subfield code="a">Gewaltenteilung</subfield>
</datafield>
<datafield tag="650" ind1=" " ind2=" ">
  <subfield code="a">Politische Philosophie</subfield>
  <subfield code="x">Bellarmino, Roberto</subfield>
  <subfield code="x">Kritik</subfield>
</datafield>
<datafield tag="650" ind1=" " ind2=" ">
  <subfield code="a">Politische Philosophie</subfield>
  <subfield code="x">Suárez, Francisco</subfield>
  <subfield code="x">Kritik</subfield>
</datafield>
<datafield tag="600" ind1="1" ind2="4">
  <subfield code="a">Suárez, Francisco</subfield>
  <subfield code="x">Politische Philosophie</subfield>
  <subfield code="x">Kritik</subfield>
</datafield>
<datafield tag="650" ind1=" " ind2=" ">
  <subfield code="a">Reformationsliteratur</subfield>
</datafield>
</record>

```

As can be seen in the provided examples, both collections contain similar basic metadata information. Depending on the provided object additional MAB2 fields are used, e.g. in case of manuscripts additional information about the condition of the book (662a) or the cover (662b) is provided.

Another characteristic of the provided collections are subject strings. This is visible in the ABO MAB2 example in the fields 902-926. These fields represent the subject key words listed as they appeared in the old card catalogues. It might be the case that duplicates appear in those fields. The handling of duplicates needs to be established for an EDM mapping. Subject strings which contain the same information in a different order can be removed. However, if a new subject topic appears this information needs to be kept (compare MAB2 subject strings 902, 912, and 917). A technical analysis of the provided conversion tools needs to be carried out with regards to this issue. It needs to be established if a duplicate check is possible during the mapping or if this needs to be done beforehand.

EDM extension requirements might be necessary for the following local fields:

- Signature (MAB2 SGA)
- Condition (MAB2 662a)
- Cover (MAB2 662b)

It became also clear that the ONB data needs to be enriched to fulfil the mandatory requirements of the EDM model. The following information needs to be provided:

- **edm:rights:** This information will be a default value and added during the mapping.
- **edm:data provider:** This also will be a default value: "ONB" (field 070a)
- **edm:isShownBy** or **edm:isShownAt:** This information will be added once the digitised representations are available; stable URNs will be probably available in the 552b MAB2 field.
- **edm:type:** This information will be the default value "TEXT".

4.1.3 Preliminary mapping of mandatory EDM fields

Property	Property range	Property description	Source Element	Source Element Description
ore:Aggregation				
edm:provider	edm:Agent entity	Mandatory element: Organisation (edm:Agent, see below) that provided this aggregation	Default value "DM2E"	
edm:dataProvider	edm:Agent entity	Mandatory element: Organisation (edm:Agent, see below) that provided the base data for this aggregation (could be the same as edm:provider)	070a	
edm:rights	remote resource	Mandatory element: URL of a resource describing licensing rights from http://pro.europeana.eu/documents/900548/1037382/Europeana_rights_201202.pdf	Will be a default value	
edm:isShownBy	WebResource entity	Either edm:isShownBy or edm:isShownAt is mandatory. Web resource (see below) describing this specific ProvidedCHO aggregated here	655e / URNs will be probably available in the 552b field	Once the digitised representations are available
edm:isShownAt	WebResource entity	Either edm:isShownBy or edm:isShownAt is mandatory. Web resource (see below) containing the ProvidedCHO aggregated here		
dct:created	xsd:dateTime literal	Creation and date of this aggregation	002a	
dct:modified	Xsd:dateTime literal	Modification date and time of this aggregation	003	
edm:ProvidedCHO				
edm:type	Literal	Mandatory element: Can be TEXT, VIDEO, SOUND, IMAGE, 3D	Default value: "TEXT"	
dc:type	Specialization of ore:Proxy	Either dc:type or dc:subject is mandatory. The most specific type that applies to the CHO, repetition of rdf:type of Proxy		

Property	Property range	Property description	Source Element	Source Element Description
dc:title	Literal with language tag	Either dc:title or dc:description is mandatory. Title of the object.	331	
dc:description	Literal with language tag	Either dc:title or dc:description is mandatory: Description of the CHO	507	Additional information to the title
dc:language	xsd:lang literal	Mandatory for textual objects: Most prominent language of the CHO	037b	ISO 639
dct:issued	xsd:dateTime literal	Date of publication, also possible: dct:created, or dc:date for generic cases	425(a)	
dc:creator	edm:Agent entity	Creator of the CHO, possibly its author. Subproperties may be used to specialize relationship	100 200 204a	100 author 200 cooperate body creator
dc:publisher	edm:Agent entity	Publisher of the CHO	412	
dm2e:publishedAt	edm:Place entity	Place of publication, subproperty of dct:spatial, other extensions possible	410u	
dc:subject	Resource	Either dc:type or dc:subject is mandatory. Resource in SKOS-Concept-Tree, e.g. Dewey Decimal Classification	902; 912; 917; 922; 927;	
dct:provenance	Literal	Description of provenance of CHO	664	

Table 4: EDM mapping ONB

4.2 SBB – Berlin State Library (Staatsbibliothek zu Berlin)

Kalliope, the Union Catalogue for Personal Paper and Autograph Collections in Germany, is hosted by the Berlin State Library (SBB). Metadata from 491 Cultural Heritage Institutions have been aggregated to date. Therefore, about 20.000 collections and 1.64 Million autographs covering a time span from the early 16th to the late 20th century can be retrieved online. Among all autographs are 1.45 Million letters, 130.000 manuscripts, and 60.000 documents such as reports, photos, or certificates. In most cases metadata describe unique cultural heritage objects maintained in library's manuscript divisions and archives. SBB's manuscript division will provide 100.000 images from its digitised holdings.

4.2.1 Underlying content

SBB provides 100.000 images at large. In detail:

1. Personal papers of Jean Paul (writer, 1763 – 1825)
Content: mainly manuscripts (17.000 images)
2. Personal papers of Theodor Fontane (writer, 1819 – 1898)
Content: diaries (7.000 images)
3. Personal papers of Arthur Schopenhauer (philosopher, 1788 – 1860)
Content: mainly manuscripts (10.000 images)
4. Collection of Margarete Hauptmann (actress, 1875 – 1957)
Content: photos (2.000 images)

5. Personal papers of Adelbert von Chamisso (poet and botanist, 1781 – 1838)
Content: mainly manuscripts, letters, Almanach des Muses, certificates (20.000 images)
6. Personal papers of Gerhart Hauptmann (dramatist and novelist, 1862 – 1946)
Content: mainly letters, notebooks (30.000 images)
7. Personal papers of Dietrich Bonhoeffer (theologian, member of the resistance, 1906 – 1945)
Content: letters, manuscripts, documents (20.000 images)

While the first four collections are already digitised, digitisation of the last three will be done by August 2013. It is to be notified, however, that the amount of images provided within the project (100.000) differs from the figures that had been stated in the project proposal (140.000). This is due to digitisation projects which had been scheduled for 2011 but were postponed to 2013.

4.2.2 Legacy Data formats

Preliminary note

Personal paper collections like archival materials in general require a multilevel description approach in line with the International Standard Archival Description (ISAD(G)). Therefore, cataloguing personal paper collections is not a plain description of single items, but also the description of the structure of a holding to provide context information. The cataloguing application provided by the Union Catalogue is based on an adapted library cataloguing system meeting this particular requirement. It enables cataloguers to set hierarchal relations between records describing various description levels. In addition to that, when the application was introduced, the data exchange format MAB2, common exchange format of libraries in Austria and Germany, had been also adjusted to special requirements. However, MAB flattens some important relationships between records and thus both MAB and the backend database have to be analysed in more detail. Moreover, an additional XML-based international data exchange format, Encoded Archival Description (EAD), is currently set up and tested. It will be available in September this year. Test records were already provided to the project for data analysis.

(a) MAB-Data

Kalliope's technical infrastructure comes with import and export services for data aggregation and delivery. Customers are mainly manuscript divisions of libraries in Germany, but also archives and museums. The Aggregation and delivery service is currently solely based on the MAB2 format mentioned beforehand. Therefore, within the first months of the project, SBB focused on a crosswalk MAB2 to EDM. Examples from catalogued collections which are exemplary for most data of Kalliope were uploaded to the project wiki and discussed among project partners:

- Personal papers of Dietrich Bonhoeffer
- Personal papers of Gerhart Hauptmann
- Personal papers of Adelbert von Chamisso
- Personal papers of Franz Brümmer
- Collection of Ludwig Darmstaedter

When exporting a collection, four files are created:

1. file: data describing the collection and sub-collections
2. file: data describing the manuscripts
3. file: authority data for persons associated with collections and manuscripts
4. file: authority data for corporate bodies associated with collections and manuscripts as well as the repository holding a collection (e.g. ISIL DE-1a for SBB)

Multilevel relationships between records in the MAB2 format are based on a bottom-up approach. Each record comes with the categories 451 and 453. While 451 provides the whole path from a respective record up to the top level (collection level), 453 provides the record number of the record superior to the respective record (453 = dct:isPartOf).

The following two records are an example from the Gerhart Hauptmann collection:

Example 1

```
### 00001nM2.01200024      h
001 01646186
002a20100408
003 20111019
004 20120307
036aXA-CH
037bger
053 b
070a611
071 1a
100 Hauptmann, Gerhart  a[Verfasser/in] a
102a118546937
104 Hauptmann, Margarete a[Schreiber/in] a
106a116534788
108bBrahm, Otto  a[Adressat/in] a
110a118514245
112fRittner, Rudolf
114a118788981
116fIbsen, Henrik
118a118555286
120fBahr, Hermann
122a118505955
124fSchnitzler, Arthur
126a118609807
128fFulda, Ludwig
130a118840452
132fSudermann, Hermann
134a118619853
136fHauptmann, Gerhart
138a118546937
410 Lugano
426a10.10.1907
433 1 Brief, 2 Bl.
451 Nachl. Gerhart Hauptmann/Briefnachlass A: Prominente, Freunde und
Bekannte/Ba-Bu
453 00032035
511 Villa Castagnola
544 GH Br NL A: Brahm, Otto, 1, 4-5
588 Die versunkene Glocke
588 Schluck und Jau
```

661cEntwurf; Schluss fehlt
 750mAntwortet auf eine von Brahm nicht erteilte Freigabe des St cks
 'Schluck und Jau'. Bittet, die Entscheidung zu berdenken, da davon
 "meine geistige Existenz, meine geistige Zukunft abh„ngt". Beklagt
 sich ber die von Brahm und dessen Regisseuren verschuldeten
 Misserfolge, die seine "Freude am Theatralischen stark abgek hlt" und
 seine "Schaffenslust nicht gesteigert haben". F hlt sich durch die
 Wiederaufnahme seiner alten St cke "vernachl„ssigt und gesch„digt",
 weil darin "ebenso wenig die alte Liebe zu finden ist wie die alte
 Kunst". 'Die versunkene Glocke' macht einen "geradezu peinlich
 verstaubten Eindruck". Verweist auf den f r Brahm beraus g nstigen
 Vertrag und fordert "Liberalit„t" ein. Brahm jedoch verweigere ihm
 "mit einer sozusagen eisernen Hand" einen "k nstlerischen
 Bewegungsversuch". F hlt sich k nstlerisch missachteten Schauspielern
 gleich, wie dem "vortrefflichen Rittner", "dem die Verh„ltnisse viel
 zu fr h die B hne verleidet haben". Bittet um Freigabe von 'Schluck
 und Jau' f r ein "Experiment auf einer fremden B hne". Verweist auf
 andere Autoren in Brahms Programm und leitet daraus das Recht ab,
 'Schluck und Jau' am Deutschen Theater auff hren zu lassen. Einer
 "Wiederaufnahme im Lessingtheater, bei dem Personal, ber das Deine
 B hne jetzt verf gt und bei der gleichen Regie, die den ersten
 Misserfolg des St ckes nicht hat verhindern k“nnen", steht er
 "hoffnungslos" gegen ber.

Example 2

```
### 00002nM2.01200024      h
001 01731557
002a20110418
003 20110816
004 20120307
036aXA-DE
037bger
053 b
070a1a
071 1a
100 Hauptmann, Gerhart  a[Verfasser/in] a
102a118546937
104bRathenau, Walther  a[Adressat/in] a
106a118598430
108fRathenau, Walther
110a118598430
112fHauptmann, Gerhart
114a118546937
410 Berlin
426a5.12.1905
433 1 Brief, 2 Bl.
451 Nachl. Gerhart Hauptmann/Briefnachlass A: Prominente, Freunde und
Bekannte/Ra-Ru
453 00032087
511 Kaiserin-Augusta-Str. 74
544 GH Br NL A: Rathenau, Walther, 2, 3-4
588 Und Pippa tanzt! Ein Glash ttenm„rchen
588 Zur Physiologie der Moral
750mDankt f r die Zusendung eines Aufsatzes und l„dt Rathenau f r den
n„chsten Tag zu sich ein um aus 'Und Pippa tanzt!' zu lesen.
```

For a better understanding of all categories and indicators of the MAB2 exchange format, SBB worked out a translation including a first crosswalk to EDM:

MAB2	Name	EDM
001	ProviderID	dc:identifier
002a	Date created	dct:created
003	Date modified	dct:modified
004	Date delivered	
036a	Countrycode	
037b	Languagecode	dc:language
053	Document type (n = collection, l =files, w = publications, b = letters)	
070a	Data Provider Code (Code/ISIL: always DE-611 (Kalliope-Verbundkatalog))	edm:provider
071	Content Provider Code (Code/ISIL, e.g. DE-1a)	edm:dataProvider
100	Person creator (author)	dc:creator
100b	Person addressee (letters)	
100f	Person subject	dc:subject
102a	Person Authority Number PND	
102b	Person ID ZKA/ Kalliope	
102c	Person ID local	
104a	Person creator (author)	dc:creator
104b	Person addressee (letters)	
104f	Person subject	dc:subject
106a	Person Authority Number PND	
106b	Person ID ZKA/ Kalliope	
106c	Person ID local	
ff.	to be continued 196/198	
200	Coporate Body creator (author)	dc:creator
200b	Corporate Body adresse (letters)	
200f	Corporate Body subject	dc:subject
202a	Person Authority Number PND	
202b	Person ID ZKA/ Kalliope	
202c	Person ID local	
204a	Coporate Body creator (author)	dc:creator
204b	Corporate Body adresse (letters)	
204f	Corporate Body subject	dc:subject

MAB2	Name	EDM
206a	Person Authority Number PND	
206b	Person ID ZKA/ Kalliope	
206c	Person ID local	
ff.	to be continued 296/298	
310	Uniform title	dct: alternative
331	Unittitle	dc: title
335	Unittitle note	
359	Title Statement	
370	Additional title	dct: alternative
365	Additional title note	
410	Geonames	dct: coverage
426	Date range note/ text	
426a	Date range normalized	dct: issued
433	Extent	dct: extent
434	Illustrations	
451	Shelfmark of superordinate record	
453	IF of superordinate record	dct: isPartOf
501	Note external	
511	Place note (410)	
516	Remarks on the CHO	
516a	language of the CHO	
516b	Font/ writing of the CHO	
524	Remarks on other CHO contained in the desc. CHO	
544	Shelfmark local	
544	Shelfmark	
544b	Old Shelfmark	
581	Acquisition Number	
582	Inventory Number	
583	Box number	
584	Folder number	
585	Folio number	
586	Further Shelfmarks	
587	Other Shelfmarks	
588	Publication title (reference)	dct: references
599	Publication title ID (reference)	

MAB2	Name	EDM
655e\$u	URL	edm:isShownBy
655e\$z	URL note	
661a	Incipit of the manuscript/ letter	
661b	Uniform incipit of the manuscript/ letter	
661c	Level of Genesis	
661d	Genre	dc:type
662a	Condition	
662b	Cover	
662c	Watermark	
662d	Support	
662e	Restoration	
663a	Edition reference	
663b	Publication reference	
663c	Other Related Publications reference	
664	Provenance general	dct:provenance
664a	Provenance remark	
664b	Acquisition remark	
664c	Loss	
669a	Use restrict	
669b	Access restrict	
750m	Abstract	dct:description
na	Location Library	edm:currentLocation

The crosswalk for at least the most important data was carried out manually since the MINT-Tool, which has been developed to simplify this task, prerequisites XML-Data. MAB2, however, is encoded with binary characters (CP850) and thus could not be processed by MINT. SBB developed a perl script to transform the binary format into XML recently and has developed a first simple EDM-Transformation script:

```
<?xml version="1.0" encoding="UTF-8"?>
<rdfs xmlns:dc="http://purl.org/dc/elements/1.1/"
xmlns:dct="http://purl.org/dc/terms/"
xmlns:edm="http://www.europeana.eu/schemas/edm/"
xmlns:ore="http://www.openarchives.org/ore/terms/"
xmlns:owl="http://www.w3.org/2002/07/owl#"
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
xmlns:skos="http://www.w3.org/2004/02/skos/core#"
xmlns:wgs84="http://www.w3.org/2003/01/geo/wgs84_pos#"
xmlns:xalan="http://xml.apache.org/xalan">
  <rdf:RDF>
    <edm:ProvidedCHO rdf:about="#UEDIN:01646186">
      <dct:identifier>01646186</dct:identifier>
    </edm:ProvidedCHO>
  </rdf:RDF>
</rdfs>
```

```

    <edm:type>TEXT</edm:type>
    <dc:description>Antwortet auf eine von Brahm nicht erteilte Freigabe
des Stücks 'Schluck und Jau'. Bittet, die Entscheidung zu überdenken, da
davon "meine geistige Existenz, meine geistige Zukunft abhängt". Beklagt
sich über die von Brahm und dessen Regisseuren verschuldeten Misserfolge,
die seine "Freude am Theatralischen stark abgekühlt" und seine
"Schaffenslust nicht gesteigert haben". Fühlt sich durch die
Wiederaufnahme seiner alten Stücke "vernachlässigt und geschädigt", weil
darin "ebenso wenig die alte Liebe zu finden ist wie die alte Kunst". 'Die
versunkene Glocke' macht einen "geradezu peinlich verstaubten Eindruck".
Verweist auf den für Brahm überaus günstigen Vertrag und fordert
"Liberalität" ein. Brahm jedoch verweigere ihm "mit einer sozusagen
eisernen Hand" einen "künstlerischen Bewegungsversuch". Fühlt sich
künstlerisch missachteten Schauspielern gleich, wie dem "vortrefflichen
Rittner", "dem die Verhältnisse viel zu früh die Bühne verleidet haben".
Bittet um Freigabe von 'Schluck und Jau' für ein "Experiment auf einer
fremden Bühne". Verweist auf andere Autoren in Brahms Programm und leitet
daraus das Recht ab, 'Schluck und Jau' am Deutschen Theater aufführen zu
lassen. Einer "Wiederaufnahme im Lessingtheater, bei dem Personal, über
das Deine Bühne jetzt verfügt und bei der gleichen Regie, die den ersten
Misserfolg des Stückes nicht hat verhindern können", steht er
"hoffnungslos" gegenüber.</dc:description>
    <dc:language>de</dc:language>
    <dct:issued>10.10.1907</dct:issued>
    <dc:creator>Hauptmann, Gerhart ¬[Verfasser/in]¬</dc:creator>
    <dct:isPartOf>00032035</dct:isPartOf>
    <dct:references>Die versunkene Glocke</dct:references>
    <dct:references>Schluck und Jau</dct:references>
  </edm:ProvidedCHO>
  <edm:WebResource rdf:about="http://kalliope.staatsbibliothek-
berlin.de/cgi-
bin/aDISCGI/w3_p/lib/adis.htm?ADISDB=HS&ADISOI=01646186">
    <edm:rights rdf:resource="http://creativecommons.org/licenses/by-nc-
sa/2.0/de/">
  </edm:WebResource>
  <edm:Agent rdf:about="http://d-nb.info/gnd/118546937">
    <skos:prefLabel>Hauptmann, Gerhart ¬[Verfasser/in]¬</skos:prefLabel>
  </edm:Agent>
  <edm:Agent rdf:about="http://d-nb.info/gnd/116534788">
    <skos:prefLabel>Hauptmann, Margarete
¬[Schreiber/in]¬</skos:prefLabel>
  </edm:Agent>
  <skos:Concept rdf:about="http://d-nb.info/gnd/4008240-4">
    <skos:prefLabel xml:lang="de">Brief</skos:prefLabel>
  </skos:Concept>
  <ore:Aggregation rdf:about="http://kalliope.staatsbibliothek-
berlin.de/cgi-
bin/aDISCGI/w3_p/lib/adis.htm?ADISDB=HS&ADISOI=01646186">
    <edm:aggregatedCHO rdf:resource="#UEDIN:01646186"/>
    <edm:provider>1a</edm:provider>
    <edm:dataProvider>DE-611</edm:dataProvider>
    <edm:isShownBy rdf:resource="http://kalliope.staatsbibliothek-
berlin.de/cgi-
bin/aDISCGI/w3_p/lib/adis.htm?ADISDB=HS&ADISOI=01646186"/>
    <dct:created>20100408</dct:created>
    <dct:modified>20111019</dct:modified>
  </ore:Aggregation>
</rdf:RDF>
</rdfs>

```

SBB will go on with transformation scenarios for the following data formats and disambiguation of literals, particularly genre and places.

(b) Encoded Archival Description

Also of interest for the project is Encoded Archival Description (EAD), a sophisticated XML-schema for archival materials and thus also suitable for personal paper collections. The development of an EAD import and export interface for Kalliope, however, delayed. Though SBB provided the project with an EAD-based data example, it has not been reliable yet and could not be tested to date. The test will be done by September 2012 when the export interface will be finalised. A first crosswalk EAD to EDM was provided by the project coordinator. It was drafted for the EU-funded project APENet.

(c) Database backend of Kalliope

As mentioned in the Description of Work, SBB will also analyse the database backend of Kalliope since neither MAB2 nor EAD will make all data, particularly relations between various kinds of data associated with a catalogue entry, available. The backend is an oracle database. Analysis and provision of these data for the interoperability infrastructure will be done by September 2012.

4.2.3 Requirements

When the first crosswalk MAB2 to EDM had been worked out, missing elements in the EDM became apparent. At least for some categories new elements need to be added to EDM:

1. Authority file number for names and corporate bodies (102a|b|c/202a|b|c)
2. Document type (053)
3. Shelfmark (544)
4. Publication title (subject headings concerning work titles, 588 und 599)
5. Incipit/ Uniform incipit of the manuscript/ letter (661a|b)
6. Level of genesis (661c)
7. Condition (662a)
8. Cover (662b)
9. Watermark (662c)
10. Support (662d)
11. Restoration (662e)
12. References to editions and publications (663a|b|c)
13. Use restrictions (669a)
14. Access restrictions (669b)

In addition to that, when analysing the MAB data it turned out that indispensable information were missing, e.g. custodial history, old shelfmarks (544b), or unique IDs for genres (661d). These data are stored in the database of Kalliope. In line with the project proposal, SBB is thus going to analyse the backend database of Kalliope which will bring about more precise data for the interoperability infrastructure. Due to delays in employing a project staff member, this work will be carried out in July, August, and September 2012.

4.2.4 Summary

Three basic challenges were identified:

1. A data model for personal paper collections and archival materials needs to consider a multilevel description approach (collection, sub-classes, and items).
2. Personal paper collections usually contain a wide variety of materials (manuscripts, letters, files, documents, photos, etc.) and thus require the provision of a broader set of database fields for in-depth cataloguing within the cataloguing application. This in turn requires additional elements in EDM.

As a result, it is to be underscored that:

The EDM model needs to be extended as mentioned beforehand and the database of Kalliope has to be analysed in more detail to provide all information necessary for the interoperability infrastructure developed by Work Package 2. This will be done by SBB in the months ahead.

Concerning the interoperability infrastructure, it is the objective of SBB to contextualise the following data with resources available in the Linked Data Cloud:

- Authority files for persons, corporate bodies, subject headings (for the first two authority types SBB provides authority numbers (MAB 102a or at least Kalliope database internal ID)
- Geonames for places (currently plain literals, 410)
- Genres (currently plain literals, 661d)
- Bibliographic information (edition, references to publications are also currently plain literals, 663a, b, c)
- Publication titles (Subject headings concerning work titles, 588/599)

Disambiguation of these data would significantly enhance user experience.

4.2.5 Preliminary mapping of mandatory EDM fields

Property	Property range	Property description	Source Element	Source Element Description
ore:Aggregation				
edm:provider	edm:Agent entity	Mandatory element: Organisation (edm:Agent, see below) that provided this aggregation	Default value "DM2E"	
edm:dataProvider	edm:Agent entity	Mandatory element: Organisation (edm:Agent, see below) that provided the base data for this aggregation (could be the same as edm:provider)	070a / 071	Codes/ISIL (always DE-611); Codes/ISIL (DE-1a)
edm:rights	remote resource	Mandatory element: URL of a resource describing licensing rights from http://pro.europeana.eu/documents/900548/1037382/Europeana_rights_201202.pdf	Will be a default value.	BY-NC-SA
edm:isShownBy	WebResource entity	Either edm:isShownBy or edm:isShownAt is mandatory. Web resource (see below) describing this specific ProvidedCHO aggregated here	655e\$u	URL

Property	Property range	Property description	Source Element	Source Element Description
edm:isShownAt	WebResource entity	Either edm:isShownBy or edm:isShownAt is mandatory. Web resource (see below) containing the ProvidedCHO aggregated here	-	
dct:created	xsd:dateTime literal	Creation and date of this aggregation	002a	
dct:modified	Xsd:dateTime literal	Modification date and time of this aggregation	003	
edm:ProvidedCHO				
edm:type	Literal	Mandatory element: Can be TEXT, VIDEO, SOUND, IMAGE, 3D	Default value: "TEXT"	
dc:type	Specialization of ore:Proxy	Either dc:type or dc:subject is mandatory. The most specific type that applies to the CHO, repetition of rdf:type of Proxy	661d	Genre
dc:title	Literal with language tag	Either dc:title or dc:description is mandatory. Title of the object.	331	
dct:alternative	Literal with language tag	Alternative title	370	Additional tile
dc:description	Literal with language tag	Either dc:title or dc:description is mandatory. Description of the CHO	750m	abstract
dc:language	xsd:lang literal	Mandatory for textual objects. Most prominent language of the CHO	037b	ISO 639-2
dct:issued	xsd:dateTime literal	Date of publication, also possible: dct:created, or dc:date for generic cases	426a	
dc:creator	edm:Agent entity	Creator of the CHO, possibly its author. Subproperties may be used to specialize relationship	100 104a 108a etc 200 204a 208a etc.	100 person (creator) 200 corporate body (creator)
dc:identifier	Literal	Provider-local identifier of the CHO, can be specialized, e.g. as dm2e:isbn	001	
dc:subject	Resource	Either dc:type or dc:subject is mandatory. Resource in SKOS-Concept-Tree, e.g. Dewey Decimal Classification	710a	rare use on collection level, no use at item level
dct:provenance	Literal	Description of provenance of CHO	664	
dct:hasPart	ProvidedCHO entity	Reference to a part of this CHO, e.g. a chapter of a book	453	ID of superordinate record
dct:references	ProvidedCHO entity	Other CHO referenced in the content of this CHO 588reference on published titles	588	reference on published titles

Table 5: EDM mapping SBB

Preliminary namespace for records identifier (001 and 453) will be:

[http://kalliope.staatsbibliothek-berlin.de/cgi-bin/aDISCGI/w3_p/lib/adis.htm?ADISDB=HS&ADISOI=\[RecordID\]](http://kalliope.staatsbibliothek-berlin.de/cgi-bin/aDISCGI/w3_p/lib/adis.htm?ADISDB=HS&ADISOI=[RecordID])

It will be replaced during the project by a persistent Kalliope-URI.

4.3 UIB – University of Bergen (Universitetet i Bergen)

4.3.1 Underlying content

The Wittgenstein Archives at the University of Bergen (WAB) contributes 5000 pages of the Wittgenstein Nachlass kept at Trinity College Cambridge (TCC), Wren Library, Cambridge, in facsimile and text editions as well as metadata.

The 5000 pages include:

- Ts-201a1 and Ts-201a2 (1913-14) from the "Notes on Logic" corpus: This contains the first expression of Wittgenstein's view on philosophy and specific material on logic.
- Ms-139a and Ts-207 (1929) from the "Lecture on Ethics" corpus: The "Lecture on Ethics" is the only "popular" lecture which Wittgenstein ever held. It is a masterpiece expressing his early views on tensions between meaningful language and nonsensical language in which we expose ourselves when speaking about ethics.
- Ms-114, Ms-115 (first part), Ms-148, Ms-149, Ms-150, Ms-153a, Ms-153b, Ms-154, Ms-155, Ms-156a, Ts-212 and Ts-213 (1931-34) from the "Big Typescript" corpus: The Big Typescript is the great "summa" of Wittgenstein's latest thoughts elaborated from 1929 to 1933. It addresses most philosophical subjects Wittgenstein has ever been interested in.
- Ms-115 (second part), Ms-140 (p. 39v), Ms-141, Ms-152 and Ts-310 (1934-36) from the "Brown Book" corpus: The "Brown Book" was first lectured to students in Cambridge and is thus laid out very pedagogically. One aspect is the introduction of "language games" which shed light on the complexity of our language. The Book was revised several times. The last revision led to the Philosophical Investigations.

In terms of philosophical development, the 5000 pages include four high points in Wittgenstein's philosophical development from 1913 to 1936; in terms of philosophical themes, they encompass all themes addressed by Wittgenstein incl. philosophy of language, logics, mathematics, psychology, ethics, metaphilosophy, etc.; in terms of type of Nachlass material they include all kinds of manuscripts written by Wittgenstein: first drafts, lecture notes, notebooks, copybooks, typescript cuttings, and elaborated materials such as typescripts, and materials prepared in cooperation with others. In terms of language, they contain both German and English materials.

4.3.2 Requirements

The main metadata format of the content provided by UIB is TEI P5 (see table on page 11). The following elements refer to the TEI P5 guidelines <http://www.tei-c.org/Guidelines/P5>.

The following list contains the tags in TEI P5 which need to be represented in EDM:

TEI Header

Requirements from the TEI header:

- <author>
- <availability> - edm rights; information about copyright holders needs to be included

- <editor><orgName>
- <idNo>
- <sourceDesc>
- <sponsor>
- <term>
- <title>

Possible additional requirements: With regard to quality and acknowledgement matters, we consider adding information about editors and transcribers.

- <editor><persName> Responsible editor
- <respStmt><name> Staff contributing with transcription and other work

Example

```
<teiHeader>

  <fileDesc>

    <titleStmt>
      <title>Wittgenstein Nachlass MS 141: Ms-141.xml - A machine-
readable transcription [Work in progress]</title>
      <author>Ludwig Wittgenstein</author>
      <editor><persName>Alois Pichler</persName> <orgName
ref="http://wab.uib.no/">Wittgenstein Archives at the University of
Bergen (WAB)</orgName>
      </editor>
      <sponsor><name>Trinity College, Cambridge; Oxford University
Press, Oxford; IntelX Corporation, Charlottesville; Uni Research,
Bergen; University of Bergen, Bergen; L. Meltzers H&oslash;yskolefond,
Bergen; COST Action A32, Brussels; eContent+ DISCOVERY,
Luxembourg</name></sponsor>
      <respStmt><name>Alois Pichler</name> <resp>coordination,
editorial guidelines, XML-TEI markup</resp></respStmt>
      <respStmt><name>Claus Huitfeldt, Kjersti Bj&oslash;rnestad Berg,
Sindre S&oslash;rensen, MLCD project</name>
      <resp>conversion from MECS-WIT to flat XML markup:
parser</resp></respStmt>

      <respStmt><name>Alois Pichler</name>
      <resp>conversion from MECS-WIT to flat XML markup: handling of
overlap</resp></respStmt>
      <respStmt><name>Vemund Olstad</name>
<resp>stylesheets</resp></respStmt>
      <respStmt><name>Tone Merete Bruvik</name> <resp>XML-TEI
validation</resp></respStmt>
      <respStmt><name>&Oslash;ystein Reigem</name>
<resp>PHP</resp></respStmt>
      <respStmt><name>Alois Pichler, Heinz Wilhelm Kr&uuml;ger,
Deirdre C. P. Smith, Florian Gst&ouml;hl</name> <resp>post BEE 2000
amendments, incl. further markup and XML-TEI markup (European
DISCOVERY project 2006-09; European COST Action A32 project 2005-
10)</resp></respStmt>
      <respStmt><name>Espen Ore, Kyrre Trohjell, Heinz Wilhelm
Kr&uuml;ger, J&uuml;rgen Gercken</name> <resp>transcription and MECS-
WIT markup: 1995-99</resp></respStmt>
```



```

</titleStmt>

<publicationStmt>
  <idno type="URL">http://www.wittgensteinsource.org/texts/BTEn/Ms-141</idno>
  <availability>
    <p>Copyright holders: The Master and Fellows of Trinity College, Cambridge; Oxford University Press, Oxford; University of Bergen, Bergen. Released under the Creative Commons General Public License Attribution, Non-Commercial, Share-Alike version 3 (CCPL BY-NC-SA).</p>
  </availability>
</publicationStmt>

<sourceDesc><p>Loose sheets Ms-141 (1933-34): The beginning of a version of the Brown Book in German. -- The manuscript is written in pen and contains 10 loose sheets similar to those used for Ms-140, measuring 300 x 470 mm, of which only seven pages contain text and one drawings. The original is in the Wren Library, Trinity College Cambridge.</p></sourceDesc>

</fileDesc>

<profileDesc><textClass><keywords
scheme="http://philpapers.org/"><term>Philosophy of Language</term></keywords></textClass></profileDesc>
  <revisionDesc>
    <change>Dating with reference to Pichler 1997, p&p; 65f (Spaetherbst 1933 bis Oktober 1934). </change>
  </revisionDesc>

</teiHeader>

<text xml:lang="de">

  <body>
    <ab xml:lang="de" n="Ms-141,1[1]" ana="pub:W-000_date:19331000-19341000" rend="blbef_0">
</code></pre>

```

TEI Body

Due to the rich information of TEI and the annotation of the Wittgenstein material, special information needs to be mapped from the TEI body. These are special annotations referring to Wittgenstein's remarks ("Bemerkungen"):

- text xml:lang= refers to the language

Relevant information from the <ab> tag, which is

- xml:lang: language
- n: relevant ID
- ana/pub: publication
- ana/date: date of remark (Bemerkung)
- persName key: person referred to
- rs type="extref" key: external primary source referred to

Examples

```
<ab xml:lang="de" n="Ms-141,11" ana="pub:W-000_date:19331000-19341000"
... <s type="es" rend="incl_4"><persName key="Augustinus, Aurelius"
corresp="commentary"> ... <rs type="extref" key="Augustinus, Aurelius:
Confessiones" ...
<ab n="Ts-213,100r2" ana="pub:W-PG1969:PartI:App:4_date:19320700-
19331000" rend="blbef_2">
<rs type="extref" key="Carnap, Rudolf: Der logische Aufbau der Welt"
n="1928" xml:id="Biesenbach_Carnap1">
<persName key="Galton, Francis" corresp="commentary">
```

4.3.3 Preliminary mapping of mandatory EDM fields

Property	Property range	Property description	Source Element	Source Element Description
ore:Aggregation				
edm:provider	edm:Agent entity	Mandatory element: Organisation (edm:Agent, see below) that provided this aggregation	Default value "DM2E"	
edm:dataProvider	edm:Agent entity	Mandatory element: Organisation (edm:Agent, see below) that provided the base data for this aggregation (could be the same as edm:provider)	<editor><org Name>	
edm:rights	remote resource	Mandatory element: URL of a resource describing licensing rights from http://pro.europeana.eu/documents/900548/1037382/Europeana_rights_201202.pdf	Will be a default value: "CCPL BY-NC-SA"	NB: The copyright holders need to be mentioned.
edm:isShownBy	WebResource entity	Either edm:isShownBy or edm:isShownAt is mandatory. Web resource (see below) describing this specific ProvidedCHO aggregated here		
edm:isShownAt	WebResource entity	Either edm:isShownBy or edm:isShownAt is mandatory. Web resource (see below) containing the ProvidedCHO aggregated here	idno	
edm:ProvidedCHO				
edm:type	Literal	Mandatory element: Can be TEXT, VIDEO, SOUND, IMAGE, 3D	Default value: "TEXT"	
dc:type	Specialization of ore:Proxy	Either dc:type or dc:subject is mandatory. The most specific type that applies to the CHO, repetition of rdf:type of Proxy		
dc:title	Literal with language tag	Either dc:title or dc:description is mandatory. Title of the object.	title	
dc:description	Literal with language tag	Either dc:title or dc:description is mandatory. Description of the CHO	sourceDesc	

Property	Property range	Property description	Source Element	Source Element Description
dc:language	xsd:lang literal	Mandatory for textual objects. Most prominent language of the CHO	text xml:lang	Is additionally specified on <ab level, i.e. for each single Bemerkung.
dc:subject	Resource	Either dc:type or dc:subject is mandatory. Resource in SKOS-Concept-Tree, e.g. Dewey Decimal Classification	profileDesc	

Table 6: EDM mapping UIB

4.4 UBER – Humboldt University Berlin (Humboldt Universität zu Berlin)

4.4.1 Underlying content

In 1820 the German chemist and industrialist J. G. Dingler first started publishing the »Polytechnisches Journal«. The journal was published over a period of 111 years and has hence become an important and European wide source for the history of knowledge, culture, and technology — at least in Germany it is unparalleled.

The project is a cooperation between the Institute for Cultural Studies (Institut für Kulturwissenschaft) at the Humboldt-Universität zu Berlin (Humboldt University Berlin), the Sächsische Landesbibliothek — Staats- und Universitätsbibliothek Dresden (SLUB – Saxon State and University Library Dresden) and an experienced service provider for digitisation, Editura GmbH. The institutions aim at digitising Dingler's »Polytechnisches Journal« as well as publishing it completely on the web. Access to Dingler-Online will be free of charge.

Besides digitisation of the images also the OCRred text is consequently encoded according to the Text Encoding Initiative Guidelines TEI-P5. On the one hand, this will enable to grasp the editors' true spirit of literally collecting poly-technical works; on the other hand, a very comfortable access to the articles can be provided.⁸

4.4.2 Requirements

The EDM does not further distinguish titles and people (title into sub- and main title, person into editor and author). Other missing elements are:

- Page and volume numbers
- Publishing place
- Publishing date of an issue
- Further person roles (cooperation partner, supporter, publisher)
- Notes

Missing elements regarding the mandatory mapping that were encountered are:

⁸ <http://www.polytechnischesjournal.de/en/home> [Retrieved 27.07.2012].

-
- edm:ProvidedCHO
 - dc:coverage, dc:type, dc:subject or dct:spatial
 - dc:language (can be replaced by a default value)
 - edm:type (default value „TEXT“)

The following figures illustrate an example mapping of ar001001.xml (created with VUE). Blue entries suggest new classes that are not part of the EDM:

4.4.3 Preliminary mapping of mandatory EDM fields

Property	Property range	Property description	Source Element	Source Element Description
ore:Aggregation				
edm:provider	edm:Agent entity	Mandatory element: Organisation (edm:Agent, see below) that provided this aggregation	Default value "DM2E"	
edm:dataProvider	edm:Agent entity	Mandatory element: Organisation (edm:Agent, see below) that provided the base data for this aggregation (could be the same as edm:provider)	<teiHeader><publicationStatement><publisher>	
edm:rights	remote resource	Mandatory element: URL of a resource describing licensing rights from http://pro.europeana.eu/documents/900548/1037382/Europeana_rights_201202.pdf	<teiHeader><publicationStatement><availability status="free"><p><ref target="">	
edm:isShownBy	WebResource entity	Either edm:isShownBy or edm:isShownAt is mandatory. Web resource (see below) describing this specific ProvidedCHO aggregated here		
edm:isShownAt	WebResource entity	Either edm:isShownBy or edm:isShownAt is mandatory. Web resource (see below) containing the ProvidedCHO aggregated here	<teiHeader><publicationStatement><idno type="URL">	
edm:ProvidedCHO				
edm:type	Literal	Mandatory element: Can be TEXT, VIDEO, SOUND, IMAGE, 3D	default value: "TEXT"	
dc:type	Specialization of ore:Proxy	Either dc:type or dc:subject is mandatory. The most specific type that applies to the CHO, repetition of rdf:type of Proxy		
dc:title	Literal with language tag	Either dc:title or dc:description is mandatory. Title of the object.	<teiHeader><titleStatement><title type="sub">	
dc:description	Literal with language tag	Either dc:title or dc:description is mandatory. Description of the CHO		
dc:language	xsd:lang literal	Mandatory for textual objects: Most prominent language of the CHO	default value: German	
dc:subject	Resource	Either dc:type or dc:subject is mandatory. Resource in SKOS-Concept-Tree, e.g. Dewey Decimal Classification		

Table 7: EDM mapping UBER

4.5 MPIWG – Max Planck Institute for the History of Science (Max-Planck-Institut für Wissenschaftsgeschichte)

4.5.1 General information

The Max Planck Institute for the History of Science (MPIWG) will provide a number of its collections to the DM2E, as it is stated in Table 1: Underlying Content (see page 10). So far, the main work focus was on two collections which will be described below: metadata of the MPIWG's rare book collection and already semantically annotated metadata of the ISMI project.

Additional collections are also in the pipeline and will be analysed as well: a semantically enriched version of the finished project "[Virtual Laboratory](#)"⁹, the ongoing project of editing the manuscripts of [Thomas Harriot](#)¹⁰, and a project about editing painters' recipes for colours.

Following general remarks about the extension of the EDM to the MPIWG's needs, two collections, along with mappings of mandatory fields to the EDM, will be presented in detail.

4.5.2 Extensions to EDM

In general, the EDM should be extended in order to describe parent and child objects. The metadata for one object could be inherited from another one. For example, significant parts of the MPIWG's digital objects (e.g. <http://echo.mpiwg-berlin.mpg.de>, http://vlp.mpiwg-berlin.mpg.de/index_html) are organised in collections. These collections describe the scientific domain or in which research context the sources are used. This includes information, e.g. on who is owner of the collection. That is, wider contexts etc. are only connected to these collections and not the cultural objects in such a collection. There are also objects – in particular archival material – where only the folder has metadata (e. g. shelfmarks), and the objects within the folder have no metadata yet.

Therefore, a collection-object with full metadata inheritance of metadata from collections to objects contained would be helpful.

A similar structure is present in the relation of Codices and Manuscripts (e.g. <https://it-dev.mpiwg-berlin.mpg.de/tracs/OpenMind3/wiki/IsmiDataModel>). Here, the codex carries information like place, collection, shelf-mark, year of compilation, but the manuscripts contain metadata like author, title, etc. It is essential to extend the EDM in the way described above to keep the data consistent.

Additionally, there are objects which have more than one web resource/ aggregation with different copyrights, creators and content providers. This could be a book from which both digital facsimiles (provided by another library) and transcriptions (made by a scholar) are present. However, they form a unity so that they have to be directly connected to the same CHO.

Furthermore, the EDM should be extended with agents, places and subjects as objects with spatio-temporal metadata describing the contexts like the duration of existence, successor and predecessor.

⁹ http://vlp.mpiwg-berlin.mpg.de/index_html.

¹⁰ <http://echo.mpiwg-berlin.mpg.de/content/search?searchSimple=harriot>.

Another important issue is the inclusion of administrative metadata: who created which (meta-)data set. This could be used on the one hand to deal with mis-attributions, different opinions and the like; on the other hand, it might constitute the act of academic authorship.

In the same vein goes the extension of the metadata structure to describe purely digital objects such as digital editions, websites, and databases as CHO. A set of standards for the integration of these objects into the Europeana would have to be developed.

4.5.3 The MPIWG's rare book collection

4.5.3.1 General remarks

This collection comprises the metadata of the institute's rare book collection which consists of approximately 2.500 volumes of which most are also available in digital form. Additionally, digitised full text is present for about 100 volumes. Metadata which will be provided to Europeana will be the set of digitised rare books.

4.5.3.2 Contents of the metadata

Metadata for each of the rare books is stored in a file called "index.meta". This file resides in a directory on the MPIWG servers. It contains the metadata of the respective CHO as well as the paths to the page scans and to the digitised text (if available).

A documentation for this file format exists (<http://intern.mpiwg-berlin.mpg.de/digitalhumanities/mpiwg-metadata-documentation/formate/indexmeta-standard>) as well as an XML schema (in RelaxNG compact).

What follows is a complete example of an index.meta file. There are pairs of fields that contain more or less the same information: "author"/ "person", "year"/ "date" and "city"/ "place", "author", "year" and "city" are included for backward compatibility.

index.meta file example

```
<?xml version="1.0" encoding="UTF-8"?>
<resource type="MPIWG" version="2.0">
  <description />
  <archive-creation-date>03.05.2007</archive-creation-date>
  <name>163127KK</name>
  <archive-path>/mpiwg/online/permanent/library/163127KK</archive-
path>
  <creator>Max Planck Institute for the History of Science,
Library</creator>
  <content-type>scanned document</content-type>
  <derived-from>
    <archive-path>/mpiwg/archive/data/library/NRNNU192</archive-
path>
  </derived-from>
  <meta>
    <bib type="book">
      <author>Benedetti, Giovanni Battista de</author>
      <person role="author" type="person">
        <name>Benedetti, Giovanni Battista de</name>
        <identifier type="GND">http://d-
nb.info/gnd/118859676</identifier>
```



```

    </person>
  <lang>lat</lang>
  <number_of_pages>[7], 426, [2] S.: Ill.</number_of_pages>
  <title lang="lat">Io. Baptistae Benedicti ... Diversarvm
specvlationvm mathematicarum, et physicarum liber: quarum seriem
sequens pagina indicabit; [annotated and critiqued by Guidobaldo Del
Monte]</title>
  <title type="alternate">Io. Baptistae Benedicti ...
Diversarum speculationum mathematicarum, et physicarum liber: quarum
seriem sequens pagina indicabit; [annotated and critiqued by
Guidobaldo Del Monte]</title>
  <year>1585</year>
  <date>1585</date>
  <place>Bevilaqua</place>
  <city>Bevilaqua</city>
  <holding-library>Max Planck Institute for the History of
Science, Library</holding-library>
</bib>
<image-acquisition>
  <image-type>rgb</image-type>
  <production-comment>Images shot in RAW-format (CR2).
Converted into DNG and Tiff-format by using Adobe CamerRaw. Image
Quality has been enhanced by applying the Photoshop action (atn)
located in the info-folder.</production-comment>
</image-acquisition>
<img>
  <original-dpi>300</original-dpi>
</img>
<access-conditions>
  <access type="free" />
  <attribution type="original">
    <name>Max Planck Institute for the History of Science,
Library</name>
    <url>http://library.mpiwg-berlin.mpg.de/</url>
  </attribution>
  <attribution type="digital-image">
    <name>Max Planck Institute for the History of Science,
Library</name>
    <url>http://library.mpiwg-berlin.mpg.de/</url>
  </attribution>
  <copyright type="original">
    <owner>
      <name>Max Planck Institute for the History of
Science, Library</name>
      <url>http://library.mpiwg-berlin.mpg.de/</url>
    </owner>
  </copyright>
  <copyright type="digital-image">
    <owner>
      <name>Max Planck Institute for the History of
Science, Library</name>
      <url>http://library.mpiwg-berlin.mpg.de/</url>
    </owner>
    <license type="cc-by-sa">Dieses Werk steht unter der
CC-BY-Lizenz.
      <url>http://creativecommons.org/licenses/by-
sa/3.0/de/</url>

```

```

        </license>
    </license></license>
</copyright>
</access-conditions>
<texttool>
    <image>pageimg</image>
    <figures>figures</figures>
    <text-url-path>/echo/1a/Benedetti_1585.xml</text-url-path>
</texttool>
<dri type="escidoc-test">escidoc:81176</dri>
</meta>
</resource>

```

4.5.3.3 Preliminary mapping of mandatory EDM fields:

Property	Property range	Property description	Source Element	Source Element Description
ore:Aggregation				
edm:provider	edm:Agent entity	Mandatory element: Organisation (edm:Agent, see below) that provided this aggregation	resource/meta/access-conditions/attribute/name or default value "DM2E"	
edm:dataProvider	edm:Agent entity	Mandatory element: Organisation (edm:Agent, see below) that provided the base data for this aggregation (could be the same as edm:provider)	resource/creator	
edm:rights	remote resource	Mandatory element: URL of a resource describing licensing rights from http://pro.europeana.eu/documents/900548/1037382/Europeana_rights_201202.pdf	resource/meta/access-conditions/copyright/license/url	
edm:isShownBy	WebResource entity	Either edm:isShownBy or edm:isShownAt is mandatory. Web resource (see below) describing this specific ProvidedCHO aggregated here		
edm:isShownAt	WebResource entity	Either edm:isShownBy or edm:isShownAt is mandatory. Web resource (see below) containing the ProvidedCHO aggregated here	http://echo.mpiwg-berlin.mpg.de/ECHOdocuViewfull?url= + resource/archived-path	The first part, up to "url=" is a constant, followed by XPath (the part after "+ ")
edm:ProvidedCHO				
edm:type	Literal	Mandatory element: Can be TEXT, VIDEO, SOUND, IMAGE, 3D	Default value "TEXT"	Constant value
dc:type	Specialization of ore:Proxy	Either dc:type or dc:subject is mandatory. The most specific type that applies to the CHO, repetition of rdf:type of Proxy	resource/meta/bib/@type	

Property	Property range	Property description	Source Element	Source Element Description
dc:title	Literal with language tag	Either dc:title or dc:description is mandatory. Title of the object.	resource/meta/bib/title	resource/meta/bib/title/@lang specifies the language
dc:description	Literal with language tag	Either dc:title or dc:description is mandatory. Description of the CHO	resource/meta/bib/title	resource/meta/bib/title/@type=alternate specifies alternate title
dc:language	xsd:lang literal	Mandatory for textual objects: Most prominent language of the CHO	resource/meta/lang	
dc:subject	Resource	Either dc:type or dc:subject is mandatory. Resource in SKOS-Concept-Tree, e.g. Dewey Decimal Classification	resource/meta/bib/keywords /	resource/meta/bib/keywords/@type specifies the vocabulary. We use <u>LCSH</u> and <u>DDC</u>

Table 8: EDM mapping MPIWG rare books collection

4.5.4 ISMI Project

ISMI is an international collaborative endeavour that is working to make available a vast array of information online about the exact sciences in the premodern Islamic world.

4.5.4.1 Structure of the metadata

Metadata for ISMI is stored in the Openmind Framework, which is a graph database and has been developed by MPIWG. This database contains Nodes (also called entities) and directed relationships, where both of them have properties (also called attributes).

The specification of ISMI defines the following kind of entities: Text, Alias, Witness (Manuscript), Codex, Collection, Repository, Person, Place, Role, Subject, and Reference; and this set is continuously growing.

4.5.4.2 Preliminary mapping of mandatory EDM fields

Witnesses (manuscripts) are the principal CHO to be exported to EDM and represent concrete written pieces of paper lying in a library. Furthermore, there are abstract texts (mostly with a specific title), which encapsulate common data from one or more witnesses. Because of this structure, not every mandatory field of the current ProvidedCHO can be found directly in the witness. Sometimes, the field should be taken from elements related to the witness.

For example the value for dc:title: A witness has no title's property associated, but the abstract text has the field ismi:fullTitle, which corresponds to the dc:title.

Therefore, the source element should be obtained using the following query:

```
dc:title = ismi:witness – ismi:is_exemplar_of - ismi:text - ismi:fullTitle
```

For this situation, two columns will be inserted in the 'edm:ProvidedCHO' table that should facilitate the understanding of mapping: "Domain" and "Range". The real entity, which contains the "source element", will be described in the column domain.

The range can contain two kinds of values: an entity or content type (like: rdfs:Literal). The range contains an entity if the property (like dc:creator) associates two entities (like: edm:ProvidedCHO dc:creator edm:Agent).

Property	Range	Description	Source Element	Domain	Range	Source Element Description
ore:Aggregation						
edm:provider	edm:Agent entity	Mandatory element: Organisation (edm:Agent, see below) that provided this aggregation	Default value "DM2E"			Constant
edm:dataProvider	edm:Agent entity	Mandatory element: Organisation (edm:Agent, see below) that provided the base data for this aggregation (could be the same as edm:provider)	Default value "MPIWG"			Constant
edm:rights	remote resource	Mandatory element: URL of a resource describing licensing rights from http://pro.europeana.eu/documents/900548/1037382/Europeana_rights_201202.pdf	ismi:rights			
edm:isShownBy	WebResource entity	Either edm:isShownBy or edm:isShownAt is mandatory. Web resource (see below) describing this specific ProvidedCHO aggregated here	ismi:uriDigitilizedCodex		ismi:codex	URL of an image
edm:isShownAt	WebResource entity	Either edm:isShownBy or edm:isShownAt is mandatory. Web resource (see below) containing the ProvidedCHO aggregated here				

Property	Range	Description	Source Element	Domain	Range	Source Element Description
edm:ProvidedCHO (ismi:witness)						
edm:type	Literal	Mandatory element: Can be TEXT, VIDEO, SOUND, IMAGE, 3D	Default value "TEXT"			Constant
dc:type	Specialization of ore:Proxy	Either dc:type or dc:subject is mandatory. The most specific type that applies to the CHO, repetition of rdf:type of Proxy	Default value "Manuscript"			Constant
dc:title	Literal with language tag	Either dc:title or dc:description is mandatory. Title of the object.	ismi:fullTitle	ismi:text	Literal	ismi:witness -> ismi:isExemplarOf-> ismi:text->
dc:description	Literal with language tag	Either dc:title or dc:description is mandatory. Description of the CHO	ismi:notes	ismi:witness ismi:codex ismi:text	Literal	We insert a description for the witness in three different level: directly in the witness, in the related codex and in the related text. The three sources will be combined together into the field dc:description.
dc:language	xsd:lang literal	Mandatory for textual objects. Most prominent language of the CHO	ismi:language	ismi:text	Literal	ismi:witness -> ismi:isExemplarOf-> ismi:text->
dc:subject	Resource	Either dc:type or dc:subject is mandatory. Resource in SKOS-Concept-Tree, e.g. dewey classification	ismi:subject	ismi:text	Literal	ismi:witness -> ismi:isExemplarOf-> ismi:text

Table 9: EDM mapping MPWIG ISMI project

4.5.4.3 Challenges with mapping to EDM

The principal object to be export is the manuscript (called in ISMI: Witness). A witness is equivalent to edm:ProvidedCHO, however there are several challenges mapping from ISMI to EDM. The following diagram helps to understand the issues:

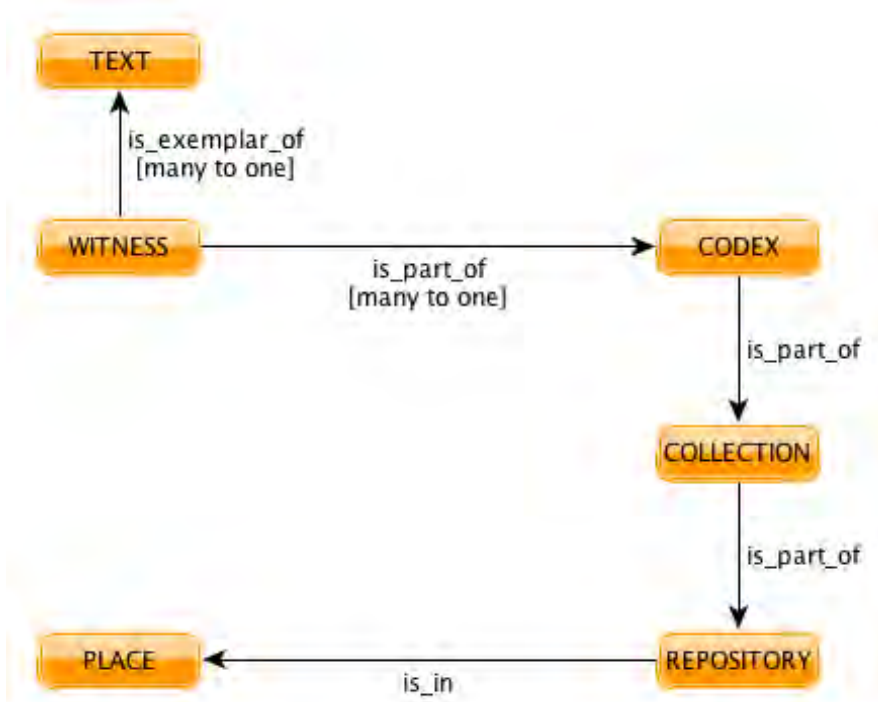


Figure 2: Diagram ISMI mapping

Either dc:title or dc:description is mandatory. For ISMI, a manuscript never has a title, but witness has a text associated, which has a title. Many witnesses can be associated to the same text; therefore, it would be beneficiary if the title could be provided as a common object (like the abstract text) and not directly to the manuscript.

Manuscripts are associated in codices or physical bound collections. For this reason, with the current vocabulary of EDM, every attribute of a codex should be repeated for every manuscript.

In ISMI, a place is not associated to a manuscript, but a codex is in a repository (like a library) and this repository has a concrete location/place. It is important to consider that in ISMI every manuscript has a codex.

The following table lists the object properties and the data properties that cannot be represented by EDM:

Object Properties

Domain	Range	Property
ismi:person	ismi:person	was_influenced_by
ismi:person	ismi:person	was_student_of
ismi:witness	ismi:person	was_taught_by
ismi:witness	ismi:person	was_studied_by
ismi:witness	ismi:person	was_owned_by
ismi:text	ismi:person	misattributed_to

Domain	Range	Property
ismi:place	ismi:place	isPartOf

Table 10: ISMI Object Properties

Data Properties

Domain	Range	Property	Description
ismi:witness	Literal	incipit	Opening notes
ismi:witness	Literal	explicit	Final words
ismi:witnes	Literal	colophon	A brief description of manuscript
ismi:witness	Literal	page_dimensions	
ismi:witness	Literal	written_area_dimensions	

Table 11: ISMI Data Properties

Relations with Attributes

This is a special case. It is possible for the ISMI scheme that a relation contains attributes. Examples are additional object properties: ismi:person, ismi:livedIn, and ismi:place. These attributes indicate the related period and place in which the person lived.

4.5.4.4 Conclusions and Requirements

Due to the fact that a set of manuscripts can have common attributes (such as title, create date, author, etc), an object is required that groups manuscripts and contains the common attributes of them (such as ismi:text). Although every mandatory field of EDM can be delivered from ISMI, from the MPIWG's perspective it is not a satisfactory solution, because fields like dc:title would be redundant in Europeana, when they are exported from ISMI.

From MPIWG's point of view, the properties title or description should not be mandatory for edm:ProvidedCHO.

An object is needed that describes a bound collection of manuscripts (such as ismi:codex).

Misattribution is a relationship between text and person whose semantically meaning indicates that somebody believes or believed that a specific person is the author of a text. This relation should also consider a property to store the reference of this information.

4.6 NLI – National Library of Israel

4.6.1 Underlying content

The National Library of Israel (NLI) has a respectable digital collection created over the past decade including various material types. Metadata for digitised maps, kettuboth (Jewish marriage contracts), and some books and manuscripts was provided to Europeana

via previous projects. Since early 2012 NLI has a new digitisation centre and has engaged in a major effort to digitise vast parts of the libraries diverse collections.

It was decided not to include the newspaper collections via DM2E at this point and time. However, NLI was able to enlarge the number of books and manuscripts slightly and add an archival collection of photographic portraits and thus the total number of images adds up to 2.907.000 as stated in the DoW. Following the changes in the selection of the materials, the records will include Hebrew and European language books; Hebrew, Arabic and European language manuscripts; as for the photographs language is not applicable.

4.6.2 Requirements

NLI has provided metadata in MARCXML for books and manuscripts. For the portraits which will replace the newspaper collection originally intended, no sample data has been provided yet, but the structure of the data is very similar to that of the books and manuscripts.

For the books as well as for the manuscripts examples, the missing mandatory elements for the EDM mapping were edm:rights (ore:Aggregation) and edm:provider. As the value for this element will be constant for entire collections, NLI will either add it globally to all records before the upload or if the MINT tool will allow for this a default value can be added during the mapping.

For the manuscripts as well the edm:isShownBy / edm:isShownAt (ore:Aggregation) were missing. This is due to the manifestation of the digitised object in the cataloguing system not as link in a MARC field but rather as ADAM object (an ExLibris propriety format). However, since the delivery of the sample data NLI has defined an expand mechanism which can be applied during export of the data to the link to MARCXML format.

Sample record books

```
<record xmlns="http://www.loc.gov/MARC21/slim"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.loc.gov/MARC21/slim
http://www.loc.gov/standards/marcxml/schema/MARC21slim.xsd">
  <leader>00000nam a22 a 4500</leader>
  <controlfield tag="001">001124667</controlfield>
  <controlfield tag="005">20061019112953.0</controlfield>
  <controlfield tag="008">050527s1480 ||| r 000 0 heb
d</controlfield>
  <datafield tag="035" ind1=" " ind2=" ">
    <subfield code="a">(OCoLC)233044456</subfield>
  </datafield>
  <datafield tag="035" ind1=" " ind2=" ">
    <subfield code="a">1124667NNL</subfield>
  </datafield>
  <datafield tag="041" ind1=" " ind2=" ">
    <subfield code="a">heb</subfield>
  </datafield>
  <datafield tag="082" ind1=" " ind2=" ">
    <subfield code="a">296.571(46)(090.2)</subfield>
    <subfield code="b">Sheelot utshuvot, responsa - Rishonim -
Spain - Before 1492</subfield>
  </datafield>
  <datafield tag="090" ind1=" " ind2=" ">
```



```

    <subfield code="a">SHOCKEN 8=RI</subfield>
  </datafield>
  <datafield tag="090" ind1=" " ind2=" ">
    <subfield code="a">FI 2322</subfield>
  </datafield>
  <datafield tag="100" ind1="1" ind2=" ">
    <subfield code="a">שלמה בן אברהם אבן (רשב) א, </subfield>
    <subfield code="d">1235-1310.</subfield>
  </datafield>
  <datafield tag="245" ind1="1" ind2="0">
    <subfield code="a">תשובות שאלות להרשב א.</subfield>
  </datafield>
  <datafield tag="246" ind1="3" ind2="3">
    <subfield code="a">תשובות שאלות להרשב א.</subfield>
    <subfield code="h">[משאב אלקטרוני]</subfield>
  </datafield>
  <datafield tag="260" ind1=" " ind2=" ">
    <subfield code="a">[רומא :</subfield>
    <subfield code="b">ל; חמו </subfield>
    <subfield code="c">מ; קודם ר.</subfield>
  </datafield>
  <datafield tag="300" ind1=" " ind2=" ">
    <subfield code="a">[160] דף ;</subfield>
    <subfield code="c">8.</subfield>
  </datafield>
  <datafield tag="500" ind1=" " ind2=" ">
    <subfield code="a">י. פקסימיל; דף [150] חסר; הושלם ע.</subfield>
  </datafield>
  <datafield tag="562" ind1=" " ind2=" ">
    <subfield code="b">עותק 2 במיקרופילם, פוזיטיב, מעותק הסמינר בניו-יורק. דפוסי צלום של הוצאה זו ו A 81 S. </subfield>
    4917 A 76 S 3719 בסינגטורה
  </datafield>
  <datafield tag="650" ind1=" " ind2="4">
    <subfield code="a">Rabbinical literature</subfield>
    <subfield code="z">Spain.</subfield>
  </datafield>
  <datafield tag="650" ind1=" " ind2="4">
    <subfield code="a">Responsa</subfield>
    <subfield code="y">1040-1600.</subfield>
  </datafield>
  <datafield tag="730" ind1=" " ind2=" ">
    <subfield code="a">א. </subfield>
    <subfield code="f">ל; ר </subfield>
  </datafield>
  <datafield tag="856" ind1="4" ind2="8">
    <subfield
code="u">aleph.nli.org.il/nnl/dig/books/bk001124667.html</subfield>
    <subfield code="z">For online version click here</subfield>
  </datafield>
  <datafield tag="914" ind1=" " ind2=" ">
    <subfield code="a">djvu</subfield>
    <subfield code="d">The Dorot Foundation</subfield>
  </datafield>
  <datafield tag="924" ind1=" " ind2=" ">
    <subfield code="a">רומה</subfield>
  </datafield>

```

```

<datafield tag="935" ind1=" " ind2=" " >
  <subfield code="a">20061019</subfield>
</datafield>
<datafield tag="999" ind1=" " ind2=" " >
  <subfield code="a">EBOOK</subfield>
</datafield>
</record>

```

Sample record manuscripts

```

<record xmlns="http://www.loc.gov/MARC21/slim"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.loc.gov/MARC21/slim
http://www.loc.gov/standards/marcxml/schema/MARC21slim.xsd">
  <leader>00000ntm a22 a 4500</leader>
  <controlfield tag="001">000040169</controlfield>
  <controlfield tag="005">20110509090914.0</controlfield>
  <controlfield tag="008">060525m18101821||| r 000 0 heb
d</controlfield>
  <datafield tag="035" ind1=" " ind2=" " >
    <subfield code="a">(OCoLC)232713868</subfield>
  </datafield>
  <datafield tag="041" ind1=" " ind2=" " >
    <subfield code="a">heb</subfield>
  </datafield>
  <datafield tag="100" ind1="1" ind2=" " >
    <subfield code="a">דנציג, אברהם בן יהואל מיכל</subfield>
    <subfield code="d">1748-1820</subfield>
  </datafield>
  <datafield tag="245" ind1="1" ind2="0">
    <subfield code="a">: חיי אדם</subfield>
    <subfield code="b">ו' &apos; נשמת אדם; ו' &apos; חיי אדם; ספר &apos; לר
    &apos; ע, עם מחיקות, הוספות ותיקונים רבים בכ' &quot; ט-תק; &quot; אברהם דנציג, דפוס ווילנא תקס
    המחבר.</subfield>
  </datafield>
  <datafield tag="260" ind1=" " ind2=" " >
    <subfield code="c">א (1821-1810) &quot; ע-תקפ; &quot; תק</subfield>
  </datafield>
  <datafield tag="500" ind1=" " ind2=" " >
    <subfield code="a">ה נדפסו על-פי ; &quot; ט, תקפ; &quot; המהדורות הבאות (ווילנא תקע) </subfield>
    תיקונים אלו ונוספים
  </datafield>
  <datafield tag="535" ind1="1" ind2=" " >
    <subfield code="a">Jerusalem - The National Library of Israel
    Ms. Heb. 4o187</subfield>
  </datafield>
  <datafield tag="535" ind1="1" ind2=" " >
    <subfield code="a">187 Ms. Heb. &quot; ירושלים - הספרייה הלאומית 4</subfield>
  </datafield>
  <datafield tag="740" ind1=" " ind2=" " >
    <subfield code="a">נשמת אדם</subfield>
  </datafield>
  <datafield tag="922" ind1=" " ind2=" " >
    <subfield code="a">1821</subfield>
  </datafield>
  <datafield tag="953" ind1=" " ind2=" " >
    <subfield code="a">OMS-0096404</subfield>

```

```

</datafield>
<datafield tag="956" ind1=" " ind2=" ">
  <subfield code="a">JER NLI 187=4</subfield>
</datafield>
<datafield tag="958" ind1=" " ind2=" ">
  <subfield code="a">אשכנזית רהוטא</subfield>
</datafield>
<datafield tag="966" ind1=" " ind2=" ">
  <subfield code="a">הלכה דינים ומנהגים</subfield>
</datafield>
<datafield tag="967" ind1=" " ind2=" ">
  <subfield code="a">כתב-יד. סריקה</subfield>
</datafield>
<datafield tag="968" ind1=" " ind2=" ">
  <subfield code="a">וילנא WILNO</subfield>
</datafield>
<datafield tag="968" ind1=" " ind2=" ">
  <subfield code="a">רוסיה</subfield>
</datafield>
<datafield tag="975" ind1=" " ind2=" ">
  <subfield code="a">187=4</subfield>
</datafield>
<datafield tag="976" ind1=" " ind2=" ">
  <subfield code="a">B 358 (187=4)</subfield>
</datafield>
<datafield tag="976" ind1=" " ind2=" ">
  <subfield code="a">CD 224</subfield>
</datafield>
<datafield tag="999" ind1=" " ind2=" ">
  <subfield code="a">MANUSCRIPT</subfield>
</datafield>
<datafield tag="999" ind1=" " ind2=" ">
  <subfield code="a">MIC</subfield>
</datafield>
</record>

```

4.6.3 Preliminary mapping of mandatory EDM fields

Property	Property range	Property description	Source Element	Source Element Description
ore:Aggregation				
edm:provider	edm:Agent entity	Mandatory element: Organisation (edm:Agent, see below) that provided this aggregation	Default value "DM2E" should also include Judaica Europeana (JE)	
edm:dataProvider	edm:Agent entity	Mandatory element: Organisation (edm:Agent, see below) that provided the base data for this aggregation (could be the same as edm:provider)	Default value "NLI"	

Property	Property range	Property description	Source Element	Source Element Description
edm:rights	remote resource	Mandatory element: URL of a resource describing licensing rights from http://pro.europeana.eu/documents/900548/1037382/Europeana_rights_201202.pdf	Will be default value	
edm:isShownBy	WebResource entity	Either edm:isShownBy or edm:isShownAt is mandatory. Web resource (see below) describing this specific ProvidedCHO aggregated here	856 or field expanded from ADAM object	
edm:isShownAt	WebResource entity	Either edm:isShownBy or edm:isShownAt is mandatory. Web resource (see below) containing the ProvidedCHO aggregated here	To be verified	
edm:ProvidedCHO				
edm:type	Literal	Mandatory element: Can be TEXT, VIDEO, SOUND, IMAGE, 3D	Default value: TEXT for books and manuscripts IMAGE for portraits	
dc:type	Specialization of ore:Proxy	Either dc:type or dc:subject is mandatory. The most specific type that applies to the CHO, repetition of rdf:type of Proxy	655	
dc:title	Literal with language tag	Either dc:title or dc:description is mandatory. Title of the object.	245 some records have alternative titles in field 246	
dc:description	Literal with language tag	Either dc:title or dc:description is mandatory. Description of the CHO		
dc:language	xsd:lang literal	Mandatory for textual objects: Most prominent language of the CHO	008 pos. 35-37 a separate virtual field can be created during export if necessary	
dc:subject	Resource	Either dc:type or dc:subject is mandatory. Resource in SKOS-Concept-Tree, e.g. Dewey Decimal Classification	650, 651, 600	

Table 12: EDM mapping NLI

4.7 CJH via EAJC – Center for Jewish History

4.7.1 Digital Collections

The Center of Jewish History (CJH) is one of the foremost Jewish research and cultural institutions in the world, having served over one million people in more than 100 countries. It is home to five partner organizations — American Jewish Historical Society, American Sephardi Federation, Leo Baeck Institute, Yeshiva University Museum and YIVO

Institute for Jewish Research. The Leo Baeck Institute (LBI) and YIVO both hold diverse collections on European Judaica and CJH will contribute metadata for digitized objects of these collections to DM2E via EJAC.

CJH's digital asset management system DigiTool currently manages gives public access to close to 45.000 digital objects (6.1 TB) from the collection of all partner organizations. Additional files have been digitized (2.5 TB) and are in process of being loaded to DigiTool. CJH Digital Collections (digital.cjh.org) is visited by 12.000 users a month.

CJH has digital objects that represent visual materials, sound recordings (mostly oral histories), movies, including 8.000 finding aids: the majority of the digital objects represent visual materials, such as books, posters, photographs, maps, museum objects, scrapbooks, albums, and drawings.

File formats are mostly JPEG2000 images TIFF images and PDF files for the visual materials and mp3 audio files for oral histories and other sound recordings. Other formats are jp2 tif, flv, pdf, txt, html, mp3, xml, wav, rtf, mpeg-4, mpeg-2, and mov.

The metadata is stored both in DigiTool and in the Aleph catalogue. When creating digital images and metadata, standards are followed, such as NARA's Guidelines for Digitizing Archival Materials for Electronic Access and DACS, AACR2, LCSH and thesauri (AAT and TGM). Each digital object also has metadata (in XML) attached. The formats used are MARCXML, DC, METS, and EAD (and occasionally MODS).

Example

```
000164102 FMT    L BK
000164102 LDR    L ^^^^nam^a22^^^^^a^4500
000164102 003    L NyNyCJH
000164102 005    L 20120509174223.0
000164102 008    L 060205s1928^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^ger^d
000164102 050    L $$aNc 1509 T7 A7
000164102 094    L $$aNc 1509 T7 A7
000164102 1001   L $$aColman, Fred A$$5LBI
000164102 24510 L $$aArtisten :$$bmit 116 farbigen Tafeln und
Abbildungen /$$cvon Fred A. Colman und Walter Trier.
000164102 260    L $$aDresden :$$bP. Aretz,$$c1928.
000164102 300    L $$a154p. illus., 4 col. plates. 25cm.
000164102 506    L $$aBook is digitized, please use link to digital
form for access
000164102 533    L $$aDigital image$$bNew York, NY$$cGruss Lipper
Digital Laboratory, Center for Jewish History$$d2010
000164102 500    L $$aOn illustrated cover: Artisten; ernstes und
heiteres Variete
000164102 655 4 L $$aArtists' illustrated books$$5CJH
000164102 650 0 L $$aGerman wit and humor, Pictorial$$5LBI
000164102 7001   L $$aTrier, Walter,$$d1890-1951$$5LBI
000164102 8564   L $$zDigital
form:$$uhttp://digital.cjh.org/R/?func=dbin-jump-
full&object_id=1182226
```

4.7.2 Preliminary mapping of mandatory EDM fields

The Center of Jewish History is associate partner of DM2E and will receive assistance from the National Library of Israel (NLI) which has taken the responsibility for technical

assistance to all Judaica Europeana members for the ingestion and mapping of their data. Both institutions will provide their metadata as MARCXML exported from Aleph due to restrictions of publishing options in DigiTool.

The mandatory mapping of MARCXML to EDM corresponds to the table of the NLI mapping (see Table 12 page 52). Only content provider specific fields have to be adapted.

4.8 BBAW – Berlin-Brandenburg Academy of Sciences (Berlin-Brandenburgische Akademie der Wissenschaften)

The Deutsches Textarchiv (DTA) of the Berlin-Brandenburgische Akademie der Wissenschaften (BBAW) provides digitised and transcribed 1st edition publications from the 17th to the end of the 19th century online. In close collaboration with SBB, BBAW will bring in these data into the interoperability infrastructure of WP2. When contextualised and transformed into EDM, SBB will make these data available for ingest by Europeana.

4.8.1 Legacy Data format – Text Encoding Initiative (TEI)

TEI is subdivided into two main sections: Metadata about the transcribed manuscript and the transcription/full-text with annotations. BBAW provides XML-data based on TEI5 XML schema. A crosswalk of the metadata section was worked out by UBER in collaboration with MPIWG who also provide TEI-based data within the project. A challenge for the time being seems to be the second section. While UBER TEI-full-text data are subdivided in paragraphs, full-texts of BBAW are subdivided on sentence level. An appropriate translation into EDM has to be agreed upon within the project.

Furthermore, SBB will aggregate all TEI-data from BBAW for the interoperability infrastructure later this year.

4.8.2 Requirements

Based on the crosswalk TEI to EDM, only few EDM elements are missing:

1. Specification for various types of titles for one document (subtitle, volume title)
2. Authority file number is missing in EDM (name key, see requirements of SBB)
3. Information concerning authors/ encoders of the TEI-document (respStmt)

This will be solved out in close collaboration with MPIWG and UBER. In addition to that few obligatory information is currently not in the data provided by BBAW:

- edm:isShownBy or edm:isShownAt

The missing information, however, can be added during the transformation and contextualisation process.

4.8.3 Preliminary mapping of mandatory EDM fields

Property	Property range	Property description	Source Element	Source Element Description
ore:Aggregation				
edm:provider	edm:Agent entity	Mandatory element: Organisation (edm:Agent, see below) that provided this aggregation	Default value "DM2E"	
edm:dataProvider	edm:Agent entity	Mandatory element. Organisation (edm:Agent, see below) that provided the base data for this aggregation (could be the same as edm:provider)	Publisher	
edm:rights	remote resource	Mandatory element: URL of a resource describing licensing rights from http://pro.europeana.eu/documents/900548/1037382/Europeana_rights_201202.pdf	Availability status	
edm:isShownBy	WebResource entity	Either edm:isShownBy or edm:isShownAt is mandatory. Web resource (see below) describing this specific ProvidedCHO aggregated here	idno type	
edm:isShownAt	WebResource entity	Either edm:isShownBy or edm:isShownAt is mandatory. Web resource (see below) containing the ProvidedCHO aggregated here	na	
edm:ProvidedCHO				
edm:type	Literal	Mandatory element. Can be TEXT, VIDEO, SOUND, IMAGE, 3D	TEXT	
dc:type	Specialization of ore:Proxy	Either dc:type or dc:subject is mandatory. The most specific type that applies to the CHO, repetition of rdf:type of Proxy	author	
dc:title	Literal with language tag	Either dc:title or dc:description is mandatory. Title of the object.	na	
dc:description	Literal with language tag	Either dc:title or dc:description is mandatory. Description of the CHO	title type	
dc:language	xsd:lang literal	Mandatory for textual objects. Most prominent language of the CHO	sourceDesc	
dc:subject	Resource	Either dc:type or dc:subject is mandatory. Resource in SKOS-Concept-Tree, e.g. Dewey Decimal Classification	langUsage	

Table 13: EDM mapping BBAW

4.9 CRNS – National Center for Scientific Research (Centre de la Recherche National Scientifique)

4.9.1 Underlying content

The *Digitale Faksimile Gesamtausgabe (DFGA)*¹¹ edited by Paolo D'Iorio and published by Nietzsche Source aims to provide a digital reproduction of the complete Nietzsche Estate, including first editions of works, manuscripts, letters and biographical documents for the first time. These are the primary sources for the study of Nietzsche's works and life and for the interpretation of his philosophy. The facsimile edition will allow scholars to verify, for example, the genuineness of the different text editions, which are the basis of any subsequent philosophical interpretations. If Nietzsche's manuscripts had been freely available, it would not have been possible to compile "The Will to Power", a book that Nietzsche never wrote. The DFGA will also allow scholars to embark upon new research enterprises like, for example, reconstructing the genesis of Nietzsche works or analysing the development of his philosophical thinking through his notebooks (what is known as "genetic criticism").

The readers will be provided with high-resolution colour facsimiles and will be able to browse, enlarge, print or download them. The DFGA relies on a specific digital classification system providing each page with a unique and stable internet address. In realizing this classification, the DFGA wanted to be compatible as much as possible with the standard critical edition and existing scholarship. Therefore, instead of adopting the more recent classification realized by the Goethe- und Schiller-Archive, which houses the Nietzsche estate, or providing a different classification, we decided to adapt the most widely used classification realised in 1933 by Hans Joachim Mette. The DFGA completes the Mette classification by correcting errors and filling in numbering gaps. The sigla used in the URLs correspond to Mette's abbreviations followed by page numbers. For example, page 194 of the manuscript M II 1 can be found at the following address: <http://www.nietzchesource.org/DFGA/M-II-1,194>.

This makes the DFGA one of the first electronic editions that can be quoted and referred to easily in academic research. In the future, a concordance will ensure compatibility with all previous classifications of Nietzsche's manuscripts as well as with the standard critical edition, enabling exact retrieval of the manuscript page in which Nietzsche wrote what the Colli-Montinari edition classified as a "posthumous fragment".

At present, 30,000 pages of the Nietzsche Estate have been digitized and about 9,000 of them have been published already in Nietzsche Source.¹²

¹¹ <http://www.nietzchesource.org/facsimiles/DFGA> [Retrieved 03.07.2012].

¹² <http://www.nietzchesource.org/documentation/en/DFGA.html> [Retrieved 03.07.2012].

Example

```
<?xml version="1.0" encoding="UTF-8"?><OAI-PMH
xmlns="http://www.openarchives.org/OAI/2.0/"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/
http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">

<responseDate>2009-11-30T10:59:10Z</responseDate>
<request>http://www.nietzschesource.org/oai</request>

<GetRecord>
<record>
<header>
<identifier>oai:talia/102732</identifier>
<datestamp>2009-11-30T10:21:20Z</datestamp>
</header>
<metadata xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/"
xmlns:ese="http://www.europeana.eu/schemas/ese/"
xmlns:dc="http://purl.org/dc/elements/1.1/"
xmlns:dct="http://purl.org/dc/elements/1.1/"
xmlns:europeana="http://www.europeana.eu/schemas/ese/"
xmlns:local="http://metadata.cerl.org/oai/namespace/local/"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/
http://www.openarchives.org/OAI/2.0/oai_dc.xsd
http://www.europeana.eu/schemas/ese/
http://www.europeana.eu/schemas/ese/ESE-V3.2.xsd">
<ese:record>
<dc:title>N II 6</dc:title>
<dc:creator>Nietzsche, Friedrich Wilhelm, 1844-1900, D'Iorio,
Paolo</dc:creator>
<dc:subject>Philosophy; 19th Century</dc:subject>
<dc:description>
In gelben Holzdeckel gebundenes Duodezheft (6,5 x 9). 64 Seiten.
'Memorabilia', autobiographische Aufzeichnungen aus Frühjahr-Sommer
1878.
</dc:description>
<dc:description>
Quaderno in dodicesimo (6,5 x 9) con copertina di legno giallo. 64
pagine. "Memorabilia", annotazioni di carattere autobiografico della
primavera-estate 1878.
</dc:description>
<dc:description>
Carnet en douzième (6,5 x 9) à reliure de bois jaune. 64 pages.
"Memorabilia", notes autobiographiques du printemps-été 1878.
</dc:description>
<dc:description>
Notebook (6,5 x 9) bound in yellow wood. 64 pages. 'Memorabilia',
autobiographical notes from Spring-Summer 1878.
</dc:description>
<dc:publisher>Association HyperNietzsche, Paris, France</dc:publisher>
<dc:type>IMAGE</dc:type>
<dc:identifier>http://www.nietzschesource.org/DFGA/N-II-6/</dc:identifier>
<dc:source>Goethe- und Schiller-Archiv, Weimar</dc:source>
```

```

<dc:language>de</dc:language>
<dc:rights>Creative Commons General Public License "Attribution, Non-
Commercial", version 3.0.</dc:rights>
<ese:provider>Association HyperNietzsche, Paris, France</ese:provider>
<ese:type>IMAGE</ese:type>
<ese:isShownBy>http://www.nietzschesource.org/DFGA/N-II-6,d1.jpeg?size=thumb</ese:isShownBy>
<ese:isShownAt>http://www.nietzschesource.org/DFGA/N-II-6</ese:isShownAt>
</ese:record>
</metadata>
</record>
</GetRecord>
</OAI-PMH>

```

A mapping of the DFGA metadata to EDM is straightforward as the source format is already ESE / DC and there are already fields that correspond to the EDM model.

4.9.2 Preliminary mapping of mandatory EDM fields

Property	Property range	Property description	Source Element	Source Element Description
ore:Aggregation				
edm:provider	edm:Agent entity	Mandatory element: Organisation (edm:Agent, see below) that provided this aggregation	Default value: "DM2E"	
edm:dataProvider	edm:Agent entity	Mandatory element: Organisation (edm:Agent, see below) that provided the base data for this aggregation (could be the same as edm:provider)	ese:provider	
edm:rights	remote resource	Mandatory element: URL of a resource describing licensing rights from http://pro.europeana.eu/documents/900548/1037382/Europeana_rights_201202.pdf	dc:rights	
edm:isShownBy	WebResource entity	Either edm:isShownBy or edm:isShownAt is mandatory. Web resource (see below) describing this specific ProvidedCHO aggregated here	ese:isShownAt	
edm:isShownAt	WebResource entity	Either edm:isShownBy or edm:isShownAt is mandatory. Web resource (see below) containing the ProvidedCHO aggregated here	ese:isShownBy	
edm:ProvidedCHO				
edm:type	Literal	Mandatory element: Can be TEXT, VIDEO, SOUND, IMAGE, 3D	ese:type	
dc:type	Specialization of ore:Proxy	Either dc:type or dc:subject is mandatory. The most specific type that applies to the CHO, repetition of rdf:type of Proxy	dc:type	
dc:title	Literal with language tag	Either dc:title or dc:description is mandatory. Title of the object.	dc:title	

Property	Property range	Property description	Source Element	Source Element Description
dc:description	Literal with language tag	Either dc:title or dc:description is mandatory. Description of the CHO	dc:description	
dc:language	xsd:lang literal	Mandatory for textual objects. Most prominent language of the CHO	dc:language	
dc:subject	Resource	Either dc:type or dc:subject is mandatory. Resource in SKOS-Concept-Tree, e.g. Dewey Decimal Classification	dc:subject	

Table 14: EDM mapping CRNS

4.10 JDC via EAJC – Joint Distribution Committee

4.10.1 Underlying content

JDC will share the earliest text collection in their archives, the Records of the American Jewish Joint Distribution Committee of the years 1914 – 1918. The material includes chronologies of events, correspondence, cables, reports, minutes and transcripts of meetings and conferences that document the origins, beginning in 1914, of systematic financial aid from American Jewish organizations to imperilled Jews abroad.

In August of that year, Henry Morgenthau Sr., then U.S. Ambassador to Turkey, cabled New York philanthropist Jacob Schiff seeking emergency aid for the Jews of Palestine. Subsequently, three relief committees – the American Jewish Relief Committee, the Central Relief Committee, and the People’s Relief Committee – formed the American Jewish Joint Distribution Committee, popularly known as the "Joint," to distribute the funds they raised for needy Jews in Eastern Europe and Palestine.

The records in this collection depict the accelerated expansion of JDC activities: fund-raising, annual campaign drives, and ongoing efforts to overcome political obstacles and military restrictions that restricted the dispatch of field workers abroad. Also documented is the developing cooperation with governments, foreign Jewish organizations and non-sectarian voluntary agencies so essential for JDC's work; it enabled JDC to transmit emergency relief funds and individual remittances, and to provide food, clothing, and shelter to endangered Jews in a nonpartisan, non-political manner.

Of note in this collection are records of prominent American Jewish leaders and philanthropists, such as Jacob Schiff, Felix Warburg, Louis Marshall, Cyrus Adler, Julius Rosenwald, and Judah Magnes, Genealogists and family researchers may find relevant correspondence, lists, and receipts for transmission of remittances to individuals.

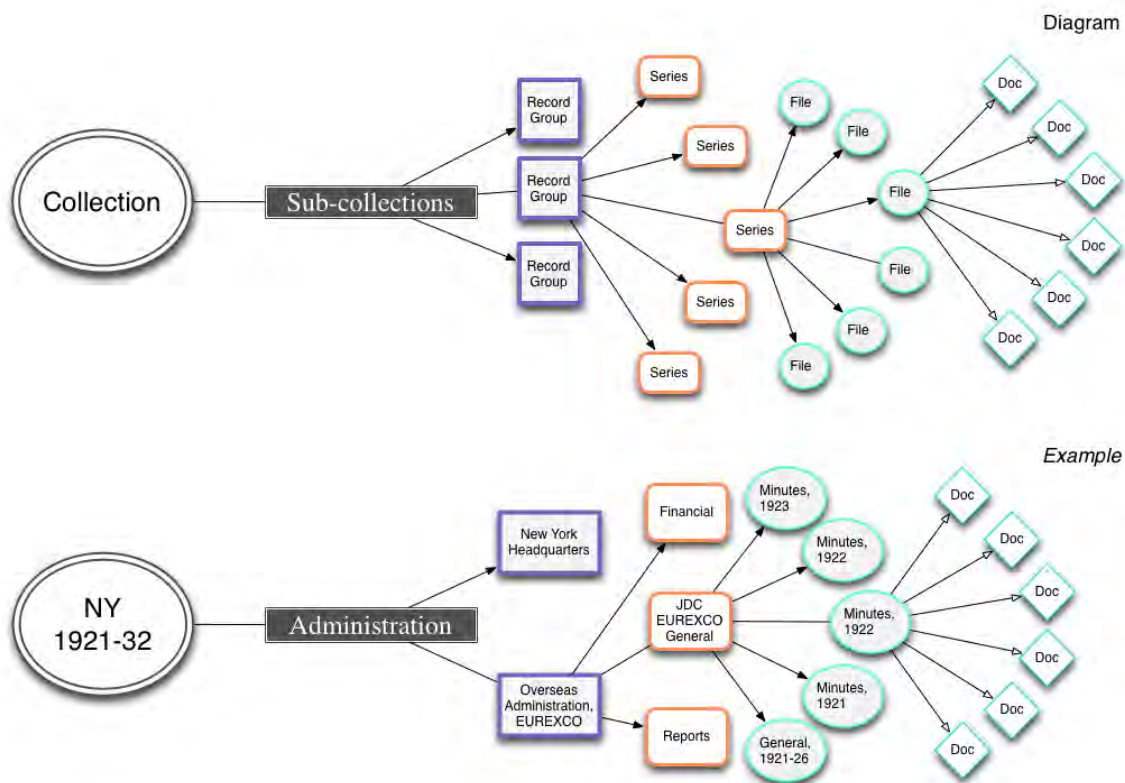


Figure 3: JDC hierarchy

JDC's collections are arranged in a hierarchy, with all digital assets located at the bottom (document) level. Each collection is divided into the same set of record types:

- Collection
 - Sub-Collection
 - Record Group
 - Series (if any exist)
 - File/Folder
 - Document

The XML files all have records at each level of the hierarchy. The corresponding JDC database record also has a record created (with no digital asset) for each level of the hierarchy. Without this association, documents would lose all contextual information. The majority of JDC metadata about collections are in the File level, which contains all subject terms/personal/corporate/geographical names.

JDC will need to work with JE to incorporate this hierarchical relationship into EDM, allowing the hierarchical nature of JDC's collections to be reflected in the flat environment of Europeana.

Example

```
<?xml version="1.0" encoding="UTF-8"?>
<ead>
  <archdesc level="Collection">
    <did>
      <unitid>NY AR191418</unitid>
      <origination label="Creator:">American Jewish Joint
```

```

Distribution Committee</origination>
  <unittitle label="Title:">Records of the American Jewish
Joint Distribution Committee of the years 1914 - 1918</unittitle>
  <unitdate from="01/01/1914" to="12/31/1918">1914 -
1918</unitdate>
  <physdesc label="Quantity:">155 folders</physdesc>
  <physloc label="Location:">American Jewish Joint
Distribution Committee, New York, NY</physloc>

```

```
</did>
```

```
<scopecontent>
```

The 1914 - 1918 Collection of the American Jewish Joint Distribution Committee is comprised of the records of the New York headquarters for the period from the Joint's origins providing emergency relief through World War I. The material consists primarily of correspondence, cables, reports and minutes of meetings and conferences among the most important American Jewish philanthropists and scholars of the early 20th century (Adler, Magnes, Marshall, Morgenthau, Rosenwald, Schiff and Warburg). The materials document JDC's evolving mission and structure as the distribution agency on behalf of the American Jewish Relief Committee, the Central Committee for the Relief of Jews and the People's Relief Committee. The materials document JDC's activities to aid distressed Jews in war-torn regions under varying occupation governments, as well as the social, economic and political conditions affecting Jewish communities in Palestine, Poland, Russia and Europe. Some of these files are incomplete. Researchers may also wish to consult the minutes of JDC meetings, which are indexed. Correspondence, lists and receipts for transmission of remittances to individuals may be useful to genealogists.

```
</scopecontent>
```

```
<addinfo>
```

There are record groups for countries and geographic regions in which JDC had interests or from which it received appeals for aid. These record groups are further arranged, as needed, into file series that cover, at a country level, topics of administration, organizations, subject matter or localities. The geographical classification of the materials presented numerous challenges. Between 1914 and 1918, national borders shifted because of the collapse of empires and military operations and occupations. Some communities or areas changed regimes repeatedly during the war era. Vilna, for example, was under Russian rule, was seized by Germany and included in a new unit Ober Ost, and was later claimed by Poland and Lithuania. Many locations were known by multiple names. Every effort has been made to cross-reference communities and to provide traditional and popular geographic names; these may not correspond to the names and national borders in the 21st century.

```
</addinfo>
```

```
<dsc type="combined">
```

```
<c01 level="Subcollection" id="1">
```

```
<did>
```

```
<unitid>NY AR191418/ 1</unitid>
```

```
<unittitle>Administration</unittitle>
```

```
</did>
```

```
<c02 level="Record Group" id="1">
```

```
<did>
```

```
<unitid>NY AR191418/ 1/ 1</unitid>
```

```
<unittitle>New York Administration</unittitle>
```

```
</did>
```

```

<c03 level="Series" id="1">
  <did>
    <unitid>NY AR191418/ 1/ 1/ 1</unitid>
    <unittitle>General Files</unittitle>
  </did>
  <c04 level="File" id="1.1">
    <did>
      <unitid>NY AR191418/ 1/ 1/ 1 -
1.1</unitid>
      <unittitle>Administration, General,
1914-1917</unittitle>
      <unitdate from="01/01/1914"
to="12/31/1917">1914-1917</unitdate>
      <physdesc>
        <extent>1 inch</extent>
      </physdesc>
    </did>
    <scopecontent audience="external">Reports,
correspondence, extracts of meeting minutes on origin of efforts to
raise relief funds among American Jews to alleviate Jewish suffering
in Europe and Palestine. Henry Morgenthau telegram, correspondence,
materials on needs of Jewish community in Palestine, including cable
(August 27, 1914) from William J. Bryan, U.S. secretary of state,
regarding U.S. plans to send relief ship. Miscellaneous on
allocations.</scopecontent>
    <controlaccess>
      <subject>Administration</subject>
      <geogname>Europe</geogname>
      <geogname>Eastern Europe</geogname>
      <subject>Financial</subject>
      <geogname>Holland</geogname>
      <geogname>Baltics</geogname>
      <geogname>Lithuania</geogname>
      <geogname>Romania</geogname>
      <geogname>Poland</geogname>
      <geogname>Germany</geogname>
      <geogname>Russia</geogname>
      <geogname>Palestine</geogname>
      <geogname>Syria</geogname>
      <geogname>Alexandria</geogname>
      <corpname>Jewish Charities and
Colonies in Palestine</corpname>
      <persname>Schiff, Jacob</persname>
      <persname>Marshall, Louis</persname>
      <persname>Morgenthau, Henry</persname>
      <persname>Bryan, William J.</persname>
      <corpname>U.S. State
Department</corpname>
      <subject>USS Vulcan</subject>
      <subject>USS Collier Caesar</subject>
      <subject>USS Sterling</subject>
      <subject>Fund-Raising</subject>
      <persname>Warburg, Felix</persname>
      <persname>Magnes, Judah</persname>
      <persname>Adler, Cyrus</persname>
      <persname>Lehman, Herbert</persname>
      <persname>Lehman, Arthur</persname>

```

```

        <subject>Albert Lucas</subject>
        <persname>Engelman, Morris</persname>
        <persname>Rosenwald, Julius</persname>
        <persname>Straus, Nathan</persname>
        <persname>Bogen, Boris</persname>
        <persname>Senior, Max</persname>
        <persname>Lowenstein,

Harriet</persname>

        <persname>Melchior, Carl</persname>
        <persname>Nierop, Frederick Solomon

Van</persname>

        <persname>Hoofien, Siegfried

(Eliezer)</persname>

        <subject>Remittances</subject>
        <corpname>American Jewish Relief
Committee for Jewish War Sufferers (Reform)</corpname>
        <corpname>Central Committee for the
Relief of Jewish Suffering Through the War (Orthodox)</corpname>
        <corpname>Jewish People's Relief
Committee (Labor)</corpname>
        <corpname>Jewish Colonization
Association (Petrograd)</corpname>
        <subject>Refugees</subject>
        <subject>Logistics</subject>
        <subject>Relief</subject>
        <subject>Reconstruction</subject>
        <subject>Rehabilitation</subject>
        <corpname>Jewish Bulletin</corpname>
        <subject>Personnel</subject>
        <subject>Banking</subject>
        <persname>Wilson, President

Woodrow</persname>

        <subject>Relief Supplies</subject>
        <persname>Tuchman, Barbara</persname>
        <subject>List</subject>
</controlaccess>
<c05 level="Item">
  <did>
    <unittitle>Form</unittitle>
    <unitdate from="">
  </unitdate>
    <origination>
  </origination>
    <physdesc>
      <genreform>Textual

Material</genreform>

    <bibref>
      <imprint />
    </bibref>
  </physdesc>

<langmaterial>English</langmaterial>
    <daogrp linktype="extended">
      <daoloc linktype="locator"
href="Documents\NY_AR1418\00001\NY_AR1418_01422.pdf" />
    </daogrp>
  </did>

```

```

        </c05>
        <c05 level="Item">
            <did>
                <unittitle>Numerical
Index</unittitle>
                <unitdate from="">
                </unitdate>
                <origination>
                </origination>
                <physdesc>
                <genreform>Textual
Material</genreform>
                <bibref>
                    <imprint />
                </bibref>
                </physdesc>
            </did>
            <daogrp linktype="extended">
                <daoloc linktype="locator"
href="Documents\NY_AR1418\00001\NY_AR1418_01423.pdf" />
            </daogrp>
        </c05>
    </c04>
</c03>
</c02>
</c01>
</dsc>
</archdesc>
</ead>

```

The main requirement of JDC towards a mapping to EDM is to be able to preserve the hierarchy structure of the EAD data format. If this is not possible it should be able to duplicate inherited elements from parent to child records.

4.10.2 Preliminary mapping of mandatory EDM fields

Property	Property range	Property description	Source Element	Source Element Description
ore:Aggregation				
edm:provider	edm:Agent entity	Mandatory element: Organisation (edm:Agent, see below) that provided this aggregation	Default value "DM2E" should also include Judaica Europeana (JE)	
edm:dataProvider	edm:Agent entity	Mandatory element: Organisation (edm:Agent, see below) that provided the base data for this aggregation (could be the same as edm:provider)	Default value "JDC"	

Property	Property range	Property description	Source Element	Source Element Description
edm:rights	remote resource	Mandatory element: URL of a resource describing licensing rights from http://pro.europeana.eu/documents/900548/1037382/Europeana_rights_201202.pdf	http://archive.s.jdc.org/researchers/terms-and-conditions.html	
edm:isShownBy	WebResource entity	Either edm:isShownBy or edm:isShownAt is mandatory. Web resource (see below) describing this specific ProvidedCHO aggregated here	<daoloc linktype="locator" href="X " />	
edm:isShownAt	WebResource entity	Either edm:isShownBy or edm:isShownAt is mandatory. Web resource (see below) containing the ProvidedCHO aggregated here		
edm:ProvidedCHO				
edm:type	Literal	Mandatory element: Can be TEXT, VIDEO, SOUND, IMAGE, 3D	Default value: "TEXT"	
dc:type	Specialization of ore:Proxy	Either dc:type or dc:subject is mandatory. The most specific type that applies to the CHO, repetition of rdf:type of Proxy		
dc:title	Literal with language tag	Either dc:title or dc:description is mandatory. Title of the object.	<unittitle label="Title:">x</unittitle>	
dc:description	Literal with language tag	Either dc:title or dc:description is mandatory. Description of the CHO	<physdesc label="Quantity:">x</physdesc>	
dc:language	xsd:lang literal	Mandatory for textual objects: Most prominent language of the CHO	<langmaterial >	
dc:subject	Resource	Either dc:type or dc:subject is mandatory. Resource in SKOS-Concept-Tree, e.g. Dewey Decimal Classification	<archdesc level="Collection">;<subject>	

Table 15: EDM mapping JDC

4.11 UBFFM – University Library JCS Frankfurt am Main (Universitätsbibliothek Johann Christian Senckenberg Frankfurt am Main)

4.11.1 Underlying content

The University Library in Frankfurt am Main holds a collection of approx. 2000 manuscripts which are scheduled to be digitised in the next years.

The digitisation of the 350 Hebrew Manuscripts has been finished and the metadata have already been exported in ESE format to Europeana as part of Judaica Europeana.

The following batch includes the 600 Mediaeval Manuscripts dating back to the 9th century with a total of 230.000 pages. It is intended to contribute these manuscripts to the DM2E project. Currently there are 450 Mediaeval Manuscripts already online.

The metadata of the digitised manuscripts should be harvested via the OAI interface using the prefix "mets". The data provide structure information using METS and title information using MODS metadata. The METS metadata always contain information about the pages, their order and the labels. In some cases a logical structure of the manuscript is also provided.

The MODS metadata generally have the following fields:

- Title containing also the shelf mark up to the substring " – "
- Authors and other persons using references to the GND where a GND record exists
- Place, date and physical description of the manuscript
- Place and date of the online edition
- URN of the online edition
- URL using a page URN pointing to the page where an extended description of the manuscript starts. (The manuscript catalogues are also digitalized, metadata can also be harvested.)
- Internal identifier of the catalogue system

There are some records of multivolume titles that are linked to each volume record.

All information on MODS level should be represented in EDM in order to have a significant step toward linked data.

Example

```
<?xml version="1.0" encoding="UTF-8"?>
<OAI-PMH xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/"
  xmlns:mods="http://www.loc.gov/mods/v3" xmlns:mets="http://www.loc.gov/METS/"
  xmlns:epicur="urn:nbn:de:1111-2004033116"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:marcxml="http://www.loc.gov/MARC21/slim"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xmlns="http://www.openarchives.org/OAI/2.0/"
  xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/
http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2012-06-19T07:04:28Z</responseDate>
  <request verb="GetRecord" metadataPrefix="mets"
  identifier="oai:sammlungen.ub.uni-frankfurt.de/mshebr:1759749">
    http://sammlungen.ub.uni-frankfurt.de/mshebr/oai/
  </request>
  <GetRecord>
    <record>
      <header>
        <identifier>oai:sammlungen.ub.uni-frankfurt.de/mshebr:1759749</identifier>
        <datestamp>2011-07-12T14:08:28Z</datestamp>
        <setSpec>ubffmjms</setSpec>
      </header>
      <metadata>
        <mets:mets xmlns:xlink="http://www.w3.org/1999/xlink"
  xmlns:vls="http://semantics.de/vls"
  xsi:schemaLocation="http://www.loc.gov/METS/
http://www.loc.gov/standards/mets/version18/mets.xsd"
  OBJID="7" LABEL="Hebräische Handschriften">
          <mets:metsHdr CREATEDATE="2012-06-19T09:04:28">
            <mets:agent TYPE="OTHER" ROLE="OTHER" OTHERTYPE="SOFTWARE">
              <mets:name>vls/2.12dev</mets:name>
            </mets:agent>
            <mets:agent TYPE="OTHER" ROLE="OTHER" OTHERTYPE="INSTANCE">
              <mets:name>ubffm-server</mets:name>
            </mets:agent>
            <mets:agent TYPE="OTHER" ROLE="OTHER" OTHERTYPE="REPOSITORY">
```

```

    < mets: name > sammlungen.ub.uni-frankfurt.de/mshebr </ mets: name >
  </ mets: agent >
  < mets: agent TYPE="OTHER" ROLE="OTHER" OTHERTYPE="BUILDER" >
    < mets: name > vd </ mets: name >
  </ mets: agent >
</ mets: metsHdr >
< mets: dmdSec ID="md1759749" >
  < mets: mdWrap MIMETYPE="text/xml" MDTYPE="MODS" >
    < mets: xmlData >
      < mods: mods version="3.4"
        xsi: schemaLocation="http://www.loc.gov/mods/v3
http://www.loc.gov/standards/mods/v3/mods-3-4.xsd" >
        < mods: titleInfo >
          < mods: title > Ms. hebr. oct. 6 - Nof 'ets ḥayim </ mods: title >
        </ mods: titleInfo >
        < mods: titleInfo type="alternative" >
          < mods: title > Nōf `eṣ ḥajjim </ mods: title >
        </ mods: titleInfo >
        < mods: name type="personal" authority="gnd"
authorityURI="http://d-nb.info/gnd/"
valueURI="http://d-nb.info/gnd/119264528" >
          < mods: namePart > Luria, Yitsḥak ben Shelomoh </ mods: namePart >
          < mods: namePart type="given" > Yitsḥak ben
Shelomoh </ mods: namePart >
          < mods: namePart type="family" > Luria </ mods: namePart >
          < mods: role >
            < mods: roleTerm type="code"
authority="marcrelator" > aut </ mods: roleTerm >
            < mods: roleTerm type="text" > Verf. </ mods: roleTerm >
          </ mods: role >
        </ mods: name >
        < mods: name type="personal" authority="gnd"
authorityURI="http://d-nb.info/gnd/"
valueURI="http://d-nb.info/gnd/119264528" >
          < mods: namePart > אִירוּל, קחצי ןב החלש </ mods: namePart >
          < mods: namePart type="given" > קחצי ןב החלש </ mods: namePart >
          < mods: namePart type="family" > אִירוּל </ mods: namePart >
          < mods: role >
            < mods: roleTerm type="code"
authority="marcrelator" > aut </ mods: roleTerm >
            < mods: roleTerm type="text" > Verf. </ mods: roleTerm >
          </ mods: role >
        </ mods: name >
        < mods: name type="personal" authority="gnd"
authorityURI="http://d-nb.info/gnd/"
valueURI="http://d-nb.info/gnd/124255957" >
          < mods: namePart > Viṭal, Ḥayim ben Yosef </ mods: namePart >
          < mods: namePart type="given" > Ḥayim ben Yosef </ mods: namePart >
          < mods: namePart type="family" > Viṭal </ mods: namePart >
          < mods: role >
            < mods: roleTerm type="code"
authority="marcrelator" > aut </ mods: roleTerm >
            < mods: roleTerm type="text" > Verf. </ mods: roleTerm >
          </ mods: role >
        </ mods: name >
        < mods: name type="personal" authority="gnd"
authorityURI="http://d-nb.info/gnd/"
valueURI="http://d-nb.info/gnd/124255957" >
          < mods: namePart > לט׳ו, ןב ס׳י ןב ס׳י </ mods: namePart >
          < mods: namePart type="given" > ןב ס׳י ןב ס׳י </ mods: namePart >
          < mods: namePart type="family" > לט׳ו </ mods: namePart >
          < mods: role >
            < mods: roleTerm type="code"
authority="marcrelator" > aut </ mods: roleTerm >
            < mods: roleTerm type="text" > Verf. </ mods: roleTerm >

```

```

        </mods:role>
    </mods:name>
    <mods:name type="personal" authority="gnd"
authorityURI="http://d-nb.info/gnd/"
valueURI="http://d-nb.info/gnd/13766365X">
    <mods:namePart>Popers, Me'ir ben Yehudah Leib</mods:namePart>
    <mods:namePart type="given">Me'ir ben Yehudah
Leib</mods:namePart>
    <mods:namePart type="family">Popers</mods:namePart>
    <mods:role>
    <mods:roleTerm type="code"
authority="marcrelator">asn</mods:roleTerm>
    <mods:roleTerm type="text">Beteiligte Person</mods:roleTerm>
    </mods:role>
    </mods:name>
    <mods:name type="personal" authority="gnd"
authorityURI="http://d-nb.info/gnd/"
valueURI="http://d-nb.info/gnd/13766365X">
    <mods:namePart>פופרס, מיר בן יהודה</mods:namePart>
    <mods:namePart type="given">מיר בן יהודה</mods:namePart>
    <mods:namePart type="family">פופרס</mods:namePart>
    <mods:role>
    <mods:roleTerm type="code"
authority="marcrelator">asn</mods:roleTerm>
    <mods:roleTerm type="text">Beteiligte Person</mods:roleTerm>
    </mods:role>
    </mods:name>
    <mods:typeOfResource>text</mods:typeOfResource>
    <mods:originInfo>
    <mods:place>
    <mods:placeTerm type="text">[S.l.]</mods:placeTerm>
    </mods:place>
    <mods:dateIssued keyDate="yes"
encoding="w3cdtf">1728</mods:dateIssued>
    <mods:issuance>monographic</mods:issuance>
    </mods:originInfo>
    <mods:originInfo>
    <mods:place>
    <mods:placeTerm type="text">Frankfurt am
Main</mods:placeTerm>
    </mods:place>
    <mods:publisher>Univ.-Bibliothek</mods:publisher>
    <mods:dateIssued keyDate="yes"
encoding="w3cdtf">2011</mods:dateIssued>
    <mods:edition>[Electronic ed.]</mods:edition>
    </mods:originInfo>
    <mods:language>
    <mods:languageTerm authority="iso639-2b"
type="code">heb</mods:languageTerm>
    </mods:language>
    <mods:physicalDescription>
    <mods:extent>68 Bl. ; Beschreibstoff: Papier ; 15x19,5
cm</mods:extent>
    </mods:physicalDescription>
    <mods:identifier type="urn">urn:nbn:de:hebis:30:2-
523</mods:identifier>
    <mods:location>
    <mods:url displayLabel="Katalognachweis"
>http://nbn-resolving.de/urn:nbn:de:hebis:30:2-17-p0031-
4</mods:url>
    </mods:location>
    <mods:recordInfo>
    <mods:recordCreationDate encoding="marc">06-04-
11</mods:recordCreationDate>
    <mods:recordChangeDate encoding="marc">07-12-
11</mods:recordChangeDate>

```

```

        <mods:recordIdentifier
source="ubffmjms">234517832</mods:recordIdentifier>
        </mods:recordInfo>
    </mods:mods>
</mets:xmlData>
</mets:mdWrap>
</mets:dmdSec>
<mets:amdSec ID="amd1759749">
    <mets:rightsMD ID="rights1759749">
        <mets:mdWrap MIMETYPE="text/xml" MDTYPE="OTHER"
OTHERMDTYPE="DVRIGHTS">
            <mets:xmlData>
                <dv:rights xmlns:dv="http://dfg-viewer.de/">
                    <dv:ownerLogo>http://sammlungen.ub.uni-
frankfurt.de/mshebr/domainresource/static/graphics/connectors/viewerLogo.gif</dv:ownerLogo>
                    <dv:ownerSiteURL>http://www.ub.uni-
frankfurt.de</dv:ownerSiteURL>
                </dv:rights>
            </mets:xmlData>
        </mets:mdWrap>
    </mets:rightsMD>
    <mets:digiprovMD ID="digiprov1759749">
        <mets:mdWrap MIMETYPE="text/xml" MDTYPE="OTHER"
OTHERMDTYPE="DVLINKS">
            <mets:xmlData>
                <dv:links xmlns:dv="http://dfg-viewer.de/">
                    <dv:reference>http://lbsopac.rz.uni-
frankfurt.de/CMD?ACT=SRCHA&IKT=6015&DB=30&TRM=ppn:234517832</dv:reference>
                    <dv:presentation>http://sammlungen.ub.uni-
frankfurt.de/mshebr/id/1759749</dv:presentation>
                </dv:links>
            </mets:xmlData>
        </mets:mdWrap>
    </mets:digiprovMD>
</mets:amdSec>
<mets:fileSec>
    <mets:fileGrp USE="DOWNLOAD">
        <mets:file MIMETYPE="application/pdf"
CHECKSUM="04043950dalff6989a7a948099500f4c71f2dff3"
SIZE="62807059"
CHECKSUMTYPE="SHA-1" CREATED="2012-06-02T19:55:10"
ID="PDF_1759749">
            <mets:FLocat
                xlink:href="http://sammlungen.ub.uni-
frankfurt.de/mshebr/download/pdf/1759749"
                LOCTYPE="URL"/>
        </mets:file>
    </mets:fileGrp>
    <mets:fileGrp USE="DEFAULT">
        <mets:file MIMETYPE="image/jpeg" CREATED="2011-04-04T19:43:16"
ID="IMG_DEFAULT_1558722">
            <mets:FLocat
                xlink:href="http://sammlungen.ub.uni-
frankfurt.de/mshebr/download/webcache/1000/1558722"
                LOCTYPE="URL"/>
        </mets:file>
        <mets:file MIMETYPE="image/jpeg" CREATED="2011-04-04T19:43:17"
ID="IMG_DEFAULT_1558723">
            <mets:FLocat
                xlink:href="http://sammlungen.ub.uni-
frankfurt.de/mshebr/download/webcache/1000/1558723"
                LOCTYPE="URL"/>
        </mets:file>
    <!-- [...] all pages in different sizes -->

```

```
<mets:file MIMETYPE="image/jpeg" CREATED="2011-04-04T19:49:00"
ID="IMG_MAX_1558975">
  <mets:FLocat
    xlink:href="http://sammlungen.ub.uni-
frankfurt.de/mshebr/download/webcache/1504/1558975"
    LOCTYPE="URL"/>
  </mets:file>
<mets:file MIMETYPE="image/jpeg" CREATED="2011-04-04T19:49:08"
ID="IMG_MAX_1558981">
  <mets:FLocat
    xlink:href="http://sammlungen.ub.uni-
frankfurt.de/mshebr/download/webcache/1504/1558981"
    LOCTYPE="URL"/>
  </mets:file>
</mets:fileGrp>
</mets:fileSec>
<mets:structMap TYPE="PHYSICAL">
  <mets:div ID="phys1759749" TYPE="physSequence">
    <mets:fptr FILEID="PDF_1759749"/>
    <mets:div ID="phys1558722" TYPE="page" LABEL="[Seite 4]" ORDER="1">
      <mets:fptr FILEID="IMG_DEFAULT_1558722"/>
      <mets:fptr FILEID="IMG_THUMBS_1558722"/>
      <mets:fptr FILEID="IMG_MIN_1558722"/>
      <mets:fptr FILEID="IMG_MAX_1558722"/>
    </mets:div>
    <mets:div ID="phys1558723" TYPE="page" LABEL="[Seite 5]" ORDER="2">
      <mets:fptr FILEID="IMG_DEFAULT_1558723"/>
      <mets:fptr FILEID="IMG_THUMBS_1558723"/>
      <mets:fptr FILEID="IMG_MIN_1558723"/>
      <mets:fptr FILEID="IMG_MAX_1558723"/>
    </mets:div>
    <mets:div ID="phys1558724" TYPE="page" LABEL="Blatt Ir" ORDER="3"
ORDERLABEL="Ir">
      <mets:fptr FILEID="IMG_DEFAULT_1558724"/>
      <mets:fptr FILEID="IMG_THUMBS_1558724"/>
      <mets:fptr FILEID="IMG_MIN_1558724"/>
      <mets:fptr FILEID="IMG_MAX_1558724"/>
    </mets:div>
    <mets:div ID="phys1558725" TYPE="page" LABEL="Blatt Iv" ORDER="4"
ORDERLABEL="Iv">
      <mets:fptr FILEID="IMG_DEFAULT_1558725"/>
      <mets:fptr FILEID="IMG_THUMBS_1558725"/>
      <mets:fptr FILEID="IMG_MIN_1558725"/>
      <mets:fptr FILEID="IMG_MAX_1558725"/>
    </mets:div>
    <mets:div ID="phys1558726" TYPE="page" LABEL="Blatt 1r" ORDER="5"
ORDERLABEL="1r">
      <mets:fptr FILEID="IMG_DEFAULT_1558726"/>
      <mets:fptr FILEID="IMG_THUMBS_1558726"/>
      <mets:fptr FILEID="IMG_MIN_1558726"/>
      <mets:fptr FILEID="IMG_MAX_1558726"/>
    </mets:div>
    <mets:div ID="phys1558727" TYPE="page" LABEL="Blatt 1v" ORDER="6"
ORDERLABEL="1v">
      <mets:fptr FILEID="IMG_DEFAULT_1558727"/>
      <mets:fptr FILEID="IMG_THUMBS_1558727"/>
      <mets:fptr FILEID="IMG_MIN_1558727"/>
      <mets:fptr FILEID="IMG_MAX_1558727"/>
    </mets:div>
    <mets:div ID="phys1558728" TYPE="page" LABEL="Blatt 2r" ORDER="7"
ORDERLABEL="2r">
      <mets:fptr FILEID="IMG_DEFAULT_1558728"/>
      <mets:fptr FILEID="IMG_THUMBS_1558728"/>
      <mets:fptr FILEID="IMG_MIN_1558728"/>
      <mets:fptr FILEID="IMG_MAX_1558728"/>
    </mets:div>
```

```

<!-- [...] all pages -->
<mets:div ID="phys1558981" TYPE="page" LABEL="[Seite 262]"
ORDER="254">
  <mets:fptr FILEID="IMG_DEFAULT_1558981"/>
  <mets:fptr FILEID="IMG_THUMBS_1558981"/>
  <mets:fptr FILEID="IMG_MIN_1558981"/>
  <mets:fptr FILEID="IMG_MAX_1558981"/>
</mets:div>
</mets:structMap>
<mets:structMap TYPE="LOGICAL">
  <mets:div ID="log1759749" DMDID="md1759749" ADMID="amd1759749"
TYPE="document"
  LABEL="Ms. hebr. oct. 6 - Nof 'ets ḥayim" ORDER="1"
  CONTENTIDS="urn:nbn:de:hebis:30:2-523">
    <mets:mptr LOCTYPE="URL"
      xlink:href="http://sammlungen.ub.uni-
frankfurt.de/mshebr/oai/?verb=GetRecord&metadataPrefix=mets&identifier=175
9749"/>
    <mets:fptr FILEID="PDF_1759749"/>
  </mets:div>
</mets:structMap>
<mets:structLink>
  <mets:smLink xlink:from="log1759749" xlink:to="phys1558722"/>
  <mets:smLink xlink:from="log1759749" xlink:to="phys1558723"/>
  <mets:smLink xlink:from="log1759749" xlink:to="phys1558724"/>
  <!-- [...] all pages -->
  <mets:smLink xlink:from="log1759749" xlink:to="phys1558981"/>
</mets:structLink>
</mets:mets>
</metadata>
</record>
</GetRecord>
</OAI-PMH>

```

4.11.2 Preliminary mapping of mandatory EDM fields

Property	Property range	Property description	Source Element	Source Element Description
ore:Aggregation				
edm:provider	edm:Agent entity	Mandatory element: Organisation (edm:Agent, see below) that provided this aggregation	Default value "DM2E"	
edm:dataProvider	edm:Agent entity	Mandatory element: Organisation (edm:Agent, see below) that provided the base data for this aggregation (could be the same as edm:provider)	Default value: "Universitätsbibliothek JCS Frankfurt am Main"	
edm:rights	remote resource	Mandatory element: URL of a resource describing licensing rights from http://pro.europeana.eu/documents/900548/1037382/Europeana_rights_201202.pdf	http://creativecommons.org/publicdomain/1.0/	Public Domain: CC0
edm:isShownBy	WebResource entity	Either edm:isShownBy or edm:isShownAt is mandatory. Web resource (see below) describing this specific ProvidedCHO aggregated here		

Property	Property range	Property description	Source Element	Source Element Description
edm:isShownAt	WebResource entity	Either edm:isShownBy or edm:isShownAt is mandatory. Web resource (see below) containing the ProvidedCHO aggregated here	"http://nbn-resolving.de/" + mods:identifier[@type="urn"]	
edm:object	WebResource entity	Preview picture for the ProvidedCHO	mets:fileSec/mets:fileGrp/mets:file[1]/mets:FLocat/@xlink:href	Via @ORDER=1
edm:ProvidedCHO				
edm:type	Literal	Mandatory element: Can be TEXT, VIDEO, SOUND, IMAGE, 3D	Default value "TEXT"	
dc:type	Specialization of Proxy	Either dc:type or dc:subject is mandatory. The most specific type that applies to the CHO, repetition of rdf:type of Proxy	mods:typeOfResource	
dc:title	Literal with language tag	Either dc:title or dc:description is mandatory. Title of the object.	mods:title	
dcterms:alternative	Literal with language tag	Subtitle; Any form of the title used as a substitute or alternative to the formal title of the resource	mods:titleInfo[@type='alternative']	
dm2e:titleTransliteration	Literal with language tag	Title transliteration	TBD	
dc:description	Literal with language tag	Either dc:title or dc:description is mandatory. Description of the CHO		
dc:language	xsd:lang literal	Mandatory for textual objects: Most prominent language of the CHO	mods:language	
dc:issued	xsd:dateTime literal	Date of publication, also possible: dct:created, or dc:date for generic cases	mods:originInfo[not(contains(mods:edition, 'Electronic'))]/mods:dateIssued[@keyDate='yes']	
dc:creator	edm:Agent entity	Creator of the CHO, possibly its author. Subproperties may be used to specialize relationship	mods:name[type="personal"]	
dm2e:publishedAt	edm:Place entity	Place of publication, subproperty of dct:spatial, other extensions possible	mods:originInfo[not(contains(mods:edition, 'Electronic'))]/mods:place/mods:placeTerm	
edm:rights	Remote resource	Possible values: http://pro.europeana.eu/documents/900548/0d423921-23e0-45fa-82a4-2ac72b3e6f38	http://creativecommons.org/publicdomain/mark/1.0/	
dc:subject	Resource	Either dc:type or dc:subject is mandatory. Resource in SKOS-Concept-Tree, e.g. Dewey Decimal Classification		

Table 16: EDM mapping UBFFM

5 Requirements gathering WP3

In the meetings of the Digital Humanities Advisory Board (DHAB) which took place at the Humboldt University among others a functional requirements gathering process took place for the prototyping platform of WP3.

The list represents a summary of the most important requirements regarding the annotations of the data which will be implemented in the project platform in WP3:

- **Granularity of Annotations:** The level of granularity of source data and the allowed annotations is important. Annotations should be doable on the transcription texts and on the images and on certain sections of a text and an image.
- **Existing annotations and markup** should also be importable to Korbo and Pundit: TEI or other entity markup already present in the original text should be supported by Pundit.
- **Access control and user groups:** The author of an annotation needs to be visible and recognisable. Generally, authorship is extremely important, as is an option to set up and specify user groups in order to enable collaboration.
- **Tracking versions of objects:** Regarding annotations it is extremely important to be able to track to which object an annotation has been attached to, independent of which specific version it originally had been actually attached to (stability of reference).
- **Discourse visualisation:** The possibility of visualising the structures of arguments or discourse (as they appear in the form of annotations) is an important feature.
- **Pattern discovery and visualisation on big data:** Exploit the machines capacity to draw inferences, to “remember things” (for the scholar who has to rely on paper sheets otherwise) and to discover and to visualise patterns (like co-occurrences) on annotated corpora and, generally, on big data. Those are things scholars can not do easily “offline” and certainly not in the traditional, analogue world and this might be the key value.
- **Enable collaboration:** In addition, working collaboratively must be possible as well.
- **Enable a “definable or customisable linking policy”.** Let the scholars choose where to look for concepts for annotations (e.g., first in VIAF, then in FOAF etc.). That is something they cannot do otherwise. Also, let scholars share vocabularies (of terms used to annotate and contextualise).
- **Grouping and filtering options:** It would be great if Korbo could automatically group similar baskets or show/adopt fitting annotations.
- **Annotations made on different representations of the same conceptual item** (e.g. different transcription, different languages, facsimile and transcription) should be possible. One crucial requirement is that such relations among resources (e.g. isDifferentVersionOf) are present in the EDM data and encoded with a well defined ontology.

A more accurate summary of the requirements can be found in the DM2E deliverable D3.1 – Initial Specification Report of WP3.

6 Summary and WP1 Outlook

This document describes the results of DM2E's WP1 Task 1.1 "Collect metadata formats and relational backend structures" and Task 1.2 "Collect requirements". As a result of two surveys based on the questionnaires, WP1 members established the basis for the work ahead, particularly of WP2 and WP3. While WP2 develops an interoperability infrastructure requiring in depths knowledge about the raw data, WP3 in turn requires precise information about quality, quantity, and accessibility of data.

As it turned out, all content providers already work with standardised metadata formats which serve each various purposes and communities:

- MAB2/Marc21 interlibrary data exchange formats
- EAD archival finding aid encoding and exchange formats
- TEI full text encoding format developed by the humanities
- METS encoding descriptive, administrative, and structural information of objects in digital libraries, often used in combination with MODS, an XML-schema derived from the MARC21 standard, for encoding bibliographic or descriptive information within a METS-File

Therefore, the surveys have been both a necessary and sustainable contribution to the development of the Europeana Data Model as well as the work of WP2 and WP3. Furthermore, the work ahead will be based on the results of survey, e.g. identification of units of information that ought to be not only transformed from one to another metadata model, but from literals to unambiguous resources. The latter significantly enhances user experiences in terms of machine supported retrieval functionalities and research done by scholars. It is in line with the main objectives of the interoperability infrastructure which aims at transforming legacy data into RDFbased information.

This deliverable, D1.1 Requirements Report, was designed to provide an overview of the material that will be contributed by WP1 partners in the course of the project. The Requirements report is also intended as a guideline for all project partners. Nevertheless, it is to be underscored that the requirements stated in this deliverable may be subject to changes with the projects' progress.

The next steps in WP1 are to provide EDM trainings for every content provider to enable a mapping of the content. This step is essential for Task 1.3 "Testing the User Interface for Creating Mapping, Interlinking Heuristics and for Configuring the Workflow". EDM trainings will be supported by UBER. It enables content providers to prepare for the final ingestion of metadata to EDM.

Furthermore, WP1 also proceeds with inviting new content providers. This includes the provision of necessary additional capacities for consulting new providers as well as analysing and transforming new metadata. Encouraging communication in this respect have been established with the recently finalised Europeana Regia project, the national Libraries of Czechia and Denmark and with the COST action IS1005 (Medieval Culture on the Web).

7 Appendix

7.1 Content questionnaire

DM2E – Content Questionnaire WP1

Distribution

Version	Date of sending	Name	Role in project
0.1	14.03.2012	Max Kaiser, ONB	
0.2	15.03.2012	Georg Petz, ONB	
0.3	21.03.2012	Stefan Gradmann, UBER	Coordinator
0.3	21.03.2012	Michele Barbera, Christian Morbidoni, Net7	WP3 lead
1.0	27.03.2012	WP1 content providers	WP1 content providers

Revisions

Version	Status	Author	Date	Changes
0.1	Draft	Ewelina Suchorzebska, ONB	14.03.2012	Initial version
0.2	Draft	Ewelina Suchorzebska, ONB	20.03.2012	Some additions, general changes
0.3	Draft	Ewelina Suchorzebska, ONB	21.03.2012	General changes
1.0	Final	Ewelina Suchorzebska, ONB	27.03.2012	Additions according to comments by Stefan Gradmann and Christian Morbidoni

Contact details

Last name	
First name	
Gender	
E-mail address	
Phone number	
Skype	
Area of expertise / Function	

Organisation

Name of the organisation	
Address	
Zip/City	
Country	

General

Please fill in one questionnaire for each collection you are contributing to DM2E.

Fill in the full name of the collection.	
Are there significant changes to the DoW regarding the objects you are making available for DM2E (e.g. adding collections, changes in the amount)?	
What license will be applied to the objects? Use one of the 12 rights statements available for edm:rights. ¹³	
Are the objects already available in digital form? If no, what is the expected date of availability?	
What languages are your full-texts in?	
Provide a link to the digital library(ies) or repositories that are currently hosting your digital objects.	
What is the number of objects (e.g. entire books, manuscripts etc.)?	
What is the number of digitised pages / single images?	
In which format will your library provide digital images (e.g. JPEG, JPEG 2000, TIFF)? In cases where images are provided in formats not widely supported by browser (such as JPEG 2000 or TIFF, refer to http://en.wikipedia.org/wiki/Comparison_of_web_browsers#Image_format_support), are you able to provide alternative formats (such as JPEG, GIF, PNG)?	
Do you expect to deliver full-text during the project?	
How was the full-text created (e.g. transcription, OCR)?	
What is the typical accuracy rate of your full-text (e.g. 70%, 99,5%)?	
In which format is the full-text available (e.g. plain text, METS-ALTO, HOOCR)?	
What character encoding is used in your full-text (e.g. UTF-8, UTF-16)?	
Do your textual objects contain TEI or other kinds of textual markup? Provide short details or links to documentation.	
To what level of granularity are your objects available (e.g. books, pages, sentences, etc.)?	
Do fragments have accessible URLs? Is the fragmentation marked up in the HTML/XML representation?	
Does your organisation already have data in Europeana as EDM (at data.europeana.eu)? Provide details or links.	
The WP3 platform will deal with distributed content recombination within and beyond the boundaries of DM2E. Do you have suggestions regarding content and data sources on the web that might	

¹³ See also: <http://pro.europeana.eu/documents/858566/7f14c82a-f76c-4f4f-b8a7-600d2168a73d>; p. 17-18 [Retrieved 30.07.2012].

complement your content or to which your content could provide context, or, in other words, that might be interesting to associate with your content?	
Can you provide a contact person for the technical aspects?	
Any additional comment	

API for accessing data

If you plan to implement one of the beneath stated measures, also indicate when you plan to do so.

Are your digital objects retrievable via URLs? Do they have unique, stable URLs? If so please provide examples, showing a representative for each media type.	
Are your objects available via persistent identifiers? If so, indicate the type of persistent identifier you are using (e.g. URN, DOI, etc.) and provide examples.	
Does your system expose web APIs to retrieve digital objects? If possible link to API documentation and/or provide example calls and information about supported data formats (e.g. JSON, XML), access control and usage limits.	
Does your system provide digital media relying on specific protocols (e.g. Internet Imaging Protocol (IIP) for efficiently delivering high resolution images or streaming servers for videos), other than plain HTTP?	
Does your system provide content search web APIs?	
Does your system provide different resolutions for the same image (or other media)? Are such resolutions addressable by a URL? Or does it address single spatial regions of large images? Give examples.	
Do the metadata in your digital library contain links to other data within your digital library (e. g. to authority data for persons and concepts) and if so, how are these links processed when creating metadata export formats?	
Is the database backend of your digital library directly accessible?	
Does your digital library support some form of user annotation of contents?	
Any additional comment	

Requirements

The following are some possible requirements we are evaluating for the content platforms (your digital libraries) to help WP3 software in accessing and meaningfully handle digital content.

What would be the effort for your organization to build support for:

(If you choose a), b), or c), also indicate until when you can implement these measures.)

A simple REST API for getting digital objects. Requirements: stable URLs for each digital object. Support for standard HTTP methods (as GET and HEAD) and headers (as last-modified, content-length, content-type, content-encoding).	a) easy b) feasible c) hard d) not possible
Implement a content embed API such as http://oembed.com/ , to deliver metadata about the digital file (e.g. format, encoding, size, editing date, etc.). This could enable WP3 to show better previews of your content.	a) easy b) feasible c) hard d) not possible
Implement a similar API for getting additional metadata about objects represented in the digital files (e.g. where there are not fully captured by EDM representations).	a) easy b) feasible c) hard d) not possible
Add metadata markup to your HTML pages to explicitly indicate atomic content (e.g. single paragraphs, pages, etc.) and provide stable URLs for them and basic metadata? A possible standard could be RDFa (http://www.w3.org/TR/xhtml-rdfa-primer/).	a) easy b) feasible c) hard d) not possible
Any additional comment	

7.2 Metadata questionnaire

DM2E – Metadata Questionnaire WP1

Distribution

Version	Date of sending	Name
0.1	22.05.2012	Julia Iwanowa, Evelyn Dröge
0.2	23.05.2012	Hannes Mühleisen
0.2	23.05.2012	Georg Petz, Max Kaiser (ONB)
0.2	23.05.2012	Chris Bizer, Stefan Gradmann, Violeta Trkulja
1.0	25.05.2012	WP1 content providers

Revisions

Version	Status	Author	Date	Changes
0.1	Draft	Ewelina Suchorzewska, ONB	21.05.2012	Initial version
0.2	Draft	Ewelina Suchorzewska, ONB	23.05.2012	Additions according to Evelyn Dröge and Hannes Mühleisen
1.0	Final	Ewelina Suchorzewska, ONB	25.05.2012	No additions

Organisation

Name of the organisation	
Address	
Zip/City	
Country	

General

Please fill in one questionnaire for each collection you are contributing to DM2E.

Is the metadata already available? If not, what is the expected date of completion?	
Does your system use specific thesauri or vocabularies to classify content? If so please provide some details or links.	
What language is your metadata in?	
Indicate the format(s) of the metadata that your library supports.	
Indicate the possible export formats.	
Can you provide a documentation of the used formats?	
Is the format you use extended by local terms/fields?	
Estimation of required update frequency	
Indicate which fields will not be relevant for the EDM mapping.	
Specify which requirements are necessary for a full mapping to EDM.	
Indicate which information is lacking in the	

metadata but is relevant for the EDM mapping (e.g. edm:rights). See also: http://141.20.126.155/projects/wp2/wiki/Requirements_for_Input_Data__Missing_Information	
Are your metadata records interlinked or should they be interlinked?	
Do references to other vocabularies exist (e.g. DC, PND,...)? If yes, which ones?	
Are there unique element descriptions (e.g. for persons or places)?	
Do default values or descriptions exist that are not explicitly mentioned in the metadata (e.g. data provider; language)?	
Are elements inherited which are not explicitly mentioned (e.g. same author for a book and its chapters or same language for a journal and article)?	
Are different values saved in fields with the same name? A several values saved in the same field?	
Are internal identifiers for books, people, places, subjects etc. exported?	
Any additional comments regarding the metadata of your collections.	

API for accessing data

If you plan to implement one of the beneath stated measures, also indicate when you plan to do so.

Does your system provide some API for accessing objects metadata (e.g. dublin core or other formats)? If so please provide details (supported formats, example calls, etc.).	
Do you think the EDM data model (along with the possible extension) will capture all you metadata? Are there other kind of metadata (e.g. about the digital representations, size, last modification, post date, etc.) that you do not plan to map to EDM?	
Any additional comment	