DELIVERABLE

Project Acronym: DM2E

Grant Agreement number: ICT-PSP-297274

Project Title: Digitised Manuscripts to Europeana

D1.2 – Final Integration Report

Revision: 2.0

Authors: Kristin Dill (ONB)
Evelyn Dröge (UBER)
Øyvind Liland Gjesdal (UIB)
Doron Goldfarb (ONB)
Esther Guggenheim (NLI)
Julia Iwanowa (UBER)
Marko Knepper (UBFFM)
Gerhard Müller (SBB)
Alois Pichler (UIB)
Kilian Schmidtner (SBB)
Klaus Thoden (MPIWG)
Jorge Urzúa (MPIWG)

Contributors:
Lena Stanley-Clamp (EAJC)
Francesca Di Donato (ECORR)
Christian Morbidoni (NET7)
Violeta Trkulja (UBER)

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
0.1	10.07.2013	DG	ONB	Initial draft
0.2	15.07.2013	DG, KS, KT, JI, ED, EG	ONB, SBB, MPIWG, UBER, NLI	Added content provider input in Section 2
0.3	16.07.2013	DG, MK	ONB, UBFFM	Added content provider input in Section 2
0.4	16.07.2013	DG	ONB	Introduction, MINT questionnaire results, Next steps
0.5	23.07.2013	DG, KD	ONB	Corrections, Text for UIB
0.6	01.08.2013	DG, JU	ONB, MPIWG	Add. Text for MPIWG, corrections by KS, MK
0.7	30.08.2013	VT	UBER	Proofreading and some additions
0.8	15.1.2014	DG	ONB	Preparation of final document
0.9	30.01.2014	DG,GM,MK,KT	ONB, SBB, UBFFM, MPIWG	Updates from SBB, UBFFM, MPIWG
1.0	31.01.2014	KT	MPIWG	SILK Evaluation MPIWG
1.1	01.02.2014	DG	ONB	OmNom Evaluation ONB
1.2	04.02.2014	DG, ØG, AP	ONB, UIB	Updates from UIB, Exec. Summary, Introduction
1.3	05.02.2014	DG	ONB	Section 3
1.4	06.02.2014	DG, ED, MK, KS, KT, LSC, EG	ONB, UBER, UBFFM, SBB, MPIWG, EAJC, NLI	Corrections by UBER, UBFFM, SBB, MPIWG, EAJC, NLI
1.5	06.02.2014	KD, AP, CM	ONB, UIB, NET7	Section 6 / Task 1.4; Corrections by UIB and NET7
1.6	06.02.2014	KD, DG	ONB	Included and extended List of Abbreviations from WP2 Deliverable 2.3
1.7	06.02.2014	DG, EG, MK	ONB, NLI, UBFFM	Corrections, Illustrations
1.8	07.02.2014	DG	ONB, UBER, UIB, NLI	Corrections (UBER, UIB, NLI)
1.9	10.02.2014	DG, KD	ONB	Corrections (ECORR, NET7)
2.0	14.02.2014	VT	UBER	Final revision and approval of version 2.0

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This deliverable contains original unpublished work except where clearly indicated otherwise.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

List of Tables

List of Figures

List of Abbreviations

ABO	Austrian Books Online
Aleph	Automated Library Expandable Program Hebrew University of Jerusalem
AlephSeq	Aleph Sequential Data Format
APEx	Archives Portal Europe network of excellence
BAS	Bulgarian Academy of Sciences
BBAW	Berlin-Brandenburgische Akademie der Wissenschaften (Berlin-Brandenburg Academy of Sciences and Humanities)
BRANDEIS	Brandeis University
CHO	Cultural Heritage Object
D	Deliverable
DC(X)	Dublin Core metadata using XML
DoW	Description of Work
DTA	Deutsches Text Archiv (German Text Archive)
DTD	Document Type Definition
CJH	Center for Jewish History
CRNS	National Center for Scientific Research
D2R Server	A tool for publishing the content of relational databases on the Semantic Web
DM2E	Digitised Manuscripts to Europeana
DM2E model	Specialisation of the EDM made by DM2E
DTA	Deutsches Textarchiv
EAD	Encoded Archival Description
ECORR	ERC AdG EUROCORR - "The European Correspondence to Jacob Burckhardt"
ESE	Europeana Semantic Elements
EDM	Europeana Data Model
GEI	Georg Eckert Institute for Textbook Research
GND	German: Gemeinsame Normdatei, also known as: Universal Authority File
GUI	Graphical User Interface
ISMI	Islamic Scientific Manuscripts Initiative
JDC	Joint Distribution Committee
KB	National Library of the Netherlands
LCSH	Library of Congress Subject Headings
MAB2	German: Maschinelle Austauschformat für Bibliotheken
MARC	MAchine-Readable Cataloging standards
METS/MODS	Metadata Encoding and Transfer Schema/Metadata Object Description Language
MINT	Metadata Interoperativity (Services)
MPIWG	Max-Planck-Institut für Wissenschaftsgeschichte (Max Planck Institute for the History of Science)

NLI	National Library of Israel
NTUA	National Technical University of Athens
OAI-PMH	Open Archives Initiative - Protocol for Metadata Harvesting
OCLC	Online Computer Library Center
OCR	Optical character recognition
ODD	One Document Does it all
OmNom	DM2Es interoperability platform
ONB	Österreichische Nationalbibliothek (Austrian National Library)
RAK-WB	Regeln für die alphabetische Katalogisierung in wissenschaftlichen Bibliotheken
RDF	Resource Description Framework
RelaxNG	Regular Language Description for XML New Generation
RNG	See RelaxNG
SBB	Berlin State Library
SILK	Link Discovery Framework for the Web of Data
TCC	Trinity College Cambridge
TEI	Text Encoding Initiative
TOC	Table-Of-Content
UBER	Humboldt-Universität zu Berlin (Humboldt University Berlin)
UBFFM	Universitätsbibliothek JCS Frankfurt am Main (University library JCS Frankfurt)
UIB	Universitetet i Bergen (University of Bergen)
UMA	University of Mannheim
URI	Uniform Resource Identifier
UTF-8	8-Bit Universal Character Set Transformation Format
VIAF	Virtual International Authority File
WAB	Wittgenstein Archives at the University of Bergen
WP	Work Package
XSD	XML Schema Defintion
XSLT	Language for XML-to-XML transformations

1 Executive Summary

This "Final Integration Report" is the second of two deliverables of WP1 "Content". It documents the results of Task 1.3 "Testing the User Interface for Creating Mapping, Interlinking Heuristics and for Configuring the Workflow", Task 1.4 "Setup a test scenario for the prototype platform" and Task 1.5 "Organise the final content integration".

During the past 18 months of the project, providers had to implement their mapping requirements expressed in Deliverable D1.1 ("Requirements Report") into transformation workflows turning their legacy metadata into the DM2E model. In order to achieve this, they used the tools provided by WP2 for creating mappings (MINT), interlinking heuristics (SILK) and for configuring the mapping workflow/data ingestion (OmNom). Task 1.3 was dedicated to record the users' experiences with using these tools in order to assess their applicability for data translation. All of these tools have in common that they offer visual interfaces for the different activities involved, seeking to support novice users in achieving these complex tasks. The results of this evaluation indicate that the visual mapping paradigm on the one hand provides an intuitive way for such users to achieve their goals, but, on the other, also comes with limitations. The input from the users was translated to suggestions to improve the tools, some of which were already addressed.

Task 1.4 was dedicated to create a test scenario for the scholarly research platform. In close cooperation with digital humanities scholars, insight was gained into how they envision the prototype platform might augment their everyday work. These findings supported WP3 in tailoring their suite of digital tools further to the scholarly activities of the user base. In particular, the prototype platform proved useful for helping scholars to come up with digital humanities solutions.

Task 1.5 was to organise the final content integration. Workshops covering the tools provided by WP2 were held during the All-WP-Meetings in Vienna (11/2012), London (06/2013) and Athens (11/2013). A consistency check of the transformed metadata revealed that the data model left providers with some room for interpretation, resulting in some inconsistences in the integrated metadata. In order to address this, a shared activity of defining strict rules for how to represent different aspects of the source metadata in RDF was initiated by WP1. In close cooperation with WP2 and WP3 this resulted in two documents, on the one hand the "Annotatable Content Specification" (Goldfarb, Morbidoni & Eckert, 2013), defining what forms of annotatable content could be provided to the scholarly research platform and how they should be represented in the metadata, on the other hand the document "Recommendations for the RDF representation of DM2E metadata" (Goldfarb & Ritze, 2013), defining how the providers' legacy metadata should be represented in RDF. This included the representation of time (including uncertain timespans), the definition of URIs, the mapping of various forms of contextual data such as index terms and the representation of hierarchies. Both of these documents effectively enabled the providers to map their data in a mutually consistent way.

As of today, seven content providers have successfully mapped and integrated their data. This includes 5 out of the 9 providers listed in the description of work, adding up to about 16.800.000 of the ~ 19.790.000 pages listed in the DoW. In addition, two associated partners were also able to finalise their mappings, adding another ~ 760.000 pages to the integrated DM2E data. Other content providers from the DoW had to postpone their contributions due to delays in digitisation, large-scale infrastructure migration or unresolved mapping issues. However all content providers have made advances with the mapping progress and these delays did not cause side-effects for the other work packages (> 15.000.000 pages integrated) and will be resolved during the next months. Having set up the various provider workflows, the integration of the remaining associated partners will be

straightforward. Regarding acquisition of additional manuscript contents, the expected progress for year one and two was already exceeded during the first year. This achievement was further extended by the enlisting five additional content providers.

2 Introduction

Within DM2E WP1, months 7-24 of the project were dedicated to work towards the final content integration (Task 1.5). This included activities to evaluate and make providers acquainted with the tools developed throughout the project, to further refine the DM2E model with respect to content provider requirements and to set up a test scenario for the digital scholarly research platform.

Content providers were introduced to the provided tools MINT, OmNom and SILK in workshops held during the All-WP-Meetings in Vienna (11/2012), London (06/2013) and Athens (11/2013). They started to implement the conceptual mappings first expressed in the D1.1 requirements report into concrete transformation workflows either in form of XSL-transformations or through services such as the D2R framework or own interfaces.

After the users collected their first experiences with the provided tools, they were asked to evaluate (Task 1.3) them in order to provide feedback to the developers. This shared testing phase resulted in valuable suggestions for improvement.

The first results of the completed data transformations were compared for consistency issues and other open questions such as how to represent annotatable content. This activity resulted in two documents as addendum to the DM2E model specification with practical and necessary directives for how to represent DM2E metadata. After resolving the open inconsistencies the content providers were able to proceed with the final content integration.

Two case studies were conducted by the ONB in close cooperation with members of WP3 (NET7, UIB, ECORR, UBER) in order to assess the prototype platform for digital humanities scholars (Task 1.4). T1.4 resulted in improvements in the technical functionality of Pundit and in an insight into how digital humanities scholars envision using the platform to facilitate their own scholarly activities.

2.1 Scope and Objective of this Document

This document wraps up the activities of the tasks 1.3, 1.4 and 1.5 within DM2E WP1 between months 7-24. It summarises Task 1.3's results of the evaluation of the provided tools and presents the outcomes of the case-study on the scholarly prototype platform carried out in Task 1.4. Moreover it states about the current state of the final content integration in Task 1.5. In the latter respect it contains detailed descriptions of each content provider's progress to date - For detailed information about mapping tables and transformation scripts, however, please consult the Section "DM2E Content Provider Mappings" on the public DM2E Wiki[1].

The final content integration is an on-going process. New Content providers who joined during the course of the project are at various stages of this process, as is also the case for some content providers affected by delays out of various reasons. Therefore, from time to time still new requirements arise that lead to adaptions to the DM2E model and the related specifications. However, the more providers finalise their mappings, valuable mapping

[1] http://wiki.dm2e.eu/Main_Page

templates emerge that can be adapted by other providers, thus reusing the experience and knowledge accumulated throughout the project.

2.2 The Structure of this Document

Chapter 3 reports about the results of Task 1.5, the Final Content Integration. It gives an overview on the integrated content to date and discusses the current state of the process.

Chapter 4 continues the report about Task 1.5 by giving detailed overview on the state of the content providers in the project.

Chapter 5 summarises the results of Task 1.3 "Testing the User Interface for Creating Mapping, Interlinking Heuristics and for Configuring the Workflow"

Chapter 6 presents the outcomes of the case study "Setup a test scenario for the prototype platform"

Chapter 7 summarises the document and gives an outlook for WP1

3 State of the Final Content Integration

This Section reports on the state of Task 1.5, the final content integration. It gives an overview on the content providers currently participating in the project and updates information gathered from newly joined partners. It outlines the current state of content ingested to date and discusses further steps. Detailed information about each content provider is given in Section 4.

3.1 DM2E Content Providers – Overview on Metadata & Content

As of February 4th, there are 16 content providers involved in DM2E. Of the 16 providers, five are members of the consortium, namely UBER, MPIWG, ONB, SBB and UIB. Consortium member EAJC serves as aggregator for three more providers, CJH, JDC and BRANDEIS. Seven more providers, BBAW, KB, GEI, BAS, UBFFM, CRNS, ECORR and NLI participate as associated partners, with NLI being also closely associated with EAJC.

Of the 16 providers, BAS, BRANDEIS, ECORR, GEI and KB joined after month 6 of the project. Table 1 therefore extends Tables 2 and 3 from Deliverable D1.1 about the content & metadata questionnaires with the respective information gathered from the new project partners. The updated Deliverable D1.1 Table 1 "Underlying Content", including the content provided to date, can be seen in Table 2.

	GEI	BAS	KB	BRANDEIS	ECORR*)
Metadata available	Yes	Yes	Yes	Yes, for parts of the collection	Yes
Language of metadata	German	English	English & Dutch	English	-
Metadata formats	METS/MODS	TEI P5	DC(X)	EAD / MARC	TEI P5
Digitised objects available	Yes	External	Yes	Yes, for parts of the collection	-
Delivery of full-texts	Yes	Yes	-	Parts of the collection	-
Format of full-text	OCR	Transcription	-	Transcription	-
Character encoding of full-text	UTF-8	UTF-8	-	-	-
Textual mark-up	No	Yes	-	No	-
Granularity of objects	Up to page level	Distinct texts within manuscripts	Page, Illumination within page	Item Level Single Image	-
Persistent Identifiers available	Yes	Yes	Yes	No	-
Usage of specific thesauri or vocabularies	GND	-	Iconclass		-

Table 1: Main results of content & metadata questionnaires from new content providers joined in 2013

*) No questionnaires received yet

Taking the metadata formats used by the new partners into account, the current distribution of formats throughout the project clearly shows that formats such as TEI, MARC and EAD are used more often than others. Having formats used by multiple content providers is of a

big advantage, as mappings and transformation workflows created in DM2E can already be reused within the context of the project itself. Figure 1 shows the current distribution of metadata formats in DM2E. The implications of this distribution on the project's progress are discussed in the next subsection.

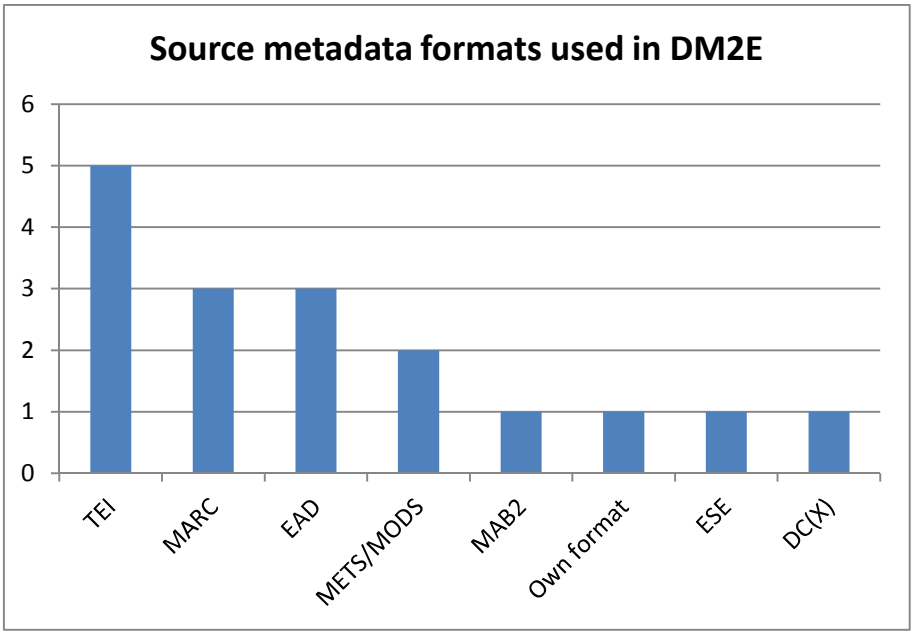

Figure 1: Distribution of source metadata formats used in DM2E.

3.2 Mapping & Transformation

After the content providers had specified their requirements in Deliverable D1.1, they proceeded with adapting their data accordingly and to set up their internal workflows for exporting the data for the subsequent mapping & transformation steps. At the same time, the DM2E model, then still called EDM+, was continuously updated to the latest requirements.

In the meantime until version 1.0 of the DM2E model became available, some content providers had already started to develop their own transformations to EDM and to the DM2E model. Some providers used D2R to provide direct RDF based access to their underlying relational databases. Others developed XSLT based workflows for converting their source data or created a direct RDF export service, reachable via command line or browser.

With the specification of the DM2E model version 1.0[2] in 04/2013, however, it was also possible to set up a test environment for the MINT mapping platform. Initial tests by UBER and ONB were performed and MPIWG and ONB designed a questionnaire for evaluating the functionality and the usability of MINT for mapping provider source data to the DM2E model. The resulting questionnaire can be found in the Appendix in Section 9.1. The questionnaire was sent out on May 28th, 2013; by July 16th 2013, five questionnaires had been returned. The results suggest that it is possible to use MINT as mapping interface to DM2E for certain source XML schemas. Other XML schemas are, however, less suitable to be mapped with MINT. A detailed discussion of the results of the evaluation can be found in Section 5.1.

As a consequence, content providers started to create their mappings in MINT, which gave them a good starting point to get acquainted with the data model. At some point, however,

[2] http://assets.dm2e.eu.s3.amazonaws.com/DM2E_Model_V1.0_Specification.pdf

they exported the XSL-transformation from MINT and continued their mapping by hand. At this point it became clear that the synergies arising from the previously mentioned fact that many of the metadata formats were used by multiple providers could be utilised for the project. Therefore, transformations created by a content provider for a specific format were reused within the project.

3.2.1 Reuse of Existing Data Mappings within the Project

To date, this was the case with TEI mappings created by SBB for BBAW, which were adapted for the data provided by UIB. UBER, also providing TEI based data, developed their own transformations in close cooperation with SBB. TEI data by BAS and ECORR will also be considered in this regard. The same took place with the transformations created by UBFFM for METS/MODS data, which were also used for the data provided by GEI.

ONB created their MAB mappings both for the codices and the abo collections. Since ONBs source metadata was exported as MAB embedded within a MARCXML container format, their mappings will be adapted to the MARC based data provided by NLI and CJH.

SBB is working on an EAD based transformation for their data. In this regard, SBB is in touch with the APEx[3] project currently developing a mapping from EAD to the EDM. The mappings arising from this process will be subsequently adapted to the data provided by JDC and BRANDEIS.

Synergies between the DC/ESE based format used by CRNS and DC(X) used by KB will be considered.

MPIWG developed a transformation for their own internal data format.

3.2.2 Refinement of the Data Model and Additional Mapping Specifications

Continuous provider feedback on the mappings created with the DM2E model up to and including version 1.0 led to its next iteration version 1.1. Its latest revision served as the foundation for the final content integration. Please consult (Dröge, Iwanowa et al., 2014) and Deliverable 2.3 for details. This version implements a number of additional provider requirements that emanated throughout the course of the project.

The first requirement dealt with the options how to provide content for the scholarly research platform developed in WP3. Some content providers were not able to support the initial WP3 requirements for providing annotatable content as specified in Deliverable D3.1 because the required provision of specific "annotatable content" html tags demanded structural changes in their data infrastructures they were not able to implement. Therefore, WP1, WP2 and WP3 worked together on specifying alternative means of providing content to WP3 which included the possibility to provide "raw" page scans or plain text instead. The resulting document "Annotatable Content Specification" (Goldfarb, Morbidoni & Eckert, 2013) defines the available options for providers to contribute their content to DM2E.

The second requirement was based on the rich possibilities originating from the linked data based paradigm of the DM2E model itself. The option to map contextual classes such as persons, places and subject terms in addition to "regular" bibliographic data enabled providers to represent their data in far more depth than it had been possible before.

[3] http://www.apex-project.eu/

Moreover, hierarchical relationships between physical or conceptual levels within the provided data had to be represented appropriately.

Although the data model catered these requirements from the beginning, a consistency check performed on preliminary mapping results from various providers in November 2013 revealed that the model left providers with some "interpretational range" regarding how to represent certain aspects of their data such as subject terms for different classes such as persons, places, timespans, works and genre. Moreover, different "levels" of controlled vocabulary (external authority, internal controlled vocabulary, uncontrolled vocabulary) had to be dealt with as well, especially with respect to the design of the URIs identifying the underlying entities. Another issue resulted from the observation that providers mapped hierarchical relationships in different ways.

In a collaborative effort to harmonise these different interpretations of the data model, WP1 and WP2 drafted the document "Recommendations for the RDF representation of DM2E metadata" (Goldfarb & Ritze, 2013) which enabled the content providers to proceed with the final data integration.

3.2.3 Documenting the Mapping & Transformation Workflows

The documentation of the providers' mappings and their specific transformations can be found on the DM2E Wiki[4] established in cooperation with WP4. This way, they can be kept up to date due to the continuous refinement of the data model and the individual provider workflows. Moreover, they will be easily accessible to interested parties.

3.3 Integrated Metadata & Content to Date

Table 2 shows a combined view on the underlying content including all additional providers who joined DM2E until February 2014. To date, seven content providers have successfully integrated (parts of) their data, covering 17,553,494 pages of content.

The datasets have been ingested using a Direct Ingestion Tool provided by WP2. Providers also evaluated the User Interface of the Web based frontend to the workflow engine, OmNom. The results of this evaluation can be found in Section 5.2. Moreover, initial tests have been performed with the SILK link discovery framework. A discussion of its user interface is featured in Section 5.3.

Due to various reasons, some content providers have not been able to integrate (all of) their data yet. Providers who joined DM2E at later stages of the project will integrate their content during the next months. The data from those providers contributing EAD records will be included once an initial transformation workflow will have been established by SBB, who are cooperating with the APEx project in that matter. Some providers had to postpone (parts of) their intended contributions due to delays in digitisation projects, while others were currently involved in large-scale migrations of their digital infrastructures. All of these delays are expected to be resolved during the next months.

[4] http://wiki.dm2e.eu/Main_Page

Institution	Type	Volumes according to D1.1	Volumes integrated 02/2014	Pages according to D1.1	Pages integrated 02/2014	Language	Format	Metadata	IPR
CJH via EAJC	books, posters, photographs, etc.	29,000	To be delivered	400,000	To be delivered	various	JPG2000, TIFF	MARCXML, DC, METS, EAD, occasionally MODS	Fair Use clause, Section 107 US Copyright Law: http://www.copyright.gov/title17/92chap1.html#107
	audio files	1,200	To be delivered		To be delivered	various	MP3		
	movies	100	To be delivered		To be delivered	various	mpeg-4. mpeg-2, mov		
JDC via EAJC *)	records of the American Jewish Joint Distribution Committee of the years 1914-1918			20,000	To be delivered	Various	PDF	EAD	http://archives.jdc.org/researchers/terms-and-conditions.html
Brandeis via EAJC *)	posters	100-200	To be delivered			Spanish/Catalan/ English	JPG	EAD / MARC	CC-BY-NC-SA
NLI	manuscripts	1,000		55,000	To be delivered	Hebrew	TIFF / JPEG	MARC21 / DC / Olive	public domain / content available free of charge for non-commercial use
	manuscripts	750-1000		50,000	To be delivered	Yiddish			
	books	7,500		2,100,000	To be delivered	Ladino			
	books	2,500		700,000	To be delivered	Hebrew			
	Archival material / portraits			3,000	To be delivered	Yiddish			
ONB	books - ABO	22,500	11,816	6,750,000	4,236,276	German	JPG2000	MAB2	Content public domain, available free-of-charge for non-commericial
		5,500	4,742	1,650,000	1,983,491	Latin			
		4,000	7,608	1,200,000	3,564,784	French			
		3,000	7,789	900,000	2,596,361	Italian			
		2,500	31	750,000	10,944	Czech			
		2,500	2,181	750,000	1,040,540	English			
		10,000	5,706	3,000,000	2,163,794	various			
	manuscripts - codices		51	18,600	14,448	Latin	JPEG or TIFF		
			25	6,200	5,176	German			
			9	4,800	4,216	Greek			
			85	12,800	22,264	various			

Institution	Type	Volumes according to D1.1	Volumes integrated 02/2014	Pages according to D1.1	Pages integrated 02/2014	Language	Format	Metadata	IPR
SBB	Personal papers of Adelbert von Chamisso			20,000 +10,000	To be delivered	German	JPEG or TIFF	EAD	CC BY-NC-SA
	Personal papers of Jean Paul			17000	Cancelled				
	Personal papers of Theodor Fontane			7000	Cancelled				
	Personal papers of Arthur Schopenhauer			10000	Cancelled				
	Collection of Margarete Hauptmann			2000	Cancelled				
	Personal papers of Dietrich Bonhoeffer			20000	Cancelled				
	Personal papers of Gerhart Hauptmann			30.000 +15,000	To be delivered	German			
UBER	Polytechnisches Journal	346		205,000	181,549	German	PNG, JPEG	TEI P5	Text:CC BY-NC-SA 3.0 Img: CC BY-NC-ND 3.0
UBFFM *)	Hebrew manuscripts and Medieval manuscripts	549	1,012	230,000	307,202	Hebrew; Latin; German; Various	TIFF, JPEG	METS/MODS	public domain
GEI *)	Textbooks	2,800	2,861	500,000	456,324	german	TIFF, JPEG	METS/MODS	public domain
UIB	Autographs - Wittgenstein	20	20	5,000	5,000	German, English	JPEG	XML; TEI P5	CCPL BY-NC-SA
MPIWG	books, newspapers	2,500	1,272	500,000	532,379	various	JPEG, TIFF, XML	own format	CC-BY-SA
	manuscripts	1,000	To be delivered	100,000	To be delivered	various			
	autographs	4	22	1,200	10,138	various			
BBAW	books / historical printings from 1650-1900	1,300	1,300	470,000	418,608	German	XML	TEI P5	CC-BY-NC
DFGA/CRNS	Nietzsche's manuscripts additional 10.000 pages in 2013			19,330	To be delivered	German	JPEG	DC/ESE	CC-BY-NC

Institution	Type	Volumes according to D1.1	Volumes integrated 02/2014	Pages according to D1.1	Pages integrated 02/2014	Language	Format	Metadata	IPR
BAS *)	Bulgarian Academy of Sciences / Codex Suprasliensis	1	To be delivered	570	To be delivered	Old Church Slavonic		TEI P5	CC BY- NC-SA
KB *)	manuscripts / illuminations	400	To be delivered	11,141	To be delivered	various	JPEG	DC(X)	public domain
ECORR *)	The European Correspondence to Jacob Burckhardt	To be delivered	To be delivered	2,834	To be delivered	To be delivered	JPEG	TEI P5	CC BY-NC-SA
		Pages total		**20,490,475**	**17,553,494**				

Table 2: Underlying content including newly acquired providers, content ingested to date --- *) Not listed in DoW

4 Current State of the Content Providers

This Section contains detailed descriptions of each content provider. Providers are grouped by the metadata format used:

- UBER, BBAW, UIB, BAS and ECORR use TEI
- SBB, JDC and BRANDEIS use EAD
- ONB uses MAB, NLI and CJH use MARC
- UBFFM and GEI use METS/MODS
- MPIWG use their own format
- CRNS use DC/ESE
- KB use DC(X)

As mentioned previously, mappings and transformation workflows produced within these groups should be reused as often as possible.

4.1 UBER – Humboldt-Universität zu Berlin

The Dingler[5] records which were provided for the first test mappings came directly from the owner and creator of the metadata, the Institute for Cultural History and Theory. They were described by the TEI-P5-XML standard without using an additional XML schema. The currently imported records are provided by the DTA and are based on the RelaxNG schema. Therefore, it was necessary to update and adapt the XSLT script that is used for the mappings to the new TEI-XML schema.

Furthermore, the current records coming from DTA do also include more contextualisation information, like the GND URIs for institutions and persons.

Overall, all mapping requirements were considered and integrated into the DM2E model v1.0. All mandatory elements of the model could be mapped. In some cases, additional text strings were used that are not part of the TEI-XML records, e.g. "de" as a language attribute. In order to fulfil the requirements of the WP3 tool Pundit, single pages of the journal will be mapped as well. Single pages have their own identifiers in the TEI-P5 record but do not link to Web resources. These links still have to be created from the given identifiers. Currently, articles and journals link via the property *dm2e:hasAnnotatableContent* to a Website with facsimile and transcription of the journal[6].

4.1.1 Background: General Introduction to the Metadata Format TEI

TEI stands for Text Encoding Initiative, which is a consortium for the contributed development of a standard metadata format for the representation of texts in digital form. The provided guidelines by the initiative are standard specifications for encoding methods for machine-readable texts. The TEI guidelines are widely used by libraries, museums, publishers and individual scholars to present texts for online research, teaching and preservation. The most recent version of the guidelines is TEI-P5[7].

Concerning the generic structure of the whole TEI tagset and the different specific projects and user groups, the guidelines can be customised to meet the particular requirements of several specific use cases. Until these days, there are some well-documented TEI customisations, which are widely used by the community. One of the most popular customisation is the so called TEI Lite[8]. The TEI Lite subset is quite simple and can be learned relatively easy. Therefore, this subset has been widely adopted and is one of the most used TEI subsets. The current version of TEI Lite is derived from the P5 version of the guidelines. Other customisations provided by the consortium are TEI Tite, Bare, All, Corpus, MS, Drama, Speech and others. All schemas are available as ODD, DTD, RNG and XSD[9].

In the case of the Dingler records the schema language that is used to describe the elements is non modified TEI. No other elements have been added to the schema. All used classes are coming from the original TEI full set. The logical description of the records is following the recommendations of the Guidelines. The only modification that has been done was the excluding of not used elements and the definition of limited value lists for some elements[10].

[5] The "Polytechnisches Journal", often named "Dingler" based on the name of its first publisher, is provided by the Humboldt-Universität zu Berlin (UBER).
[6] See http://dingler.culture.hu-berlin.de/article/pj001/ar001001 as an example (12.07.2013).
[7] Definition of the TEI-P5 guidelines: http://www.tei-c.org/Guidelines/P5/ (20.01.2014).
[8] Definition of TEI-Lite: http://www.tei-c.org/Guidelines/Customization/Lite/ (16.01.2014).
[9] For detailed Information please refer to: http://www.tei-c.org/Guidelines/Customization/ (20.01.2014).
[10] Documentation of the Dingler XML schema: http://dingler.culture.hu-berlin.de/Schema/dingler.html (20.01.2014).

4.1.2 Adaption of Mappings to Different Model Revisions

For the finalised version of the mapping to the DM2E model, DM2E got local copies of the last modified TEI-XML metadata records of the complete journal on volume and on article level. The mappings were created and tested with these records. The current mapping is based on the first test mappings which were carried out using the "DM2E v1.0 Fixed Ranges" schema in MINT. Two different *ore:Aggregation* and *edm:ProvidedCHO* classes were created: one for a journal issue, another for a journal article. After the first mapping circle with MINT, which already included about two-thirds of the first mapping, further mapping steps were carried out by manually working on the MINT output (supported by the Oxygen editor). This was mainly done due to readability reasons (the output file was split up into different files for the creation of journal issues and articles), to reduce redundant steps in the mapping workflow (URIs of all classes were created as variables instead of typing them repeatedly) and to include steps that were not possible to proceed with MINT (e.g. normalising URIs or the creation of titles for smaller CHOs). Furthermore, the mappings were first created for the DM2E model v1.0 and then manually adapted to DM2E v1.1, revision 1.0. It was much easier and faster to do this step by hand than by repeating the whole mapping in MINT.

The structure of the XSLT custom script is based on the XSLT script provided by SBB and further developed for the requirements of the institute for cultural studies at UBER.

The TEI data of the Dingler records are mapped on journal, issue, article and page level since almost all TEI documents encode full texts. Basic provider descriptive metadata from the TEI header is transformed in DM2E without any loss of data. Missing mandatory elements that the DM2E model requires are completed by default values.

Although all TEI-encoded full texts of the DTA are based on philological methods, there are almost no semantically marked up persons, corporate bodies, or other subjects. In order to produce not only RDF literals, but URI references (resources), full text literals have to the transformed into URIs during the mapping or have to be extracted and processed by SILK in a second step, the contextualisation.

4.1.3 Representation of Hierarchical Levels

The TEI records include a representation of the hierarchical structure of the journal. The top-level is described within the TEI-header on article level and includes the basic metadata about the physical journal and about the online journal as well. The metadata on the journals is mapped to the top-level CHO, which is related to the sub-level-CHOs on the next level, the issues of the journal, via the *dcterms:hasPart* property. Issues include articles, which in turn gather CHOs on the lowest representational level in the object hierarchy: the pages. All top-down hierarchical relations are described by *dcterms:hasPart* and respectively with *dcterms:isPartOf* for all bottom-up relations, as these are inverse properties. Figure 2 illustrates the hierarchical concept in the Dingler records. The linear relations between the resources on one level are defined with the property *edm:IsNextInSequence* as proposed in the Europeana Data Model specification.

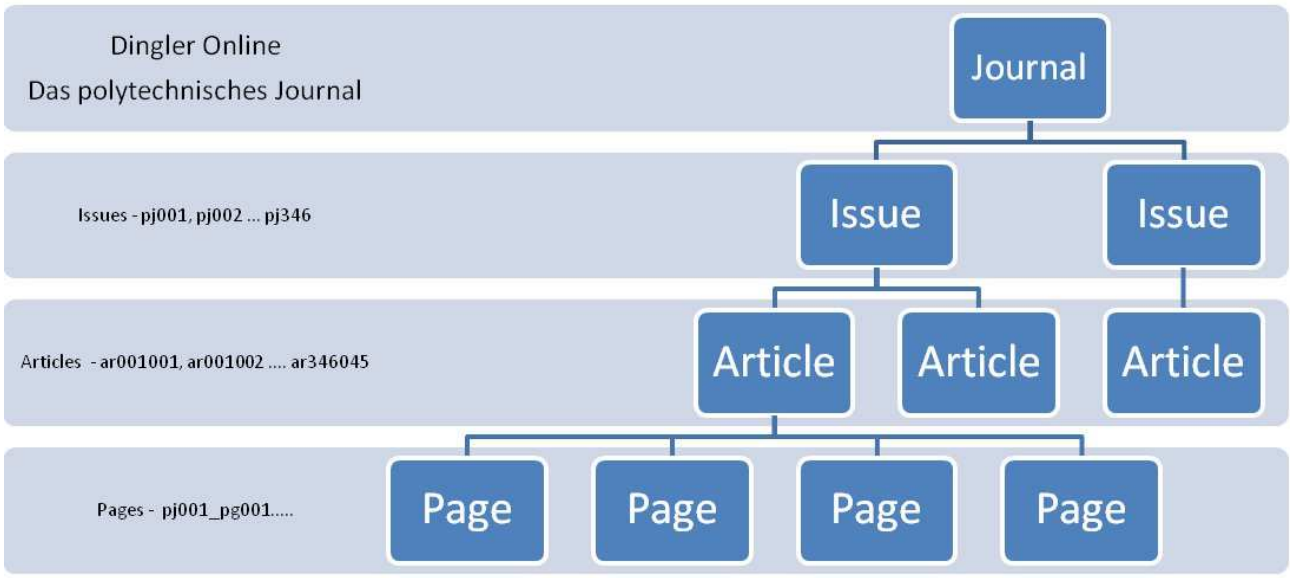

Figure 2: Hierarchical Structure of the Dingler Journal.

4.2 BBAW – Berlin-Brandenburg Academy of Sciences

The German Text Archive (Deutsches Textarchiv, DTA) of the Berlin-Brandenburg Academy of Science provides over 800 manuscripts published from the 17th to the 20th century. All full texts of these manuscripts are encoded as TEI-XML.

The analysis and the transformation of the TEI-XML documents have also been carried out by the Berlin State Library (SBB).

4.2.1 Tool-Development: TEI2EDM, DM2E-EDM.xsd

SBB developed an XSLT-script called TEI2EDM for transforming the TEI-XML documents to the EDM/RDF and DM2E formats.

The script supports scenario testing by creating examples for transformations and the demonstration of an XSLT-Module implemented in OmNom recently. It is also a means for analysing differences between TEI documents from various institutions using several different dialects of TEI. These dialects often result from case-by-case digital edition project requirements.

Therefore, main objective was the development of a single reliable and generic TEI transformation scenario that can be applied to various TEI dialects of different content providers such as UIB and BBAW.

After the Europeana Data Model was finalised, the code for the transformation was reviewed and tested. Three modules constitute the transformation code:

- A set of core modules processing the data,
- Switches for the export format (EDM-DM2E/RDF) and
- Modules for provider specific transformation requirements.

When transforming input data (TEI-XML documents), there are three different export qualities to choose from:

- EDM: mainly transforming input data into RDF literals
- EDM-RICH: when transforming input data, literal values will be replaced by URI references whenever possible
- EDM-DM2E: application of DM2E specifications incl. DM2E namespaces, sub-properties, granularity pages, and paragraphs

As a result of the code revision and scenario testing, it must be underscored that the application of a standard like TEI is not a sufficient condition for comparability of data using the same standard. In almost every project, an individual approach to the way a standard will be used by project members is defined. This also holds true for various projects within a single institution.

Nevertheless, it turned out that two content providers (SBB, UIB) can apply the same transformation code with individual adjustments but without interfering with each other. The code is available via a private, joint repository.

In addition, SBB maintains a DM2E-XML-Schema which is derived from the EDM-Schema. It is used for quality assurance by validating transformed data.

4.2.2 The Crosswalk

SBB compiled a crosswalk for TEI data of the DTA on file and page level since almost all TEI-documents encode the full text of manuscripts or books.

Basic metadata from the TEI-header are transformed in EDM without any losses of data. Missing mandatory elements are completed by default values.

Although all TEI-encoded full texts of the DTA are based on philological methods, there are almost no semantically marked up persons, corporate bodies, or other subjects. In order to produce not only RDF literals, but URI references, terms of the full text have to be extracted and processed by SILK in a second step.

4.3 UIB – University of Bergen

The Wittgenstein Archives at the University of Bergen (WAB) contributes 5.000 pages of the Wittgenstein Nachlass kept at Trinity College Cambridge (TCC), Wren Library, Cambridge, in facsimile and text editions as well as metadata.

After the definition of the relevant metadata for DM2E, UIB proceeded with their improvement. This was followed by a conceptual mapping to the DM2E model. For the technical mapping, SBB's TEI conversion scripts were adapted to UIB's requirements. With the adapted transformation, UIB was ready for data ingestion.

The Wittgenstein Archives refer to no outside extensions of the TEI standard and use TEI P5 elements and attributes. However, there are some specific practices for encoding the documents. For instance, text tokens and prefixes in attributes like @ana and @xml:id (for example on tei:ab elements) and @key (for example on tei:rs elements) are used in order to represent multiple values and different properties. For instance: The @ana attribute for Bemerkung xml:id="Ms-114,145r[2]et145r[3]» looks like ana="field:PhilosophyOf Language_pub:W-PG:PartI:IX:sect114_date:19331001*-19331231*". It is composed of the values

- field: which is mapped to subject,
- pub: which is not mapped,
- date: which is mapped to dcterms:issued as timespans.

The same type of practice is also used on other items like for example references to external works: there the rs element and the @key attribute are used.

The Wittgenstein data use the structure of the document itself to generate the xml:ids that represent the different digital objects. For instance a "Bemerkung" xml:id="Ms-114,145r[2]et145r[3]» is composed by the top id, plus, separated by a comma, the page id(s) (145r) (r or v for recto or verso), plus the sequential number of the text block(s) out of which the Bemerkung is composed ([2]et[3]). Given this clear structure with id's on the item, and id's on the tei:ab descendants, the represented hierarchical depth was decided to be on the level of (1) Item and (2) Bemerkung. The issue of building hierarchies with overlapping pages etc. was avoided by only using the two mentioned levels.

Knowledge embedded in the id's was used to map each page mentioned in a Bemerkung as a Web resource, and the Bemerkung itself in different versions and views (normalised, diplomatic, facsimile), reusing the URI's for the mappings. Thus, much of the mapping process was utilizing the knowledge of how the id's were used by other Web applications.

4.4 BAS – Bulgarian Academy of Sciences

The Bulgarian Academy of Sciences (BAS) intends to provide DM2E with the full electronic corpus and digitised images from the 10th century Codex Suprasliensis, listed in the UNESCO's Memory of the World Register since 2007.

The content- and metadata questionnaires have been received. A preliminary mapping for the metadata for the Codex Suprasliensis was completed and will soon be followed by the implementation of a corresponding transformation workflow.

4.5 ECORR – EUROCORR

EUROCORR is an ERC AdG project which aims at publishing the European Correspondence to Jacob Burckhardt (1842-1897). The project's main output is the platform Burckhardtsource.org, which hosts the manuscripts facsimile as well as XML TEI P5 transcriptions and semantic annotations on the letters.

EUROCORR intends to provide metadata and facsimiles from its digital collection to the project.

4.6 SBB – Berlin State Library

The Union Catalog for Personal Paper and Autograph Collections, Kalliope, at the Berlin State Library (SBB) provides access to several thousand finding aids mainly from holdings in archives, libraries, museums, and research facilities in Germany. For the last deliverable the initial situation was analysed and documented. This included available metadata formats like MAB2, EAD, and the relational database backend.

Furthermore, the deliverable stated that the Berlin State Library would make some of its outstanding collections accessible via Europeana during the project. Among these are the

personal papers of Gerhart Hauptmann, who won the Nobel Prize in Literature in 1912, and of Adelbert von Chamisso, who is not only commonly known for his poetry and research in Botany, but is also recognised for his liberal European ideas.

All objectives of the last period reporting period were accomplished. As agreed upon in the DoW, an in-depth analysis of the proprietary relational database backend, which is composed of 201 database tables with 4932 columns in total, was completed. In order to develop a crosswalk between the relational database and the Europeana Data Model (EDM), it had been necessary to identify the relevant tables and columns in the first place.

After the first draft of the crosswalk was finalised, the D2R-Server was implemented and tested internally. After several adjustments, the server has been running for test purposes since March 2013. The crosswalk has also been implemented and, as a result, all data from the database are accessible as RDF. However, it is still necessary to identify a small number of D2R server errors for it currently cannot handle load by too many requests.

As a preliminary summary it can be stated that in close collaboration with cataloguers all parts of the database backend have been analysed. The crosswalk between database columns and EDM fulfils all requirements. In addition to that, a D2R-Server was set-up and configured. In line with the DoW, all data are accessible as EDM/RDF, now.

Notwithstanding, it should also be mentioned that neither the problem of the stability of the D2R-Server nor of the security of the data for over 900 institutions, not being partners of DM2E themselves, has been sufficiently solved yet. It is therefore crucial to deal with both tasks over the next few months.

Last but not least, the Berlin State Library is currently working on implementing a new data model for the Union Catalogue based on the XML-Schema Encoded Archival Description, EAD. This Schema is commonly used by archives for encoding archival collections such as personal paper and autograph collections. It should therefore be examined if an RDF/EDM workflow can be adopted and integrated into the new XML-based backend infrastructure. This will be a sustainable approach and will be done in close collaboration with the APEx[11] project.

4.7 JDC via EAJC – Joint Distribution Comittee

JDC will deliver the Records of the New York Office of the American Jewish Joint Distribution Committee of the years 1914-1918[12] as EAD records, with the support of EAJC and NLI. The following changes to the current EAD XML files are necessary in order to make them valid EAD:

- The component element ID attribute must be an alpha-numeric string beginning with alpha characters but is currently a number
- Component element level attributes prohibit capitals and spaces both of which are present in the current XML; ideally, valid values should be used for the attribute to represent JDC's hierarchy (subfonds for subcollection, recordgrp for record group, file for file)
- All of the dates in JDC's XML finding aids are formatted MM/DD/YYYY. Valid EAD requires that these data conform to the ISO 8601 standard for dates, which is YYYY-MM-DD format with a forward slash separating date ranges.

[11] http://www.apex-project.eu
[12] http://archives.jdc.org/researchers/finding-aids/new-york-office/1914-1918/

A very important aspect of JDC's mapping process will include inserting both a copyright statement and a note acknowledging the lead donor into all shared records. JDC is currently working on the specific wording of these statements.

After export of the full data set, NLI will test the APEx tool with the JDC data in order to assess its usability for DM2E.

4.8 BRANDEIS via EAJC – Brandeis University

The University Archives & Special Collections Department of Brandeis University will provide DM2E with data on selected collections to DM2E via Judaica Europeana with the support of EAJC and NLI.

The content- and metadata questionnaires were filled out and returned to ONB. Most of the provided digital items are photographs of artefacts and other objects held in the collections.

Brandeis is in the process of unifying their catalogues into one central system. For most collections the archiving system Archon[13] is used. The metadata can be exported from Archon as EAD records, which include collection level metadata and information about single items in the collection. There are, however, no image links in the exported EAD files. Brandeis is currently examining ways to include image links into the export and will be supported by NLI and ONB in this regard. Mapping of EAD to DM2E and EDM still has to be tackled by other providers, e.g. JDC and SBB, and the Brandeis records will be reviewed in this process.

Other collections are currently in the state of being uploaded into an institutional repository called DSpace[14]. This repository will allow for the exportation of metadata as MARC records. The mapping of the Brandeis MARC21 metadata will be based on the work done for NLI and CJH, with support from NLI.

4.9 ONB – Austrian National Library

Throughout the project, ONB will provide material from two of its collections. One part consists of 170 digitised manuscripts from ONB's manuscript collection[15] (CODICES), while the other consists of 50,000 books digitised throughout ONB's collaborative project with Google, Austrian Books Online[16] (ABO).

The metadata for both collections reside in ONB's Aleph[17] library system and are catalogued according to the RAK-WB rules used in German-speaking countries. The internal metadata model of Aleph based library systems is called AlephSeq, but the cataloguing and data exchange takes place using the MAB2 data model that is closely tied to the RAK-WB cataloguing rules.

For Austrian libraries, country specific extensions[18] have been added to the common MAB definition and thus have to be taken into account when providing data to DM2E or Europeana in general. Moreover, some differences between the internal MAB model used within Aleph

[13] http://www.archon.org
[14] http://www.dspace.org
[15] http://www.onb.ac.at/ev/collections/7692.htm
[16] http://www.onb.ac.at/ev/austrianbooksonline
[17] http://www.exlibris.co.il/category/Aleph
[18] http://www.obvsg.at/fileadmin/files/kat/formate/Sonderkategorien.pdf

based systems and the general MAB2 specification[19] have been discovered during ONB's mapping process to the DM2E model. Therefore, the MAB2-DM2E mappings produced by ONB are "out-of-the-box" mainly applicable to libraries in German-speaking countries using Aleph based library systems.

Nevertheless the effort to adapt ONB's mapping to other MAB2 based source models will be manageable if necessary. The metadata is exported from the library system using Aleph's X-Server[20], downloading an XML record for each provided item into a collection-specific folder. The X-Server returns the metadata in the Aleph specific MAB2 model encoded using the MARCXML[21] format. This has the advantage that metadata mappings from MAB2 to DM2E based on MARCXML can easily be adapted to other source data models that use the same underlying data format. The transformations generated could for example be adapted to mappings from MARC21 encoded in MARCXML mainly by adjusting the respective field numbers and indicators. One disadvantage of using the MARCXML data format is, however, that it is not ideally suited for creating mappings with the MINT platform – see Section 5.

Although both collections are catalogued based on the same rules and using the same data model, there are quite some differences in the actual metadata records of the respective items. Moreover, the metadata exported from the library system only cover information on item level and not on page level. In order to provide the granularity required by the WP3 annotation platform for annotating manuscripts from the CODICES collection, additional data sources have to be mapped accordingly. The following two subsections describe the individual differences of the two provided collections with respect to the metadata records.

4.9.1 Codices

Each of the 170 items provided from the CODICES collection represents one manuscript. There are no hierarchies between single items in the sense of multivolume works or collections, although some of the items themselves contain different works that are not represented in the metadata except for in a general table-of-content (TOC) field. The option of an automatic splitting of such metadata records based on the TOC field will have to be evaluated separately – currently, such items will be provided as a whole including the respective TOC field.

The metadata for the 170 items contain detailed information about the roles of persons involved in the history of each item, including authors, painters and writers, but also current and previous owners. All of these different roles are based on a controlled vocabulary and each role has been implemented in the DM2E model.

Entries for presumed person roles such as "presumed author", "presumed previous owner", etc. are currently handled as if they were correct, i.e. mapped to the respective "author", "previous owner" roles. Moreover, dates related to manuscripts are sometimes not known exactly and therefore represented in form of estimates such as "First third of the 13th century", subsequently being catalogued as timespans such as 1200-1233. The need for representing such 'uncertain' information has also been expressed by other DM2E content providers and has already been included in or will be addressed in future versions of the DM2E model.

[19] http://www.ubka.uni-karlsruhe.de/hylib/mab/mab2.html
[20] http://www.exlibrisgroup.com/category/MetaLibXServer
[21] http://www.loc.gov/standards/marcxml

Some data fields are specific to cataloguing manuscripts such as information about the writing material, etc. ONB's requirements with respect to these fields have already been incorporated into the current DM2E model.

The digitised representations of the provided manuscripts are stored in ONB's Digitool[22] system. The structures of the digitised manuscripts are encoded in the form of specific METS/XML files that have to be manually extracted from Digitool for each item. In a separate mapping step, information on page level is then mapped to the DM2E model and the result attached to the transformed metadata on item level.

The content on page level is available in form of Jpeg images. Links to the images hosted by ONB are provided in the *dm2e:hasAnnotatableVersionAt* fields of the DM2E model according to the DM2E Annotatable Content Specification[23].

4.9.2 Austrian Books Online

In contrast to the CODICES collection, the amount of provided items from ONB's ABO collection[24] is much larger and consists of about 50,000 items in various languages. The variety of items provided is very rich and spans books, atlases, illustrated books and periodicals such as scientific journals.

This variety is also reflected in the item metadata, which ranges from single, independent metadata records to comprehensive multivolume or collection hierarchies. Implications arising from such multilevel/hierarchical item-item relationships need to be considered when mapping the ABO metadata to DM2E. RAK-WB/MAB2 based cataloguing in general takes such hierarchies into account. It provides explicit interlinking fields for connecting related items, thus reducing data redundancy and allowing a rather straightforward translation of hierarchies into the DM2E model. While many of the ABO metadata records make use of these fields, some records represent hierarchies in a rather flat way, needing additional processing to be adequately translated to DM2E.

In addition, the metadata records feature a rich set of subject indexing terms from the controlled GND[25] vocabulary, providing a good means for contextualisation within the collection itself but also with the collections contributed by other content providers. About 10% of the person names within the records are also linked to the GND vocabulary, leaving much headroom for further contextualisation of the remaining person names by using the SILK framework.

4.9.3 Technical Details

Mapping process

The initial mappings from the MAB2 records to the DM2E model were performed with MINT. The visual editor offered a convenient introduction to the data model. Due to the nature of the source data format, however, the creation of XSL-transformations had to be continued by hand from a certain point on. The structure of the XSLT file generated by MINT served as a good framework for the manual extension of the code.

[22] http://archiv.onb.ac.at:1801/webclient/DeliveryManager?pid=2743137
[23] http://wiki.dm2e.eu/File:DM2E_annotatable_content_spec.pdf
[24] http://search.obvsg.at/primo_library/libweb/action/search.do?scp.scps=scope%3A(ONB_aleph_abo)
[25] http://www.dnb.de/DE/Standardisierung/GND/gnd_node.html

Hierarchical structure

For the CODICES collection, mapping consisted of the creation of one title CHO for each manuscript and children CHOs for each page of the manuscript. The resulting structure can be seen in Figure 3. It conforms to the Annotatable Content Specification[26] as agreed upon in the project. As noted previously, however, the metadata for the manuscript and for the pages were in two different source files. Since page CHOs needed information such as the title from their parent, the manuscript-level source file and the page-level source file were merged together into one compound source XML file for each manuscript.

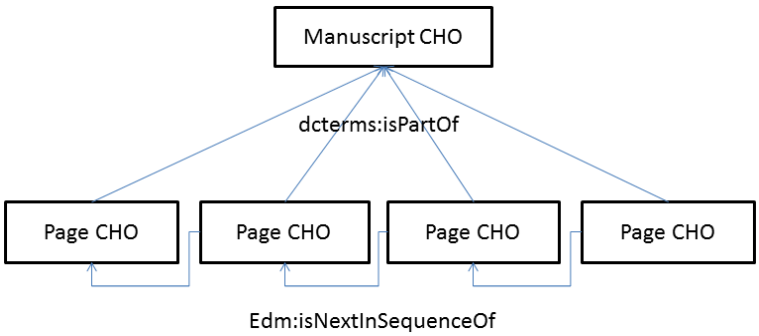

Figure 3: Hierarchical structure of the CODICES CHOs

Mappings of ABO items needed to take more complicated hierarchical structures into account. Many of the provided titles are multivolume titles that have appropriately been catalogued following the RAK-WB/MAB rules, thus being represented by different metadata records for the different hierarchical levels. Figure 4 shows the possible levels for such cases involving multiple physical items to be connected to an upper level series record. Such multivolume records themselves can again be part of a superordinate record.

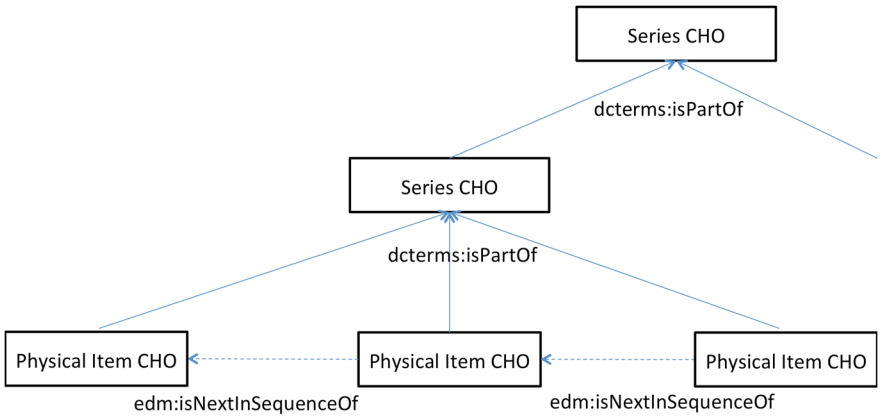

Figure 4: Hierarchical levels for ABO content

The order of volumes within a multivolume series could not always be resolved. As a consequence, physical item CHOs are only connected by edm:isNextInSequence if the order could be derived from the metadata records.

[26] http://wiki.dm2e.eu/File:DM2E_annotatable_content_spec.pdf

Links to External Authorities

As mentioned before, both CODICES and ABO metadata records contain rich links to the GND. While the CODICES feature detailed links to the person authority including person roles, the ABO titles feature a smaller proportion of linked persons but a high proportion of links to controlled index terms. The linked index terms are often distinguished by type such as person, place, genre etc. In order to preserve this information, specific instructions how to represent these relationships in the DM2E were included in the mapping recommendations.

4.10 NLI – National Library of Israel

NLI will provide records from its manuscripts, books and archives collections in Hebrew, Yiddish and Ladino. The bibliographic records are maintained in Aleph[27]. The digital objects are currently stored in Digitool, but currently being converted to Rosetta. The bibliographic metadata is catalogued in MARC21, with local fields as expansions, and can be exported as MARC XML.

NLI has added VIAF Id's to a large part of the personal names in the authority library[28] and will expand this data to the bibliographic records during export. This will allow creation of URIs for enrichment from VIAF in DM2E. This is especially important as the cataloguing language of the records that will be provided is Hebrew (with exception of subject headings, see below) and enrichment from VIAF will allow to retrieve the records by searching with English (Latin characters) personal names.

For the book collection[29], NLI uses LCSH subject headings with some adoptions for Judaica and hopes to publish these as RDF in the near future. The published data would also include translations to Hebrew. This would enable subject headings to be exported with IDs that can be used to create URI's as well. With subject headings as an only English element, this would make the records fully accessible in Hebrew. For the manuscript collection[30] there are local keyword style subjects in Hebrew, most of which have an English translation in a designated authority file. Exporting the records with the English translation should allow enrichment from an external source.

As MINT currently doesn't fully support MARC, NLI may have to use XSLT instead. WP2 included test records from NLI for the development of the DM2E model and there are no apparent mapping problems. For minor remarks, questions and requests please refer to the attached mapping table.

Currently, NLI objects do not meet the technical requirements for Pundit. However, after conversion to Rosetta this should be feasible. Export procedures from Rosetta are currently being developed.

[27] http://aleph.nli.org.il/F/?func=find-b-0&con_lng=eng
[28] http://aleph3.libnet.ac.il/F?func=find-b-0&local_base=nnl10&con_lng=eng
[29] http://web.nli.org.il/sites/NLI/English/collections/jewish-collection/Pages/rare-books.aspx
[30] http://web.nli.org.il/sites/NLI/English/collections/manuscripts/Pages/default.aspx

4.11 CJH via EAJC – Center for Jewish History

CJH will provide records for the digitised objects of the collections of the Leo Baeck Institute[31] and the YIVO Institute for Jewish Research[32], two member institutions which hold material on European Jewry.

CJH uses the same ILS and digital repository as NLI, namely Aleph and Digitool. The metadata is standard MARC21 and will be exported as MARC XML.

CJH will receive technical support from the EAJC and NLI. According to the sample records, CJH uses standard MARC21 without local fields. Mapping will be carried out upon receiving the full data sets, which may be expected in the near future. Due to the similarity with NLI, no problems are anticipated.

The provision of digital objects for Pundit needs further follow up.

4.12 UBFFM – University Library JCS Frankfurt am Main

The University Library Frankfurt am Main provides two collections including more than 1.000 digitised manuscripts: The collection 'Hebrew Manuscripts'[33] is already ingested in Europeana via the aggregator (provider) 'Judaica Europeana'. The new collection ('mediaeval manuscripts'[34]) includes more than 600 additional manuscripts for Europeana.

The digitisation projects at the University Library Frankfurt am Main are performed using the presentation and workflow software "Visual Library" (Semantics). The primary source of the metadata is the library system "Pica" (OCLC). The Pica records are mapped to MODS and embedded in a METS structure record according to the DFG guidelines. The combined METS/MODS records may be harvested on an OAI interface.

The METS/MODS format is an XML format that combines two standards of the Library of Congress[35]. Metadata Encoding & Transmission Standard (METS)[36] is the container that holds one or more metadata sections and sections for the logical and physical structure relating to the images that constitute the digital object (e.g. book). Also a link structure between different digital objects may be included in the logical structure section. The metadata sections hold the bibliographic information using Metadata Object Description Schema (MODS)[37]. There may be also links between digital objects in the MODS section using the relatedItem[@type="host"] element.

A complete mapping to the DM2E model satisfying the requirements of the Deliverable 1.1 has been developed. Due to the complex structure of the METS part, a mapping using a tool such as MINT is not possible. Algorithms are necessary to address the proper elements of the record and to distinguish different cases. Furthermore the METS/MODS metadata encoding is quite flexible and often contains partly redundant information.

The implementation of the mapping has been carried out in two steps using a XSLT sheet 'METS2DM2Ev9.xsl': In the first step, the triples are generated following the structure of

[31] http://access.cjh.org/home.php?type=repository&term=TGVv#1
[32] http://access.cjh.org/home.php?type=repository&term=eWI2bw==#1
[33] http://sammlungen.ub.uni-frankfurt.de/mshebr
[34] http://sammlungen.ub.uni-frankfurt.de/msma
[35] http://www.loc.gov/
[36] http://www.loc.gov/standards/mets/mets-schemadocs.html
[37] http://www.loc.gov/standards/mods/mods-schemas.html

the source record. In the second step, the merging and deduplicating of properties and resources is done in order to simplify further processing including tests and validation.

The mapping includes two levels of CHOs. On the title level, the corresponding MODS fields are mapped to edm:ProvidedCHO and ore:Aggregation. On this level edm:isShownAt is provided using the URN and a resolver URL:

```
<edm:isShownAt rdf:resource="http://nbn-
    resolving.de/urn:nbn:de:hebis:30:2-11547"/>
```

The available MODS fields mods:identifier, mods:recordInfo, mods:titleInfo, mods:name, mods:language, mods:originInfo, mods:physicalDescription, mods:subject are mapped to the corresponding properties dc:identifier, dcterms:title, pro:author, dc:language, dcterms:issued, bibo:editor, dm2e:publishedAt, dc:format, dc:description, dc:subject of ProvidedCHO.

Values with GND resources in the MODS part

```
<mods:name type="personal" authority="gnd" authorityURI=http://d-
    nb.info/gnd/ valueURI="http://d-nb.info/gnd/120313170">
<mods:namePart>Lauber, Diebold</mods:namePart>
<mods:namePart type="given">Diebold</mods:namePart>
<mods:namePart type="family">Lauber</mods:namePart>
<mods:role>
<mods:roleTerm type="code" authority="marcrelator">scr</mods:roleTerm>
<mods:roleTerm type="text">Schreiber</mods:roleTerm>
</mods:role>
</mods:name>
```

are mapped as resources – dm2e:copyist in this case – in the DM2E model:

```
<dm2e:copyist rdf:resource="http://data.dm2e.eu/data/agent/ub-
ffm/authority_gnd/120313170"/>
<foaf:Person rdf:about="http://data.dm2e.eu/data/agent/ub-
    ffm/authority_gnd/120313170">
<owl:sameAs rdf:resource="http://d-nb.info/gnd/120313170"/>
<skos:prefLabel>Lauber, Diebold</skos:prefLabel>
</foaf:Person>
```

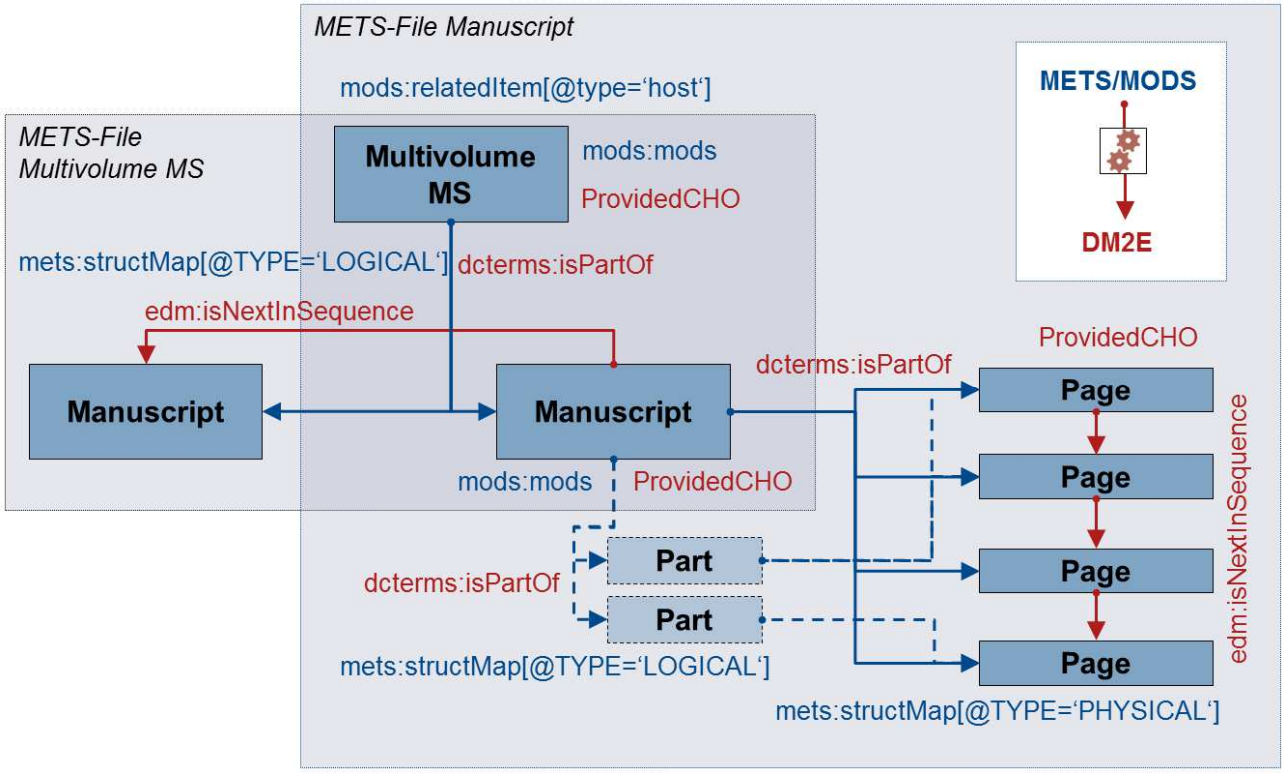

Figure 5: METS/MODS mapping to the DM2E model

The page level CHOs are based on the METS information. On this level, edm:isShownBy and dm2e:hasAnnotatableVersionAt are provided pointing directly to the page JPG. For the single pages there is no MODS section available, but a page label, which is based on intellectual recording, can be used for dcterms:title. Figure 5 shows the various hierarchical levels present in the METS/MODS data. The representation of hierarchical structures – pages and works in multivolume cases – is accomplished following the "Recommendations for the representation of hierarchical objects in Europeana". The METS and mods:relatedItem informations are used for the modelling of the hierarchical structure on title level.

Further development will include the mapping of the logical structure information (parts, chapters etc.) in the METS section.

4.13 GEI – Georg Eckert Institute for Textbook Research

The Georg Eckert Institute for Textbook research contributes metadata and digital objects from its digital collection GEI-Digital. The collection contains 400,000 scanned images and full text in METS/MODS.

Since the metadata are available as METS/MODS, it was possible to adapt UBFFM's transformation workflow for the GEI data. Thus, the provided content could already be integrated successfully.

4.14 MPIWG – Max Planck Institute for the History of Science

The Max Planck Institute for the History of Science (MPIWG) provides three different sources to the DM2E project: the institute's rare book collection, the manuscripts of Thomas Harriot and data from the Islamic Scientific Manuscripts Initiative (ISMI).

4.14.1 Rare Book Collection

The rare book collection consists of 1266 books. These books are, for the most part, scientific writings of the early modern period and are used as electronic sources of reference by the scholars of the institute. For the institute's web presentation on http://echo.mpiwg-berlin.mpg.de, the metadata for each CHO is stored on a server, along with the digital facsimiles. This metadata is an export from the library's Filemaker database into the MPIWG's proprietary metadata format (XML files, called index.meta. Documentation: http://intern.mpiwg-berlin.mpg.de/digitalhumanities/mpiwg-metadata-documentation/formate/indexmeta-standard). However, these data contain no information on the digital facsimiles other than rights information and their location. Thus, information about the amount of images per CHO has to be obtained via a web service. It is important to include information about the digital facsimiles in the RDF export, so that links can be generated to each page of the CHO to make it annotatable.

This two-fold approach makes it difficult to use existing solutions like MINT. First attempts were made with MINT, dealing only with the index.meta part, and yielded good results. Some fields were difficult to map which would have required a row of custom XSL functions, applied either via MINT itself, or by editing the mapping via an XSL export locally.

Finally, the decision was made to develop a tool, accessible by Web Service or command line, that reads the input files and directly writes DM2E model RDF data ready for ingestion.

4.14.2 The Manuscripts of Thomas Harriot

With the advent of 2 additional volumes from a digitising campaign this spring, the number of manuscripts volumes containing the works of Thomas Harriot has grown to 10 (for more information, see http://echo.mpiwg-berlin.mpg.de/content/scientific_revolution/harriot). The number of pages is about 10000. The situation here is similar to the one described above, as the resources are stored on the MPIWG's servers in the same way as the rare book collection.

Additionally, the varied content of the manuscripts is being classified by topics. For this purpose, a set of hierarchical maps, grouped by topics, has been created by scholars and can as well be exported to RDF. The maps can be browsed starting at the URL stated above. The relations described in this RDF display information about e.g. original pagination, the usage of mathematical results from other pages and conjectural relations. Also, quotations and references to other persons and works have been added and would be a nice addition to add this via the appropriate properties in the DM2E model.

4.14.3 ISMI

As of December 2013, the ISMI collection is composed of over 13,000 CHOs, whose metadata can be visualised in https://openmind-ismi-dev.mpiwg-berlin.mpg.de/om4-ismi. From the ISMI database, object of the two classes: witness and codex are exported as CHO to DM2E.

The ISMI database is stored by the Openmind framework. This framework is also responsible for the mapping to DM2E.

Although Openmind is currently able to write DM2E data ready for ingestion without help of the MINT platform, the WebResources that are generated from ISMI have not yet permanent

URLs. For this reason, the ingestion cannot actually be executed. These WebResources with permanent URLs will be available in late March 2014.

4.15 CRNS – National Center for Scientific Research

The metadata from the *Digitale Faksimile Gesamtausgabe* will be mapped to DM2E by ONB. According to the information in Deliverable 1.1, the metadata should be available in form of Europeana Semantic Elements (ESE). Currently, however, only Dublin Core (DC) formatted metadata is available via the Nietzschesource OAI server.

NET7 was able to provide the DC records enriched with links to the digitised objects, thus meeting the requirements for ingestion into Europeana.

4.16 KB – National Library of the Netherlands

The National Library of the Netherlands will provide access to digitised illuminations from their Medieval Illuminated Manuscripts collection to DM2E. The data consists of metadata for the 400 digitised manuscripts and metadata for 11,141 illuminations in these 400 manuscripts.

The content- and the metadata questionnaires have been filled out and returned to ONB. KB's data seems well suited for a mapping with MINT and ONB will support KB with their first steps in MINT.

5 Task 1.3: Testing the User Interface for Creating Mapping, Interlinking Heuristics and for Configuring the Workflow

5.1 Results of the MINT Questionnaire

This Section presents the results of the evaluation of the returned MINT questionnaires as of July 16th 2013. As of this date, five questionnaires have been filled out, therefore these results should be regarded as preliminary, although it is expected that they already cover most aspects relevant for DM2E.

The questionnaire can be found in Appendix A. It has been designed by ONB and MPIWG and seeks to provide users with a general guideline on how to proceed with their mapping and to assess their experience with MINT's mapping functionality according to four major steps:

- Data import
- Data mapping
- Validation of the created mapping
- Export of the resulting XSLT transformation

In general, MINT's visual mapping functionality was accepted by the users. This concept is very intuitive and helps users become familiar with the target data schema. It is, however, limited to source XML schemas that do not rely on repetitive xml elements distinguished only by attribute values or on schemas that require a lot of conditional mappings. For example, MARCXML based formats with their repetitive <varfield>-<subfield> structure are therefore not very suited to be mapped with MINT. NTUA offers a separate transformation that can be used to transform MARCXML data into a format suitable for a mapping with MINT.

For complex mappings in general, it might be easier to write one's own XSLT code, although MINT can serve as a valuable starting point providing an initial, automatically generated code to be refined by hand.

The results for each of the four steps are presented in the following subsections.

5.1.1 Import

Users had to upload an XML file with their source metadata. They had to define the structure of the file by using the provided MINT functions for defining elements separating items in the XML file. They were asked if they were able to fulfil this task and what ways the documentation and the usability of the user interface supported them in doing so.

In general it can be said that the provided drag & drop functionality for defining items within the XML files works well. The supporting documentation is sufficient, although there were some problems in understanding the terminology regarding the required fields "item level", "item ID" and "item label". A fundamental problem arises if the structural XML elements for separating items and defining their respective IDs are based on conditional values such as special attributes. As there are no means provided for defining conditional mappings during this step, it is not possible to fulfil this task in such cases. One solution would be to allow providers to specify custom XPath expressions at this point.

In addition, the user interface is very helpful in supporting users when defining the required elements. However, some browsers seem to have problems with correctly displaying all the field labels within the provided source element tree, which cannot be resised or moved. It would be beneficial if there was a possibility to interrupt the import process if the amount of imported data accidentally exceeds a reasonable size.

5.1.2 Create Mapping

Users had to create a new mapping of the source fields of their uploaded XML file to the DM2E target schema by using the provided drag & drop functionality. They were asked to distinguish between problems arising from lack of functionality, usability issues or problems with the DM2E target schema. Moreover, they were asked if they had used any of the advanced mapping functionalities and navigation aids offered and if the documentation supported them with their task.

For the mapping, the drag & drop functionality proves to be very effective and the provided navigation aids such as bookmarks and the source/target element search are very much appreciated. In general it can be said that for some XML data schemas the provided functionality is sufficient to create mappings to the DM2E model. As stated in the beginning, however, it works well with source XML schemas that have distinguishable element names while it is less suited for XML schemas relying on repetitive elements distinguished by attribute values. Once mappings are dependent on such conditional constraints, the mapping process becomes significantly more difficult. This is currently the case with MARCXML but also with METS/MODS based XML schemas.

MINT provides very useful advanced mapping functionalities such as functions, conditional mappings and value mappings and these are used extensively within DM2E. In general, the feedback on these features was very good, although there were issues with concatenation (Not possible to use spaces in concatenation), with custom functions (No documentation) and with conditional mappings. The latter seem to be not very intuitive (visual layout) and cannot cover all cases encountered in the DM2E mappings. For mappings to target attributes such as @rdf:about, the conditional mapping functionality does not seem to work. Moreover, with MARCXML it would for example be necessary to test for a certain value of one subfield and then do a conditional mapping from a sibling subfield. Such combinations of conditional mappings would be possible if users were able to enter custom XPath expressions instead of having to rely on the visual configuration of such constraints. Therefore, one important requirement is to be able to enter XPath expressions by hand if necessary.

The nested structure of the target schema helps users that are unfamiliar with the DM2E model, giving them a good understanding of how the different DM2E classes are related to each other. This comes with the disadvantage that sometimes certain class instances have to be declared over and over. For example, although one and the same edm:Agent can be linked to both dm2e:dataProvider and dm2e:rightsHolder, it has to be mapped twice. Similarly, URIs are sometimes constructed through concatenation of different source fields and are used in various places of the mapping, but can only be distinguished using different prefixes (Such as URIs for ore:Aggregation and edm:ProvidedCHO) . Therefore, another great improvement to MINTs functionality would be if users were able to copy already mapped items to other target elements.

Users were concerned that, in some browsers, the "Backspace" key keeps its standard action which is "navigating back". If a user presses the Backspace key in MINT, the browser steps back to the login page, making the user believe that he or she has been kicked out of MINT

and that all the mapping is lost. Therefore it would be helpful to catch the activation of the Backspace key and to supress its standard action.

The red dot with an x marking mandatory fields that have not been mapped yet has a visual appearance that suggest a close/delete functionality. It might be an improvement to choose a different symbol for unmapped mandatory fields.

MINT sometimes has noticeable lags and seems to stop taking instructions. Loading times are sometimes quite high. One significant problem is that if two MINT sessions are simultaneously opened in the same browser, there are random interactions between the sessions and mappings get corrupted. Since it is sometimes desirable to open two mappings simultaneously, this should be fixed.

5.1.3 Validation

Users had to use the "preview" functionality of MINT in order to perform a test validation. They were asked to look at the validation result and to iteratively resolve validation errors.

The validation functionality of MINT is a valuable tool within the whole mapping process. It provides visual feedback about mapping errors with respect to the target schema. The provided documentation is helpful, the process very intuitive and the interface easy to use. From the mapping view it is, however, not very intuitive to find the validation interface. Users have to click "Preview" and are then directed to a separate preview pane that is initialised for viewing their source xml document. In order to see the validation results, users first have to choose the "Mapped item" view. Since the validation component is a very important feature, it should be placed more prominently and be easier to find.

In the validation view, the "Show/Hide Report" link is not clearly identified as a clickable button, while the red squares with an x have the appearance of being buttons, but are not. Moreover, it does not seem to be possible to scroll to the left or to the right.

Of the returned errors, one error message "cvc-complex-type" was not understood. In general, working with MINT and its validation also raised questions about the DM2E target schema itself. Some target properties that are "alternatively mandatory" (such as edm:isShownBy/edm:isShownAt or dc:description/dc:title) trigger errors although one of the options is correctly mapped.

Some users find it annoying to have to choose language tags so often, as it is very often the case that it cannot be clearly stated what language a string has. Therefore it could be helpful to drop this requirement or to initialise with "unknown language" tags. Rules for creating identifiers are sometimes not followed, such as the initialised concatenation of the prefix for ore:Aggregations, obviously not following the proposed "dm2e:aggregation/[provider]/[collection]/[identifier]" pattern.

Values for "xsd:dateTime" require specially formatted strings of the form "YYYY-MM-DDThh:mm:ss". Users ask for a best practice approach for filling in the values if only years are stated in the source metadata. For example, should a stated year "1176" be formatted as "1176-01-01T00:00:00"? Another request for formatting guidelines arises from the question how URIs should be created for string resources. Should a string "John the Baptist" result in a URI "dm2e:item/[provider]/[collection]/John_the_Baptist" or should they be formatted differently? What about preserving uppercase characters?

5.1.4 Generated XSLT

Users were asked to export the generated mapping as XSLT and to have a look at the generated code. They were asked to note anything that they might have done differently.

The generated code is readable and functional and provides a good base for manual additions. In fact, some content providers generate initial XSL transformations with MINT and then continue editing the code by hand.

User comments suggest the need for more variables, especially for repeated mappings such as the edm:Agent example mentioned in Section 5.1.2. Another comment suggested leaving out the "position()=1", constraints in the generated code.

5.1.5 Conclusion

The results suggest that for metadata schemas with distinguishable XML elements such as TEI, DC(X) or ESE, mappings to DM2E can already be created with the current MINT version. With other formats such as MAB, METS and MARCXML, the use MINT for mapping depends on the functionality of future versions. Such mappings would, however, require advanced knowledge from the user with respect to XPath expressions etc.

In cases of complex mappings, MINT can still be used for the creation of initial mappings that can subsequently be extended by hand.

For formats such as MARC or MAB, NTUA can provide a separate transformation that can be used in a pre-processing step to expand MARCXML data to a format suitable for MINT. This procedure turns the repetitive <varfield> <subfield> element structure of these formats into unique XML elements by creating their names from the attribute values in the respective <varfield> and <subfield> elements. This makes it possible to use MINT for MARCXML data, but its success depends on the complexity of the source metadata. In cases where the full range of MARC/MAB fields and indicators are used, such an expansion results in a very large number of unique XML elements that each have to be mapped separately. In such cases, manual XSLT creation is currently still the preferred choice.

In general, the user feedback collected through the questionnaires has resulted in suggestions for improvements of MINT, both with respect to its functionality and its usability. Moreover, they have revealed some open questions with respect to the DM2E model such as best practices to format dateTime strings and instance URIs, addressed in the document "Recommendations for the RDF representation of DM2E metadata"

5.2 Results of the Evaluation of OmNom

As of now, most data ingestions in DM2E have been performed using a specialised client ("Direct ingestion") that is restricted to file- and OAI-PMH-based ingestions using an XSLT transformation. While this fits to the workflows of the core data providers, it prevents users from creating their own workflows and from using the high flexibility provided by the underlying workflow engine. OmNom, the Web frontend to the Web-based workflow engine developed in WP2, provides this flexibility. For evaluation purposes, we asked the providers to create their XSLT-based ingestion workflows in OmNom. Like with the direct ingestion content providers use this tool to upload their legacy metadata, to transform them to a new target data model – the DM2E model in the context of this project – and to publish the results as linked data on the Web. The transformations in form of XSLT files can either be also uploaded by the user or directly transferred from the MINT mapping tool discussed in

Section 5. The concrete data transformation and publication steps are represented in the form of workflows. A workflow consists of a specified but generic sequence of transformation/publishing operations on or with the data. OmNom's user interface allows the users to create such workflows based on their individual needs and to instance them in form of so-called configurations, assigning explicit input data (file names, descriptions, etc.) to a workflow. Once such a configuration has been created, it can be started as a job. Users can monitor the progress of the job and get feedback on success or failure. Successfully transformed and published data can be previewed in the integrated linked data presentation platform Pubby.

One part of Task 1.3 was to evaluate the functionality of this tool for content providers to create workflows based on their source metadata files and the mappings created for their transformation to the DM2E model. Being a tool developed from scratch in DM2E, some content providers were able to participate early in the development process in order to test the User Interface in various stages while constantly being in touch with the developers. This process proved to be very effective with respect to the early detection of bugs and to the incorporation of additional UI features. Figure 6 shows a short snippet of this process as it is represented on OmNom's Github[38] site.

Once the UI had reached a state to be evaluated by the providers, a detailed walkthrough[39] was created in order to guide users through the necessary steps to create workflows for their data to be published in the DM2E Triple-store. This step-by-step tutorial also served as a guideline for the evaluation of the tool by the content providers. Based on the steps featured in the walkthrough, a questionnaire was created in order to record user experiences with the tool involving six tasks:

- User registration
- Upload & manage files
- Create workflows
- Create configurations
- Start and monitor transformation jobs
- Preview published results in Pubby

[38] https://github.com/DM2E/dm2e-gui/issues?page=2&state=closed
[39] https://github.com/DM2E/dm2e-gui/raw/master/doc/OmNom_WalkThrough_v2.pdf

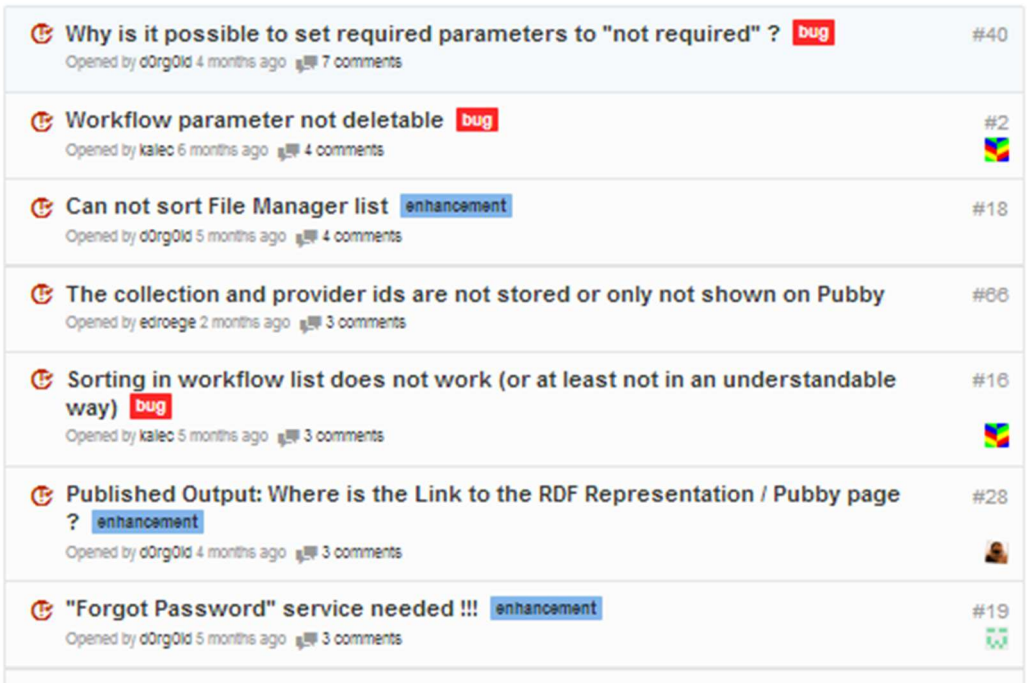

Figure 6: OmNom Issue list on GitHub.

Content providers were introduced to the most recent version of OmNom during a training session at the All-WP meeting in Athens on November 28th-29th 2013. After the introduction they were asked to perform the required steps to publish their data and to record their experiences by filling out the questionnaire and to send it back after the meeting. Up to date, four questionnaires have been received and form the basis for the results presented in this Section.

5.2.1 User Registration

In this task, users had to create a new user for the system and to log in to OmNom. All users successfully managed to create a user and to log into OmNom using the generated credentials. One user noted that she did not receive a new password via email after using the "forgot password" service.

5.2.2 Upload & Manage Files

This task was dedicated to uploading legacy metadata files and respective XSL-transformations and to assess the functionality of the file management screen.

File Upload

All users were successfully able to upload their legacy metadata files and their respective XSL-transformations. One user mentioned that he was not immediately able to understand the functionality of the existing textbox for applying labels to uploaded files but stated the importance of this feature in later stages of the ingestion process.

File Management

The ability to filter the displayed files to only those uploaded by user worked well for most of the users and they suggested making this the default setting. The display of internal IDs assigned to uploaded files was considered as confusing and stressed the importance to make self-assigned file labels mandatory during the upload process. For the publishing of data already present as RDF, users also suggested to include the ability to upload N-Triples serialisations.

5.2.3 Create workflows

In this task, users had to create workflows defining the necessary steps for data transformation and publication. Users were asked to report on the usability of the visual configuration editor. In general, users were able to understand the drag & drop based configuration of ingestion workflows and described the process to be straightforward. All users were able to create their intended workflow. Some usability issues, however, were raised and the following paragraphs discuss the users' remarks in detail.

Drag & Drop

One user mentioned that it is very uncommon to drag & drop text labels instead of icons. The functionality of highlighting relevant GUI elements on mouse-over in the "Connectors" section was not immediately understood.

Visual Layout & Appearance

The visual layout of the different areas was sometimes not immediately understood. One user mentioned the visual order of the functional areas of the GUI to be disturbing, being oriented from right to left. Users have to start dragging elements from the rightmost column to the center of the screen and from there to the input/output parameters on the left. This is confusing for users used to left-to-right reading settings. The list form of displaying the connectors between input/output parameters and operators and between different operators within the workflow is not very intuitive. This experience is also reinforced by some unclear labelling issues such as using the same labels for input parameters and input arguments. Moreover, the meanings of some of the labels were not immediately clear to the users. Especially abbreviations such as "WF" for workflow or "Pos#1" for the first operation in the workflow were mentioned in this regard and some labels were considered being too technical for general users. The function to visualise the "wiring" of the different workflow elements, however, was of great help to clarify many of these issues.

General Functionality

It was not immediately clear to some users that some the input parameters to the different operators in the workflows represent files such as XML and XSLT files for the transformation operator and that other input parameters represent "simple" descriptive strings. Another remark was raised that the general idea of creating generic workflows interfered with the possibility to assign default values for input parameters within the workflow configuration step. Last but not least it was mentioned that context based help such as info buttons or mouse-over balloons would be of great support.

5.2.4 Create Configurations

After the creation of the workflow, users had to create a specific instance of the workflow for a set of XML and XSLT input files. All users were able to create specific configurations of the previously created workflows. The following paragraphs report about the details encountered by the users.

Relationship between workflows and configurations

The conceptual difference between workflows and their specific configurations was understood by the users. As pointed out in the previous section, the ability to define default values for generic workflows slightly interfered with this concept.

Drag & Drop

The drag & drop functionality of the configuration editor to drag files from the file list to the respective input fields worked well for all of the users. As for the workflow editor, the fact that text labels have to be dragged to their respective target fields was considered as being a bit unusual. On user suggested to highlight valid target fields based on the type of file being dragged (XML file dragged -> XML input field highlighted). Some users found it difficult to distinguish between target field accepting files and those accepting strings. One user suggested separating such fields into two visually distinct sections.

General functionality

The ability to filter and sort input files by type and owner was considered as being very helpful. As for the workflow editor, users suggested to include context sensitive help.

5.2.5 Start and Monitor Transformation Jobs

In this task, users had to run their created configurations as "jobs". They were asked to monitor the job-progress and to report if they finished successfully. All users were able to start the jobs from the configuration editor, but they reported problems on tracking job progress and some encountered problems with successfully finishing ingestion.

Starting Jobs & Monitoring Job Progress

Starting jobs from the configuration editor was reported as being straightforward, one user mentioned to use the label "run" instead of "execute" in the configuration editor.

Users reported problems with tracking job progress. One issue was raised regarding the need to constantly push the "refresh" button in order to update the display of the job's state. Moreover, users reported that they sometimes did not get immediate feedback after refreshing the view. This happens because the job progress is updated in discrete steps based on the different positions in the workflow - feedback is only returned to the user after one position has finished. With larger datasets, this can lead to the impression that the whole process got stuck and users are left uninformed about the actual state of the process.

Managing Jobs in the Job List

Users reported some problems with finding their jobs in the job list. They were not able to filter jobs by user name and all jobs were displayed with the state "not_started", even if they are currently running or have already finished.

General Functionality

Starting, monitoring and successfully finishing ingestion jobs currently only works for small datasets. Larger datasets, especially those containing multiple source data files, currently seem to take very long to ingest and leave users with the impression that the ingestion has failed. It is unclear whether the reason for this lies in the User Interface or in the processing backend.

5.2.6 Preview Published Results in Pubby

After a successful ingestion users can open the integrated Linked Data preview by clicking the "Show in Pubby" button. Users were asked to explore the ingested dataset and to report about the functionality of the preview.

In general, Pubby proved to be a very practical tool for viewing the created linked data sets. It gives a very good overview on what has been ingested and allows users to browse through the generated linked data. Many user requests for additional preview features were immediately addressed by adjusting the underlying query templates.

One user mentioned that the Pubby preview was perceived as being of very technical nature and that the complete display of all triples sometimes resulted in very long lists. At the same time, however, he also acknowledged this to be the "raw view" of the triples and to fulfil its purpose in this regard.

5.2.7 Conclusions

In general it can be said that the GUI for the ingestion platform fulfills its purpose quite well. Guided by the step-by-step tutorial, users quickly adopted to the visual editing paradigm of the workflow and configuration editors. Some usability issues were encountered and leave room for future improvements. Especially the job monitoring section and the job list view need some revision in this regard.

During the time of this evaluation, problems have been encountered with large scale data ingestions using multiple input files, despite the fact that such iterative workflows should be supported by OmNom. Bug reports have been filed and will be addressed by the developers.

Once these issues will have been resolved, the results of this evaluation suggest that OmNom will be able to fulfil all the requirements for providers to successfully ingest their data.

5.3 SILK

The Silk workbench is a web application to configure and run contextualisation tasks for the enrichment of data. In DM2E it is used for enriching the data providers' content with links to external authority repositories for the enhancement of the Web of Linked Data.

The basic functionality is as follows. The data of two repositories (in this case, one is the DM2E data, the other is an external repository with authority data) is compared to each other field-wise (e.g. the URI of a place-name) and if the agreement between the two sources is above a certain level (based on measures like the similarity of the two strings or proximity in terms of geo-coordinates), a link is created and the item in question is contextualised. In the DM2E context, this is used for the de-duplication of ingested data and for enrichment. The two scenarios are described in detail in D2.3 ("Final version of the interoperability infrastructure").

There are essentially two ways configuring the tool: one is by using the graphical editor that forms the integral part of the Silk Workbench (Figure 7). Here, the data sources and the measuring and comparing parameters are displayed as boxes, which can be visually pieced together by connecting them with lines, creating an easily understandable way of how the data will be processed. For beginners, this is a good way of getting to know the capabilities of the system.

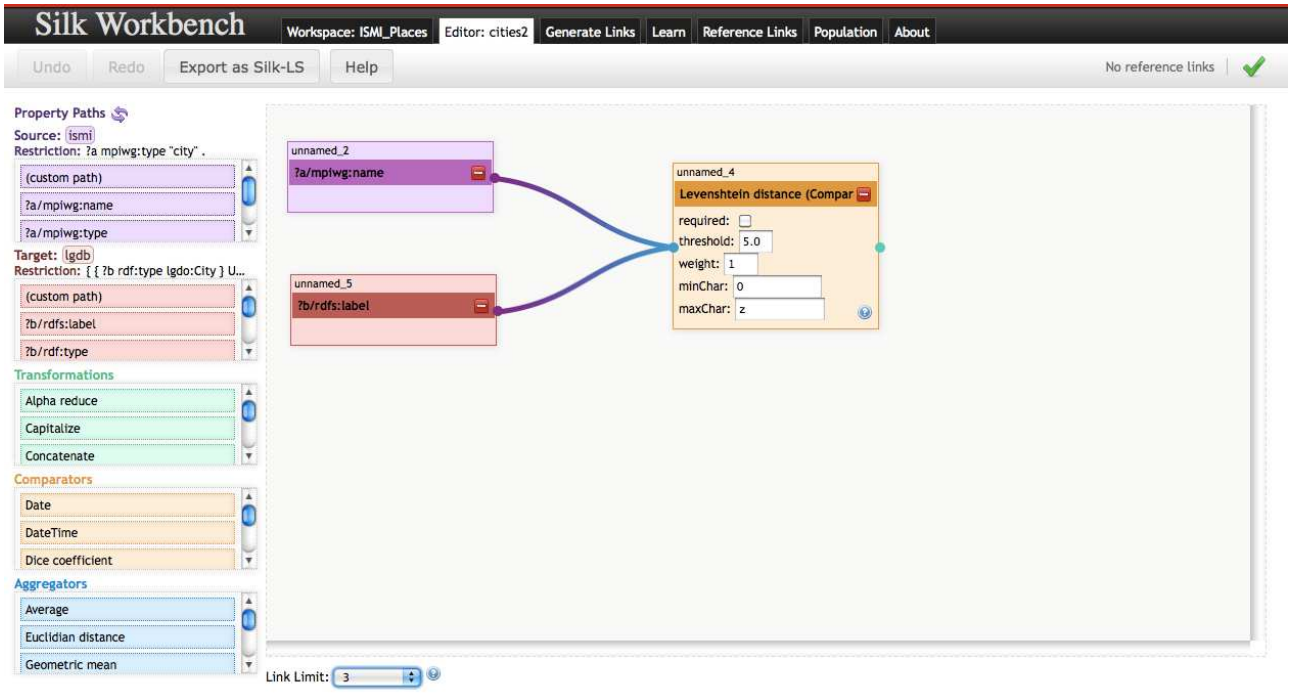

Figure 7: The SILK Workbench

The other method is by writing a configuration file in XML using the Link Specification Language[40] (Figure 8). This requires a shell access to the machine, as the linking process is then started using the command line. These files can also be imported into the graphical editor and be further refined there.

[40] https://www.assembla.com/spaces/silk/wiki/Link_Specification_Language

```
    <DataSource type="file" id="ismi">
      <Param name="file" value="ismi_metadata_100_texts.rdf"></Param>
      <Param name="format" value="RDF/XML"></Param>
    </DataSource>
    <DataSource type="sparqlEndpoint" id="lgdb">
      <Param name="pageSize" value="1000"></Param>
      <Param name="pauseTime" value="0"></Param>
      <Param name="retryCount" value="3"></Param>
      <Param name="endpointURI" value="http://linkedgeodata.org/sparql"></Param>
      <Param name="retryPause" value="1000"></Param>
      <Param name="graph" value=""></Param>
      <Param name="login" value=""></Param>
      <Param name="entityList" value=""></Param>
      <Param name="password" value=""></Param>
      <Param name="additionalParameters" value=""></Param>
    </DataSource>
  </DataSources>
  <Interlinks>
    <Interlink id="cities">
      <LinkType>owl:sameAs</LinkType>
      <SourceDataset dataSource="ismi" var="a">
        <RestrictTo> ?a &lt;http://www.mpiwg.de/ismi/type&gt; &quot;city&quot; . </RestrictTo>
      </SourceDataset>
      <TargetDataset dataSource="lgdb" var="b">
        <RestrictTo>
          { { ?b &lt;http://www.w3.org/1999/02/22-rdf-syntax-ns#type&gt; &lt;http://linkedgeodata.org/ontology/City&gt; } UNION { ?b
&lt;http://www.w3.org/1999/02/22-rdf-syntax-ns#type&gt; &lt;http://linkedgeodata.org/ontology/Town&gt; }UNION { ?b &lt;http://
www.w3.org/1999/02/22-rdf-syntax-ns#type&gt; &lt;http://linkedgeodata.org/ontology/Village&gt; } } .
        </RestrictTo>
      </TargetDataset>
      <LinkageRule>
        <Compare weight="1" threshold="2.0" required="false" metric="levenshteinDistance" id="unnamed_4">
          <Input path="?b/rdfs:label" id="unnamed_5"></Input>
          <Input path="?a/mpiwg:name" id="unnamed_2"></Input>
          <Param name="minChar" value="0"></Param>
          <Param name="maxChar" value="z"></Param>
        </Compare>
      </LinkageRule>
      <Filter></Filter>
      <Outputs> </Outputs>
    </Interlink>
  </Interlinks>
</Silk>
```

Figure 8: SILK Link Specification Language

A drawback of the first method is the lack of logging. Especially when dealing with huge datasets, in the first stages, there is no way of telling if or what is going on. The command line output is very verbose in that respect. Furthermore, if the linking process is expected to run for a long time, this can be safely run on the server, while the browser-based approach seems not so stable.

In a first experiment, one geo-location property of a small set of 100 RDF triples was matched against the Linked Geo-data (http://linkedgeodata.org/About) with a simple Levenshtein comparator set between them. After starting the task, the webpage shows no change for about 10 seconds, after which the linking results are displayed (Figure 9). The generated links can be controlled by immediately visiting the resources and their correctness can be set which serves as an input for the learning task.

It would also be possible to return then to the linking editor to adjust some settings for better results. The result can then be written to file or ingested into a triple-store via a SPARQL update – or they can merely be shown, which is good for learning or experimenting with settings.

Keeping the drawbacks mentioned in mind, interaction with the workbench can be described as successful, even for a novice user.

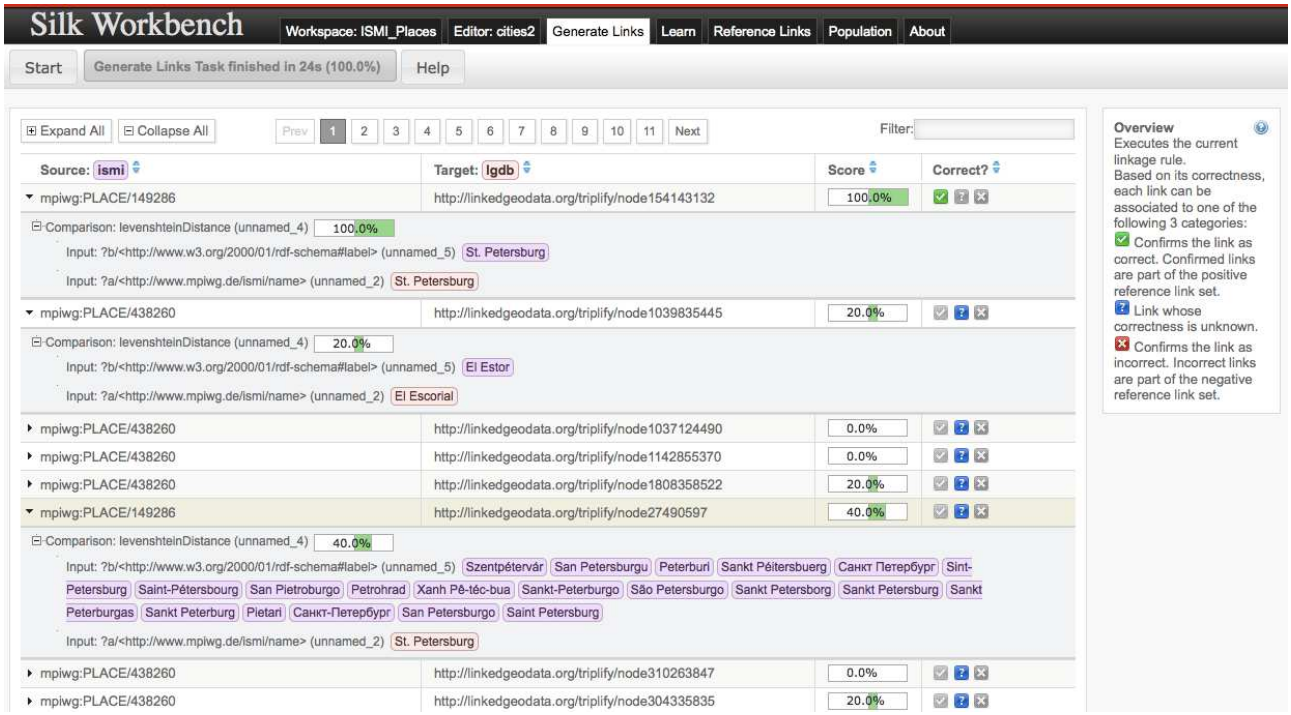

Figure 9: Results of SILK linking process

6 Task 1.4: Setup a Test Scenario for the Prototype Platform

6.1 Overview of Task 1.4: Setup a Test Scenario for the Prototype Platform (ONB)

The following Section provides an overview and the results of the test scenario carried out by the ONB (WP1) in close cooperation with members of WP3 (NET7, UIB, UBER) and to a further extent with OKFN (WP4). The purpose of the test scenario was to assess the prototype platform in its ability to facilitate the creation of semantic enrichments of digital material and extraction of rich features from digital objects by digital humanities scholars. This test scenario concentrated less on evaluating the annotations that were made, and instead focused in the scholars' opinions of the platform.

In order to measure the usability and usefulness of the prototype platform for digital humanities scholars, two case studies were conducted. In the first, complex retrieval experiments enabling the extraction of rich features from digital objects using the DM2E prototype platform of WP3 (Pundit and http://ask.thepund.it) were carried out by digital humanities scholars in the field of Wittgenstein scholarship, organised by UIB partner Wittgenstein Archives at the University of Bergen (WAB) - the "Wittgenstein Pilot Group" or "Wittgenstein Incubator". Questionnaires and a discussion with the scholars followed these experiments and served as the basis of assessment for the platform's ability to facilitate scholarly activities in general and the creation and reuse of scientifically meaningful annotations in particular. In the second use case scenario, a second group of digital humanities scholars already using Pundit for the purpose of linking entities in a digital edition on the semantic web (Burckhardtsource) were given a survey with questions geared at ascertaining their perceptions of the platform.

The evaluation of the prototype platform in the context of the Wittgenstein Pilot showed that scholars not intimately familiar with the new way of working facilitated by the prototype platform may have some initial difficulties using the platform at first. As a whole, however, this group could generally see the possibilities of using the platform in their own digital humanities research. In addition, this use case scenario strongly shows that the prototype platform can facilitate the creation of digital humanities solutions regarding the linked open data paradigm. The Burckhardtsource use case revealed that the prototype platform is being successfully used by digital humanities scholars to create rich document descriptions, which can be reused by other scholars. These scholars showed a predominantly positive attitude towards the functionality of the platform as well.

6.2 Current State of the Prototype Platform

Over the course of the project, the prototype platform has come to stand for a collection of tools with diverse functionalities developed to aid the content providing institutions of DM2E in storing, visualising, aggregating, augmenting and publishing digital humanities related (meta)data on the semantic web. Although many elements of the prototype platform were already being developed by NET7 before the start of the project, they have been extensively expanded upon during the first two years of DM2E (see D3.2). In its current form, the prototype platform consists of annotatable content from UIB and ECORR and the following tools developed by NET7: Muruca Digital Library (a tool that enables curators of (digital) materials to present "collections" online), Pundit (a web application that lets curators and researchers annotate digital materials in various ways including, but not limited to, identifying named entities using text mining and drawing outlines of shapes or objects in

digital images), http://ask.thepund.it (lets users search their own notebooks and the published notebooks others have created), Feed.ThePundit (lets the user add URIs directly to the Pundit so that they can be added to the Linked Open Data Cloud), and Korbo (lets users create "baskets" of data taken from existing data providers so that they can browse annotate and augment them). Since the first batch of content is now integrated into the DM2E WP2 Platform (Interoperability Infrastructure), it will also be accessible and annotatable using tools developed in the context of the WP3 prototype platform.

With Pundit, users are able to create semantic annotations (RDF-triples) out of content published on the web. These annotations are then aggregated into public or private notebooks and stored on the Pundit Annotation Server. The notebooks are then accessible through http://ask.thepund.it. A feature of Pundit's triples generator is the use of existing ontologies and vocabularies such as those from DBPedia and Freebase to create meaningful annotations. In theory, users can also create their own ontologies and upload them to Pundit. This was done by the Wittgenstein Pilot Group (UIB) with their WAB Wittgenstein ontology. Annotations stored on the server can then be queried and reused for quantitative data analysis such as statistical analysis or to create powerful data visualisations. This focus on semantic annotations is consistent with one of the primary goals of DM2E, which is to enable the integration of metadata (and eventually annotations of the full-texts) and content provided by participating institutions into the semantic web.

The test scenario concentrated on digital humanities researchers affiliated with DM2E institutions whose content was already annotatable using the prototype platform. At the time of the test scenario, these institutions included the Wittgenstein Pilot Group at the Wittgenstein Archives in Bergen (UIB) and philologists working on the digital edition Burckhardtsource.org of the newly acquired (October 2013) associate partner ECORR. One outcome of the test scenario, however, was to create a method of assessing the usability and usefulness of the platform which can be utilised in the future for other content providing institutions and the digital humanities scholars using their content in the associated scholarly domains. The results of these methods should also be useful in the context of the work on scholarly activities of WP3.

6.3 Test Scenario Goals

From the beginning of the project, the purpose of the prototype platform was to develop a proof of concept for DM2E's wider goal of facilitating the creation and reuse of semantically enriched data and metadata extracted from the digital content--in this case manuscripts--to be provided to Europeana in the EDM format. Using the platform, scholars are able to semantically augment full texts and other digital material provided by content providers, thus aiding reuse of the content.

With regards to the developments made in WP1 and WP3 over the course of the project, it was decided that T1.4 should not only highlight the ability of the platform to allow scholars to annotate and therefore provide rich document descriptions of content provided by WP1, but also serve as a bridge between the technical (prototype platform) and scholarly (SDM, research on scholarly primitives) activities of WP3. The goal of the test scenario was therefore to determine the platform's usability and usefulness for the scholarly activities the digital humanities scholars in the test group are engaged in.

Regarding platform usability, T1.4 set out to determine how easy it is for scholars to create, store and reuse semantic annotations they make using Pundit. The results would not only help NET7 to refine their software, but also to determine if the usability affected the scholars' perception of the usefulness of the software. As far as usefulness was concerned, the test

scenario was aimed at discovering the scholar's opinions about the platform as a research tool, the potential reuse of semantic annotations created using the platform and its ability to aid their scholarly activities.

6.4 Challenges for Task 1.4

There were some challenges involved in carrying out the test scenario. For one, there are varying levels of awareness regarding the concept of the semantic web in the digital humanities community. Many scholars, even those in the digital humanities, have not been introduced to the concept or only have vague ideas about its applications. In addition, even if the scholars understood the concept of semantic annotation in principle, they often lacked the technical expertise to know how to analyse the resulting data. This lack of familiarity proved challenging for determining the platform's usefulness, as some scholars could not yet imagine reasons for using the new method of working over their current one.

6.5 Use Case 1: Wittgenstein Pilot Group at WAB

6.5.1 Method

The case study of the Wittgenstein Pilot Group at the University of Bergen, the larger of the two in the test scenario, was conducted by ONB working closely with other members of WP3: NET7, UIB, ECORR, and UBER. In a first step, fourteen (digital) humanities scholars who were familiar with wittgensteinsource.org (DM2E content) and the WAB Wittgenstein ontology but had not yet conducted their own research with the Pundit environment (Wittgenstein Pilot), were given a series of exercises using Pundit to complete on their own. These exercises were created in a collaborative effort by UIB (Alois Pichler) and members of NET7 (Christian Morbidoni, Simone Fonda) and facilitated the creation of semantic annotations and extraction of rich features from digital objects on Wittgensteinsource.org. In addition, these experiments were carefully modelled after actual research questions based on the WAB Wittgenstein ontology. They were then followed up by two means of assessment. One was a survey (ONB, NET7, UBER) with structured and open questions to gather feedback on the scholars impressions of the usability and usefulness of the tool. This survey also included questions created by UBER (Steffen Hennicke) about the scholarly activities the users engaged in. The answers to the scholarly activities questions will be used in the context of Task 3.4 in the last year of the project. The second method of assessment for this first step was an open discussion round with the survey participants at the University of Bergen on the 6th of December 2013.

The second step of the Wittgenstein Pilot Group use case also took place on the 6th of December in Bergen at WAB. It involved a second round of experiments, this time with the platform tool http://ask.thepund.it, which allowed the participants to retrieve the semantic annotations that had been created in the first stage of the experiments. The results of this experiment were assessed in two ways: by real-time observation on the day of the experiment and by a survey sent out shortly after the meeting in Bergen.

The test scenario workflow can be described as follows: Pilot members, who could be reached by a mailing list setup for this purpose, were sent a first round of exercises (UIB, NET7) in September 2013. Data was then collected through an online survey sent out in October and a discussion round (recorded) at the University of Bergen on the 6th of December. At Bergen, a second round of exercises regarding the platform tool http://ask.thepund.it were sent to members of the group. This was followed by another

survey (ONB, UIB, UBER) on the 10th of December. Table 3 (below) provides a list of the test scenario activities for the Wittgenstein Group case study.

Date	Assignments	Survey/Feedback	Link to relevant document	Comment
17.09.2013	Wittgenstein DM2E Pilot: Exercises with Pundit – September 2013		https://docs.google.com/document/d/1SeTbj4LlIrOu4mxxVsbw2km55sSZlMe06UY2G5-sN5Q/edit#heading=h.msrf391ffj2d	First round of assignments using Pundit sent to members of the Wittgenstein Pilot
21.10.2013		Survey for the Assignments with Pundit – (Wittgenstein Pilot)	https://docs.google.com/forms/d/1qe6B3CmeoCRufCa1OMu7LTaG28PvtqySc77FJ_3W014/viewform	First round of questionnaires sent out about Pundit
06.12.2013	Exercises with Ask		https://docs.google.com/document/d/1aGxTmC1lQkMwpmm543vejgfGNpfH_yaXYdo7WuCVHEw/edit	Exercises regarding the tool http://ask.thepund.it
10.12.2013		Survey for the assignments with Ask – (Wittgenstein Pilot)	https://docs.google.com/forms/d/19TeB7o8u-p12wrCH9IrPNmELp6NtvfDGSK5IyNWXQQ0/viewform	Second round of questionnaires regarding the tool http://ask.thepund.it

Table 3: Timetable of task 1.4 activities

6.5.2 Wittgenstein DM2E Pilot: Exercises with Pundit - September 2013

Since the exercises with Pundit are discussed in detail in D3.2, this Section will only provide a brief overview of their content for T1.4.The first set of experiments in the test scenario, which include four assignments created by UIB and NET7, simulate potential research activities of Wittgenstein scholars and are based on actual topics within the particular field of Wittgenstein scholarship. In particular, participants were asked to annotate specific philosophical ideas discussed in two texts of Wittgenstein's Nachlass: Ms-115 (Philosophische Bemerkungen) and Ts-310 (The Brown Book). UIB's digital edition of the Nachlass not only includes content that will be provided to Europeana through DM2E, but it is also accessible, searchable and annotatable with Pundit on Wittgensteinsource.org[41]. In addition, the assignments were built around the WAB Wittgenstein ontology[42], a controlled vocabulary currently being developed at WAB to facilitate the "sharing and debating of Wittgenstein". Not only did the exercises help participants "understand the different ways of annotating a document in Pundit"[43], but by following the directions, participants created meaningful semantic annotations, i.e. rich document descriptions.

In the first assignment, "Annotate some Ts-310 (Brown book) passages which are relevant for the discussion of what grammar is" the participants were asked to "establish a link between a portion of text and a concept taken from a controlled vocabulary. The final shape

41 http://www.wittgensteinsource.org/

42 http://wab.uib.no/wab_philospace.page

43 https://docs.google.com/document/d/1SeTbj4LlIrOu4mxxVsbw2km55sSZlMe06UY2G5-sN5Q/edit#

of the annotation [was]: "Portion of text" : discusses : grammar,"[44] where the participants were free to choose the portion of text from the entire Brown Book. The second assignment was based on Bemerkungen pairs in the Ms-115 and Ts-310. In this exercise, participants "establish[ed] one link between two portions of text [of their choosing] and another link between one of these two texts and a free comment. The final annotation [included] two triples, of the following shape: "Portion of text" : has revision : "Another portion of text" and "Portion of text" : has comment : "Originally it said … and got changed because ..."[45].

The third assignment asked the participants to "connect one language game [in Ts-310] to another Bemerkung's language game, which is a further development of the first one." Here they "establish[ed] two links between two portions of text coming from different pages and the concept "Language game" taken from a controlled vocabulary. This [was to state] that the identified portions are both language games. A third link [...] then connect[ed] the two texts together. The final annotation [...]l include[d] three triples, of the following shape:"Portion of text" : discusses : Language game; Another portion of text" : discusses : Language game and "Portion of text" : refers to : "Another portion of text""[46]. In the fourth and last assignment the participants were able to "annotate all passages in the Brown Book complex where Wittgenstein, implicitly or explicitly, refers to another author or work." In this assignment they were allowed to try out "the different annotation functionalities available in Pundit".

Participants were encouraged to each create a total of 20 annotations. In the end, there were a total of fifteen participants in the experiment (including Kristin Dill (ONB) and Alois Pichler (UIB)) who created a total of 336 annotations.

6.5.3 Results of the Survey for the Exercises with Pundit

The first assessment of the Wittgenstein Pilot Group use case scenario was carried out using a large online survey (Link in Table 5.1) created by ONB in close cooperation with UIB, NET7 and UBER. The survey contained around sixty questions geared at determining the participants' perceptions of the usability and usefulness of the tool. In order to provide context for the answers, participants were first asked about their research preferences and previous knowledge of the semantic web and linked open data. These questions were modelled on those established by UBER for the narrative interviews in the context of T3.4 (WP3). A series of questions designed to assess the usability of pundit and elicit responses that could improve the software were also included. In addition, a number of questions attempted to determine the subjective usefulness of the software by asking for the participants opinions about using Pundit in their scholarly domain. A last set of questions regarding the scholarly activities carried out in each assignment was added to the survey by Steffen Hennicke from UBER. In this set of questions, the participants were asked to select the scholarly activities (provided in the form of a list) they felt they had carried out while doing the assignments.

The Test Group

A total of fourteen respondents, including Alois Pichler from the Wittgenstein Archives, completed the assignments and filled in the survey over the course of two months (from 21 October to 6 December 2013). The participants ranged from the 25-30 age group to the 50-

[44] Ibid.
[45] Ibid.
[46] https://docs.google.com/document/d/1SeTbj4LlIrOu4mxxVsbw2km55sSZlMe06UY2G5-sN5Q/edit#

55 age group. Three respondents were female, nine were male and two did not provide this information. The group was international with members living in or affiliated with research institutions in France, the Czech Republic, Norway, Germany and Canada. As far as academic experience goes, there were master's and PhD-students, post grads, researchers and research associates, university lecturers, assistant professors, university professors as well as a university administrator and a system developer among the informants.

Twelve of the fourteen scholars who took the survey were humanities scholars, one was a computer science scholar (working with humanists) and one was in administration but had an interest in both linguistics and computer programming. Of the twelve humanities scholars, eleven were particularly interested in Philosophy and/or Wittgenstein and one had studied information and library sciences, but was not a philosopher. Six of the twelve humanities scholars can also be considered to be digital humanities scholars and have digital humanities as a research interest, whereas eight of the total fourteen respondents can be said to either be digital humanities scholars or have digital humanities as a research interest. In addition, all of the respondents were familiar with at least one of the two basic concepts at the heart Pundit: semantic annotations and ontologies. Eight of the ten scholars who said they worked with semantic annotations also said that ontologies such as the Wittgenstein ontology were relevant or very relevant to their own work (see Figure 10); two said that they were not relevant.

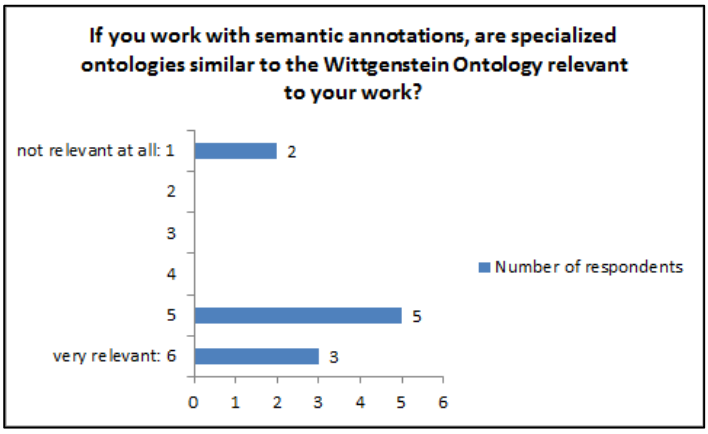

Figure 10: Relevance of ontologies to the scholars' research

Figure 11 and Figure 12 (below) summarise the participants' answers to the survey questions regarding their knowledge of and experience with ontologies and semantic annotations. In general it can be said that all fourteen of the scholars had at least some knowledge of the basic semantic web concepts related to Pundit. Figure 13 (below) gives a breakdown of all informants according to their experience and or knowledge of the digital humanities semantic web topics relevant to Pundit.

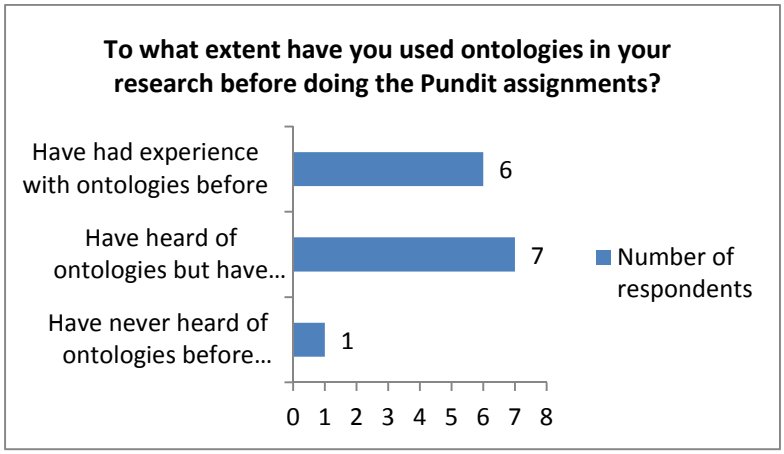

Figure 11: Scholars' experience with ontologies

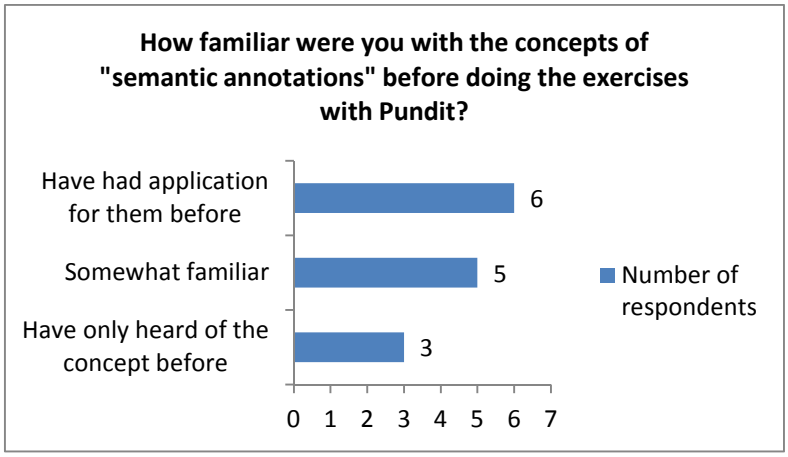

Figure 12: Scholars' familiarity with semantic annotations

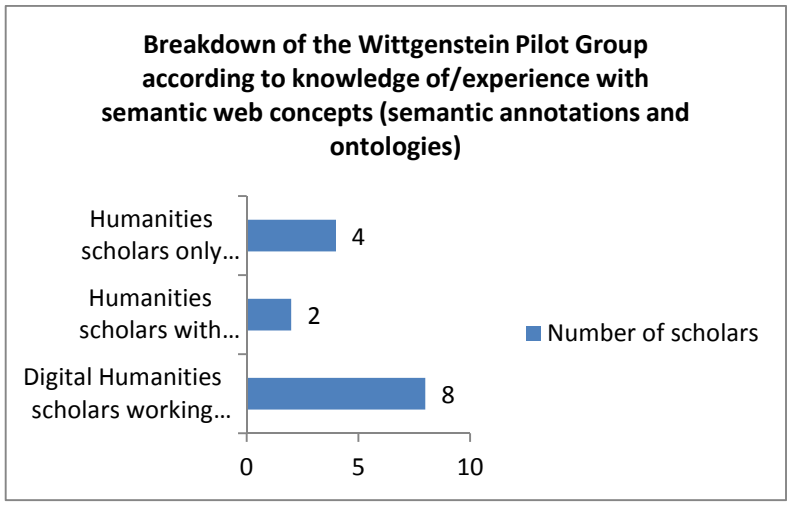

Figure 13: Wittgenstein Pilot Group experience with Digital Humanities and semantic web concepts

Although not all of the respondents can be considered digital humanities scholars in the strictest sense, as a group they represent scholars aware of the field and in different stages of acceptance of digital humanities in general and the semantic web in particular. For example, when asked what digital humanities meant to them, an optional question in the survey, all but one provided a definition. In this context, the informants included admitted

digital humanities scholars working with semantic web tool such as Pundit, traditional humanities scholars who have had experience using semantic annotations or ontologies and are interested in the emerging field of digital humanities as a topic, and at least one respondent who had a negative view of digital humanities in general.

Usefulness of the Prototype Platform (Pundit)

A group of survey questions targeted at discerning the scholar's perceptions of the usefulness of the tool concentrated on three aspects: 1) the relevance of the methodology and techniques (scholarly activities) used in the Pundit assignments to the scholars' own research, 2) the scholars' general willingness to use Pundit in the future and 3) the ways in which the scholars might use the platform.

1) Relevance/Appropriateness of the Concepts and Techniques used in Pundit to the Scholar's Domain

The majority of the scholars (nine out of fourteen) reported it being likely that they would work with semantic annotations in the future. Five answered this question negatively with four saying that it was not likely at all (Figure 14 below).

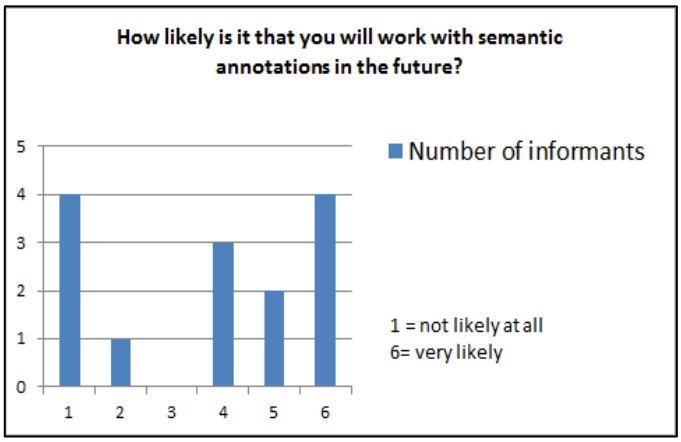

Figure 14: likelihood of working with semantic annotations

When asked what might keep the scholars from using semantic annotations and structured data in future research, the scholars gave the following reasons, which provide insight into their attitudes towards the semantic web paradigm:

- Doubt about the usefulness of semantic annotations
  - *I don't find them practical enough*
- Lack of knowledge about the potential applications of LoD
  - *it is difficult to see an immediate use*
- Deficiencies in available tools
  - *deficiencies in semantic annotation and visualization software*
- Reservations about the quality of the resulting research
  - *if bias is created by necessity* [because the vocabulary is restricted]
  - *provenance, versioning and authorization would need to be dealt with adequately* [this is also a technical issue]

- Incompatibility with Personal research activities
    - *would rather discuss research in person* [instead of publishing notes on the web]
- Technical issues
    - *the software would have to be usable enough*
    - *if there is no long term preservation*
    - *whether the time used learning the platform pays off*
    - *if I have to sign up with companies that use my data*

Scholars were also asked about their previous experience with semantic annotations and ontologies. Six respondents reported having experience using digital tools to create, store and visualise triples before the start of the experiment. The scholars not only mentioned tools, but also models and ontologies:

- Tools for creating triples
    - *XML and Hypertext tools*
    - *XML database with queries from Xpath and Xquery*
    - *Classic spreadsheets* [Excel?]
    - *Standard text processing programs*
    - *Swicky notes* (NET7) [three respondents]
    - *Pundit*
    - *Protege 3.X ontology editor* [two respondents]
    - *Lodlive*
    - *SQL database*
    - *Diigo*
    - *Tools in the EU Discovery Project* [two respondents]
    - *Tools in the EU Agora Project*
- Models and ontologies
    - *Relational models such as ER, EER*
    - *Owl files*

The respondents were also asked about their familiarity with the annotation techniques specifically used in the exercises with Pundit. Figure 15 shows then number of scholars who had experience using digital tools for the activities they engaged in during the assignments. Two of the respondents reported having used digital tools only to add free comments to a text before (one mentioned using Adobe Acrobat for this purpose), four had had not used any digital tools for these types of tasks before.

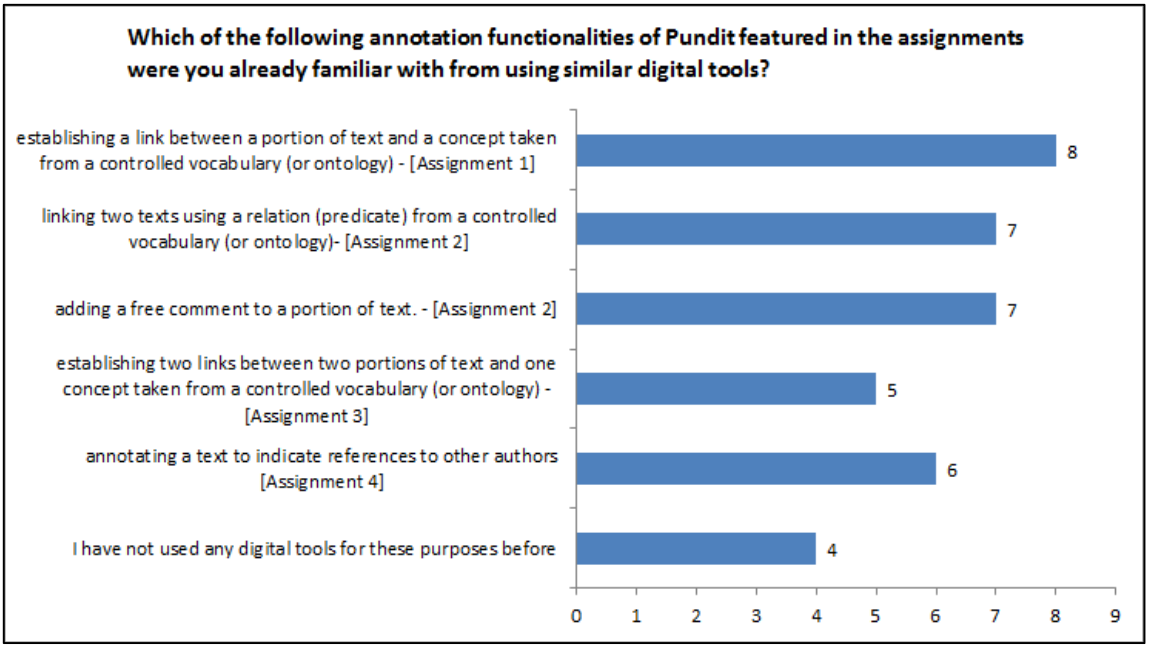

Figure 15: Familiarity with annotation functionalities similar to those found in Pundit

In contrast to the large number of tools mentioned by the scholars in the previous question, the respondents (six in total) reported using very few tools to for these tasks:

- Tools used for all tasks
  - XML and Hypertext tools
  - Standard text processing programs
  - SwickyNotes (Two respondents)
  - Pundit (two respondents)
- Tools used for "Adding a free comment to a portion of text [Assignment 2]"
  - Adobe Acrobat

This is a telling result because it suggests that the scholars, although familiar with semantic annotations and the creation of triples in general, were not used to the way of working facilitated by Pundit and exemplified in the exercises. Since this way of working is seemingly new to them, it is possible that this affects their reactions to Pundit and willingness to use the tool in their research.

Related to the relevance of the activities in the exercises to the scholars' research, we asked the scholars how willing they would be to use Pundit for their research in the future. In particular, we wanted to know if the research activities in the exercises were useful to them. It is safe to say that most of the scholars found all of the functionalities of Pundit highlighted in the exercises useful. This is expressed in Figure 16:

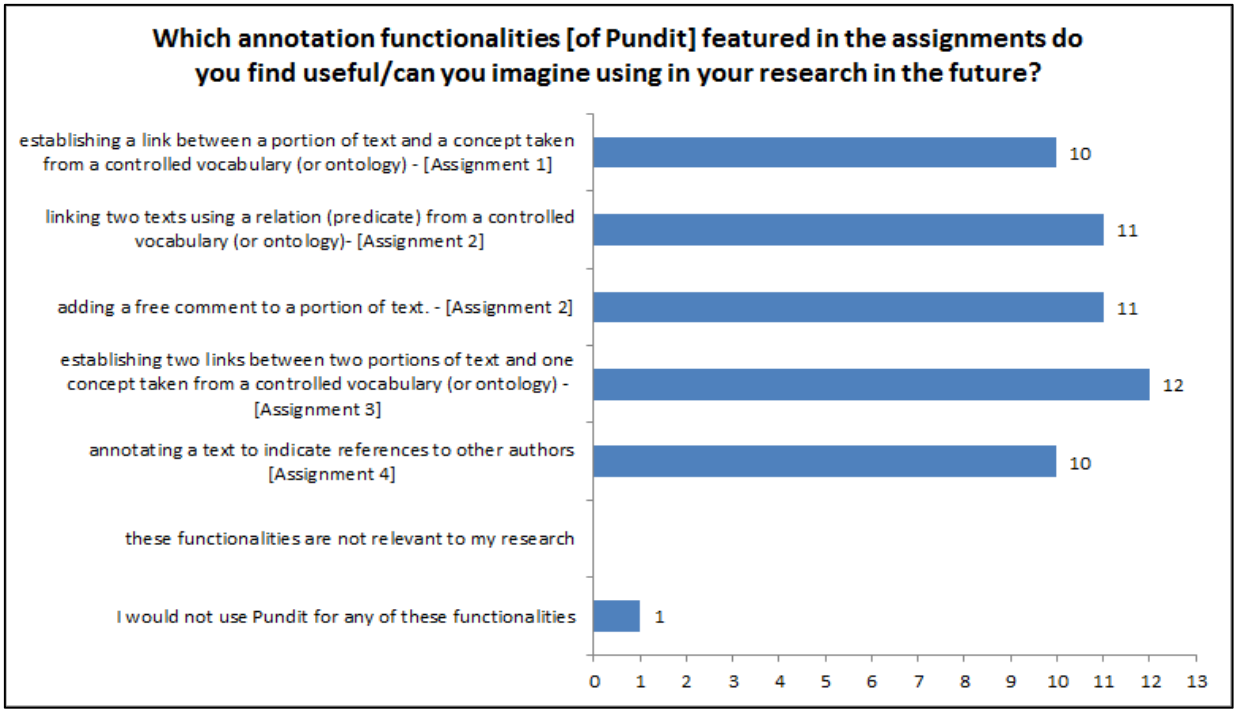

Figure 16: Pundit annotation functionalities found useful by scholars

This question was followed up by asking the scholars how they might use the techniques in these assignments or the annotation functionalities of Pundit in their own research. Eight of the fourteen scholars provided the following answers:

- For research on ontologies and the semantic web
  - *to give concrete examples of the [Wittgenstein] ontology*
- For research in general
  - *to reach a clearer view of the subject*
  - *to add comment[s] or tags*
  - *for propaedeutic reasons - preparing the material, getting a good overview of it, setting up navigational paths/shortcuts, sharing notebooks*
  - *for text genetic work*
  - *tracing the development of [...] concepts in other works*
- For organising research
  - *Assuming that scholarly editions as well as the discourse of the disciplines I am involved in leaves behind the paper mode or its emulation, the manipulation of the vocabulary as well as tools for named entity recognition are implemented as well, I would annotate everything I work with connecting what I work with like in a web based literature administration software, e. g. Endnote or with a better knowledge management Citavi. This of course articulates the need of an efficient way to work with aggregations of semantic annotations and the configuration of the software*
- For teaching purposes
  - *as an assignment in a classroom setting*

- For research on Wittgenstein I
    - *investigating pairs of Bemerkungen that could be useful*
    - *identifying and saving and commenting on Bemerkungen pairs*
    - *systematizing information on Wittgenstein*
    - *tracing the development of [...] concepts in Wittgenstein's Nachlass*

2) Scholars' General Willingness to use Pundit in the Future

The scholars were then asked if they would use Pundit for their own research purposes; ten of the fourteen giving a positive answer, although only two said that they would "definitely" use Pundit. See Figure 17 below:

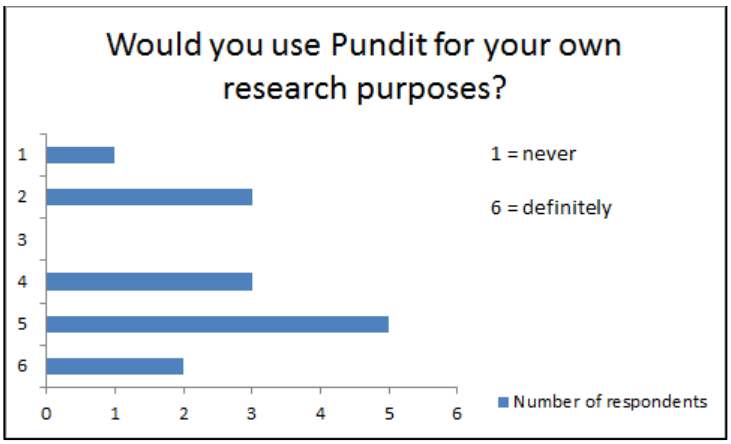

Figure 17: Likelihood of the scholars to Pundit

3) The Ways in which the Scholars might use the Platform

The scholars gave the following examples of research projects they could image carrying out using Pundit. Their answers can be put into the following categories:

- Collaborative scenarios
    - *Collaborative work on the [Wittgenstein] ontology*
    - *Collaborative annotation*
    - *Collaborative project work with people who are in different places*
- Refining the Wittgenstein Ontology / Ontologies in general
    - *competing ontologies*
    - *finding a more effective way of integrating terms to the controlled vocabulary via its "usage"*
    - *Correction and addition to existing annotation (ontologies)*
    - *to further model the Wittgenstein ontology*
- Teaching
    - *This will be an excellent tool to use for most aspects of research and teaching!*
    - *Own research, collaborative research, teaching scenarios.*
    - *teaching of philosophy students at the university in Bergen*

- Meta-Projects
  - *"meta-projects" in the sense of continuing to work on digital humanities solutions*
  - *as a service for similar datasets as Wittgenstein*
  - *it could be any kind of project, just a new way of working*
- Wittgenstein Research
  - *Especially Nachlass-related study of development of Wittgenstein's central notions. The development of his ideas from "middle" to "late" period*
  - *Close scholarly work on Wittgenstein*
  - *research on Wittgenstein*
  - *to annotate the relations between the Nachlass and the discourse around the literature from Nordic Wittgenstein Review to the Nachlass with Annotations*
- Annotating
  - *experiment with Annotations representing the actual citation networks in open educational resources or other corpora of text that are available via open access licences*
  - *Argumentation scenarios*
  - *Correction and addition to existing annotation (ontologies)*

When asked what might keep the Wittgenstein scholars from using Pundit in their research, eight cited the functionality of Pundit, five stated that they did not know what they could do with the annotations and two did not find the creation of annotations to be relevant to their research. Of these three reasons, the first can be tackled by obtaining information about the scholars' perceptions of the usability of Pundit. The next Section deals with exactly the first of these obstacles and discusses the scholars' observations of Pundit's functionality as reported in the survey. As far as the second reason goes, the hope was that the introduction to "Ask" and the visualisations in the second set of exercises would help the scholars to see the applications for the semantic annotations.

Usability of the Prototype Platform (Pundit)

The usability section of the survey concentrated on three aspects: 1) the user experience, 2) the particular problems the scholars had working with Pundit, and 3) suggestions for improving pundit.

1) User Experience

When directly asked about user experience, the feedback was mostly on the positive end of the spectrum. Although ten out of the fourteen respondents said that Pundit was more easy than not to use, most gave the tool a rating of 4 (five respondents) or 5 (four respondents) on a scale of 1 to 6, where 1 means that Pundit is "very difficult" to work with and 6 that it is "very easy" to work with (see Figure 18). When asked about the respondents user experience, nine out of fourteen found it to be more positive than negative, although most of the respondents rated it either as a 4 or 5 on a scale from 1 to six (1 = poor, 6 = excellent) (Figure 19). Both charts below indicate that Pundit has some room for improvement regarding usability, since some scholars found it to be very difficult to work with Pundit and the user experience to be poor. However, the experience of most of the Wittgenstein scholars was generally more positive than negative.

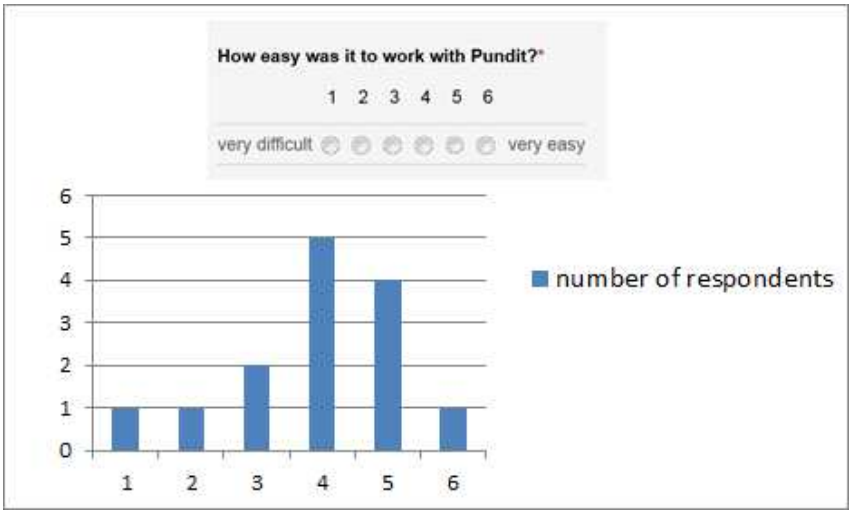

Figure 18: Scholars' perception of Pundits operability

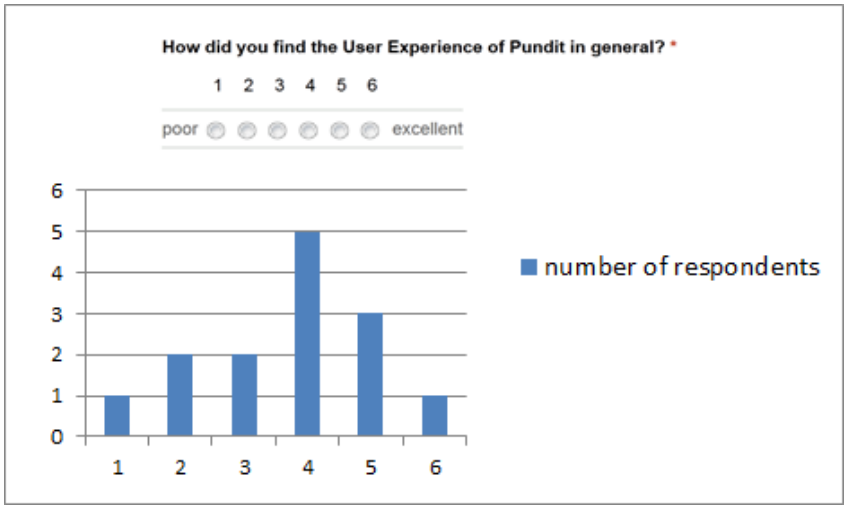

Figure 19: Scholars' user experience with Pundit

2) Technical Difficulties when Working with Pundit

The scholars mentioned the following difficulties when working with Pundit:

- Understandability
  - Problems with the interface: respondents did not know what happened to the annotations after they had made them
  - Respondents could not edit the triples after having made them. An edit feature has been added by NET7 as a result
  - Difficulties logging in, viewing and finding the annotations that were made
- Operability
  - Bugs
    - problems with the text-marking functionality
    - overlapping text in the drop-down menus
    - pop-up menu out of reach (outside of frame)
    - Pundit could not save enough triples at once

- Malfunctioning
  - screen freezing
  - losing triples
  - needing to refresh the page

3) Respondents Suggestions for Improvements

➢ Suggestions for improving current functionalities

- improve the functionalities for searching and browsing vocabularies
- the program should be more intuitive
- a search function should be integrated in Pundit
- better integration with Muruca-Suite
- make a simple way to show text fragments/whole texts side by side
- create standard options for creating tree triple-connections,
- advanced editor option
- my items folder should automatically update in each open browser with Pundit

➢ Suggestions for additional functionalities

- integrated translation tool
- creating triples by drag & drop
- annotation features of Pundit integrated into the Wittgenstein site
- autocomplete for searching and browsing vocabularies
- annotations should be annotatable
- a graph visualisation of the triples
- search function to move freely between pages of a text
- be able to use more than one window in Pundit annotation environment
- list of standards
- menu of possibilities
- editing of annotations
- reification
- search function for my items
- named entity recognition

6.5.4 Wittgenstein Pilot Meeting in Bergen

A second form of assessment was set up with regards to the Wittgenstein Pilot experiment. On the 6th of December 2013, members of DM2E involved in T1.4 met with the scholars involved in the Wittgenstein Pilot who had completed the online survey in Bergen, Norway. The meeting was organised by OKFN (Sam Leon), UIB (Alois Pichler), ONB and NET7 (Christian Morbidoni, Simone Fonda). The purpose of this meeting was to discuss the exercises and the results of the survey and to elicit further responses about Pundit from the group. In addition, participants were also introduced to http://ask.thepund.it, another component of the prototype platform. For evaluation purposes, the discussion was recorded and a transcript was made.

During the morning session, which had originally been organised with the goal of facilitating an in depth discussion of the scholars' reactions to the exercises with Pundit, the scholars

concentrated instead to a great extent on fundamental theoretical and practical issues of the linked open data paradigm. In this context, the discussion gives insight into how the scholars assessed the usefulness of Pundit for their scholarly activities. It also provides a solid basis for the scholarly domain activities of the next year of the project and will therefore only be briefly discussed here.

- Many of the issues discussed by the scholars uncovered challenges of using the linked open data paradigm in the digital humanities in general. These challenges included:
  - **The limited availability of resources in the LoD cloud**. Scholars desired to be able to access the bibliographical information and content of standard works in the field of Wittgenstein to augment their annotation activity with Pundit.
  - **A perceived difference between current scholarly activities and the modality of working with linked open data**. Some scholars mentioned a workflow of close reading, intense production of notes and face to face scholarly exchange of ideas that they felt was not entirely compatible with the strict annotation workflow of making of triples to publish on the web. These scholars felt a more informal method of annotation such as a chat function would therefore be needed in addition to the annotation environment to more appropriately represent these activities.
  - **Uncertainty about the exact nature of the semantic annotation and its publication status**. There was some discussion around the question of whether scholarly "guesses" and "notes" should be available as semantic annotations on the web or if only publishable representations of knowledge were appropriate. Others wanted to publish the entire discourse around a certain triple including the more informal discussion.
  - **Concerns about creating/sustaining credibility in an open medium**. Some scholars expressed the importance of knowing the level of certainty a scholar has about the accuracy of an annotation he has made. This issue is related to the issue of uncertainty about the exact scholarly nature of the semantic annotation. One respondent said "When [an] author starts to publish all his guesses, all his notes, then I cannot any longer rely on the validity of these three lines."
  - **Concerns about intellectual property rights:** Scholars wanted to be able to license their notebooks, for example.
  - **The frustration of working with an already heavily annotated text.** Some scholars were wary of the fact that a text might become too heavily annotated by one group of scholars thus reducing the attractivity to others who would like to create their own annotations.
  - **Problems with ontologies in general.** Those not familiar with the WAB Wittgenstein ontology needed descriptions of the entities. Users don't necessarily know the domains and ranges of the predicate in the ontology without having to look them up. Some scholars mentioned that an ontology creates bias and limits to scholar.
  - **Complex annotation needs.** The more complex the annotation needs of the scholars, the more problems they had using Pundit to create the triples. Scholars described a workflow in which they first map out the concepts they

would like to annotate for many pages or a whole text and then they would like to apply this on a larger scale to a whole text or groups of texts.

- On another level, there were differences between the expectations of the humanities scholars and the expectations of the developers of Pundit about what Pundit should be able to do. These differences mostly revolved around the limitations caused by Pundit being a tool designed to flexibly annotate webpages and that relies on urls to create the triples.
    - **Scholars did not want to have to leave the environment for browsing and discovering activities.** At the moment, scholars have to open up a separate window and load the Pundit for each page they want to annotate.
    - **Some scholars wanted Pundit to facilitate the scholarly activity of discussing.**
    - **Scholars want to create entities and resources on the fly without being forced to use preconstitued vocabularies.**
    - **Scholars wanted to be able to use Pundit to augment the ontology**
    - **Scholars wanted integrated visualisations**. It was not enough for the scholars to simply create triples, they also wanted visualisations of the metadata, for example in the form of a map of the Nachlass.

6.5.5 Wittgenstein DM2E Pilot: Exercises with Ask - November 2013

During the meeting in Bergen on the 6th of December 2013, the Wittgenstein scholars were introduced to and given the opportunity to work with http://ask.thepund.it.[47] The purpose of this tool was explained to the scholars as follows: "Ask is a web platform for sharing annotations and notebooks created with Pundit. While Pundit's role is that of enabling people to annotate web pages of interest, Ask is where such annotations, once made public, can then be searched and explored. Additionally, in Ask you find a MyAsk tab: it offers single users a view of their personal (public or private) notebooks."[48]

In this second experiment, the participants were given four assignments to complete. The first assignment showed the scholars how to search for public notebooks using 'Ask'. In this assignment they were able to load all of the public notebooks on the platform and to sort the notebooks by date, author, name and title. They were then asked to look for the notebooks created by the Wittgenstein Pilot Scholars in first assignments and to explore the annotations (triples) within them. In the second assignment, participants were asked to explore the options within their own notebooks such as changing the name or visibility (from public to private) of their own notebook. A third assignment explored the 'faceted search function' of the tool. This function allows users to "query the system" and "to deeply explore the structured data [...] from a number of notebooks"[49]. The fourth assignment invited scholars to take a look at visualisations of philosopher's influences created by Christian Morbidoni.

[47] http://ask.thepund.it
[48] https://docs.google.com/document/d/1aGxTmC1lQkMwpmm543vejgfGNpfH_yaXYdo7WuCVHEw/edit#
[49] https://docs.google.com/document/d/1aGxTmC1lQkMwpmm543vejgfGNpfH_yaXYdo7WuCVHEw/edit#

6.5.6 Results of the Survey for the "Exercises with Ask"

This Section deals with the results of the survey about the "Exercises with Ask". It is divided into four parts. First, there is a short discussion of the observations made during the meeting of the scholars as they were using http://ask.thepund.it. Then, there is a section about survey results, which can be divided into 1) the scholars' user experience 2) the usefulness of 'Ask' and a section about 3) the scholars' opinions of the visualisations explored in the "Exercises with Ask".

Observation

The afternoon session was smaller with seven of the original fourteen scholars being able to stay. During this session, NET7 had the opportunity to directly observe the scholars in real time as they were learning how to use the tool. Christian Morbidoni and Simone Fonda took note of a few small bugs not worth mentioning here; the scholars suggested features they would like to have in 'Ask' such as the possibility of organising annotations by the date they were made and grouping the statements by selected facets. The feedback given in the discussion was mostly positive with respondents saying that they found the tool to be easy to use. The 'facets' feature was especially well received.

Survey Results

1) User Experience with 'Ask'

A total of five Wittgenstein scholars including Alois Pichler (UIB) took the survey about Ask.thePund.it. In general, the scholars reported having a positive user experience, with four of the five respondents giving 'Ask' a 5 on a scale of 1 (= poor) to 6 (= excellent). The scholars also all found 'Ask' to be relatively easy to work with, with two respondents giving the tool a 4 and three respondents giving the tool a 5 on a scale of 1 (= very difficult to work with) to 6 (= very easy to work with). This can be seen in Figure 20 and Figure 21.

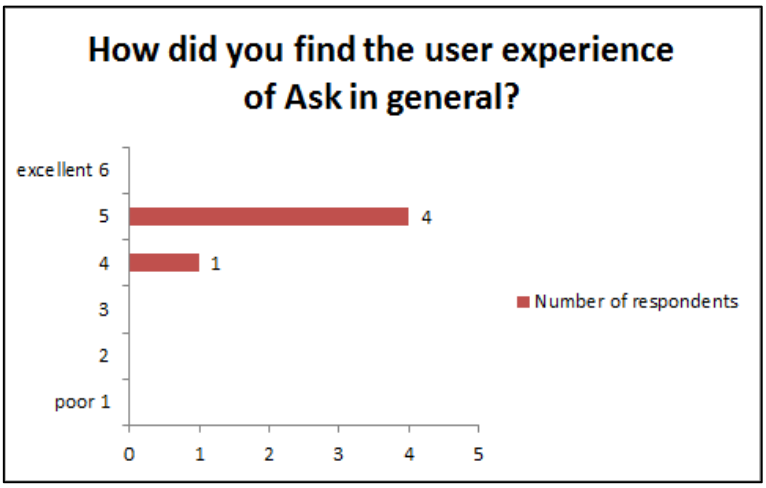

Figure 20: User experience of 'Ask'

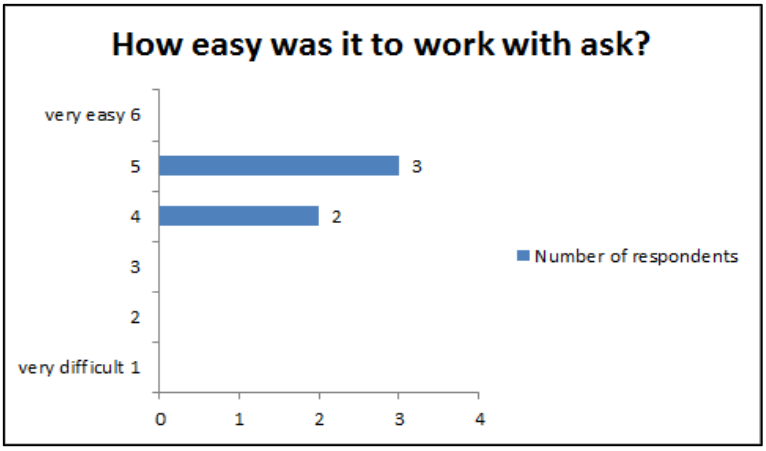

Figure 21: Ease of using 'Ask'

When asked about their impressions of Ask regarding usability, the scholars gave the following answers:

- **Understandability (Layout/Intuitive Design)**
  - *The layout is very clear, everything was quite simple to find*
  - *layout is convenient in my opinion*
  - *I found it intuitive. I think it is not possible to make this tool much simpler or more easy - because the information structure that it handles is complex*
  - *I could easily find everything.*
  - [it was] *unclear how to open several notebooks at the same time*
- **Operability**
  - **Speed**
    - *It is important that it runs fairly quick. This goes for every web tool today, of course. I felt it maybe ran a little slow at some points*
  - **Malfunctioning**
    - [the respondent had to reload Pundit after having already]
- **Attractiveness**
  - *Facets is a good concept, I find. I think it will be recognised by most people who have worked with information science or library*

2) Usefulness of Ask

In the survey about the 'Ask' tool, scholars were asked four open questions relating to the usefulness of 'Ask': 1) If they could imagine using 'Ask' to store and share annotations, 2) if they could imaging using 'Ask' as a collaborative research tool, 3) which functionalities of 'Ask' they would like to see improved upon and 4) which functionalities of 'Ask' they think should be added to the tool.

The first two questions were answered positively by almost all of the scholars, but with some qualification (more search functionalities). In response to the first question mentioned in this Section, one respondent said he/she would still like store annotations offline, but the rest had very positive feedback:

- Can you imagine personally using Ask to store and share your own annotations? Why or why not?
  - *Yes. It works, it is intuitive, it is powerful.*

- *Not now. Ask is still a beta tool. I prefer to store my annotation offline.*
- *Yes, I would. Primarily because it's the only good way of gaining an overview of my own (and other's) annotations. But I think also with a good search function and more functionalities for managing notebooks it would be a good way to store annotations and use them.*
- *Yes. Because it provides the overview needed: where are my annotations; who else has annotated and where and what.*

When asked if the scholars would use 'Ask' as a collaborative research tool, the response was also positive:

- Can you imagine personally using Ask as a collaborative research tool? Why or why not?
  - *Yes, I could. Especially if I knew I were to work on the same material as someone else, e.g. with a view to writing a joint paper. But also more generally just to exchange annotations as ideas.*
  - *I can imagine. For example a group exercise of annotation. Maybe a small group of annotators, why not students? Then one could easily navigate the contents by using Ask, I think.*
  - *Yes, but the corpus must be larger.*
  - *Fairly intuitive.*
  - *Yes. Because it eases communication between different annotators.*

The last two questions were answered with ideas for functionalities that give insight into how the scholars work and would make for a better tool. In summary, the scholars would like improvements in the help function and more powerful search functions. In addition, they would like to be able to merge notebooks and to have some kind of a visualisation of the facets. The scholars' answers are given in full below.

- Which functionalities of Ask would you like to see improved and in what ways
  - *The help function. Often I am lost in the tool and need help.*
  - *The possibility to merge notebooks and annotations (to get an actual collaborative research output).*
  - *Facets is great; maybe implementing some visualization there.*
  - *Search functionality is very, very important. It should be possible to search for all elements of the WAB ontology (once that's implemented properly) and then get to know what remarks have been annotated with these in triples (both for my own and shared notebooks). It should also be possible to string search annotational comments. Furthermore, it is vital that it is easy to manage notebooks: to delete and move annotations between my notebooks. The possibility to merge notebooks and annotations (to get an actual collaborative research output).*
  - *I don't know. It would require that I had worked more with it, I think, than I have today.*

- Which additional functionalities of Ask would improve the tool?
  - A search function. I would like to have additionally "querying" functions: Give me all the triples that are identical. Give me all the triples that contradict each other / disagree with each other. Same point as to searching and management, I think.
  - merging functions I would like to have there a search function similar to that in the Bergen Electronic Edition.
  - Not sure.

Usefulness of the Visualisation Demos

Visualisations are powerful ways of structuring, querying and generally making sense out of the data captured by semantic annotations. The prototype platform does not have the capability to create visualisations such as graphs and timelines of a scholar's data automatically, as this requires a technical step out of the bounds of the tools as well as annotations carefully tailored for the purpose. NET7 does, however, have demo visualisations available, so that scholars can get an idea about the possibilities of semantic annotations. In this context, an exploration of the visualisation demos was built into the exercises with 'Ask': a simple timeline of city of Ancona and an EdgeMap graph showing philosopher's influences. We asked the scholars if they could imagine personally visualising their own data/results using similar digital tools. Four said yes and one didn't know. We then asked them how they might visualise their results using tools similar to TimelinJS and EdgeMaps. The answered in the following ways:

- If you [can imagine personally visualising data/results of your own research using similar tools] please describe the data you would collect and the visualisations you might create based on this data.
  - *concepts and sources*
  - *Timelines on the basis of datings for Wittgenstein's texts. Maps of influence for textgenetic paths. Maps of influence for influence on Wittgenstein. Maps of influence for annotators' interaction.*
  - *I would like to visualise different ontologies in order to compare them.*
  - *For me I'm thinking more along the lines of having triples visualised as graphs, sort of like in SWickyNotes. It ought then also to be possible to navigate on the basis of the graph.*

6.6 Use Case 2: Burckhardtsource

6.6.1 Method

A second, much smaller, survey based on the Wittgensteinsource survey was carried out with the scholars from the DM2E associate content provider ECORR, which stands for the ERC AdG project EUROCORR - "The European Correspondence to Jacob Burckhardt". Within the project, ECORR is also referred to as with Burckhardtsource, which is the name of the project's website hosted by the Scuola Normale Superiore of Pisa on which they have made their digital edition of the European Correspondence to Jacob Burckhardt freely accessible: "The aim of this project is to map and publish in a critical edition the extensive

correspondence of European intellectuals with the Swiss cultural historian Jacob Burckhardt over a period of more than half a century, from 1842 to 1897."[50]

The two scholars interviewed in this case scenario represent a different set of test subjects than those from the Wittgenstein Pilot Group. For one, whereas the Wittgenstein scholars were primarily from the domain of philosophy, the Burckhardtsource group consisted of scholars interested in literature, art and art history. Another more important difference is the fact that the Wittgenstein Group had not used Pundit to systematically augment the Wittgensteinsource corpus or prepare the digital library for publication. In addition, most of the members of the Wittgenstein group (all but three) were not involved in editorial philology. As a result, the focus of the participants was more on the theoretical and practical implications of the Ontology and using the corpus for deep philological reading by scholars not involved in digital editorial philology. In contrast, the scholars at Burckhardtsource systematically created 3513 annotations using Pundit. Most of these (3000) are annotations of persons, places and dates but another 500 were used for proof-reading while constructing the edition. At the moment, Burckhardtsource has not published visualisations of the metadata, because the research group is still working to make the data more representative. They do, however, have two demos that show the types of visualisations they are working on:

- EdgeMaps demo of Burckhardtsource letters (See Figure 22)
  http://metasound.dibet.univpm.it/edgemaps/maps/test.html#letters;map;;Giottino
- TimelineJS demo with the letters involving Wilhelm von Bode (See Figure 23)
  http://metasound.dibet.univpm.it/timelinejs/examples/bur-bode.html

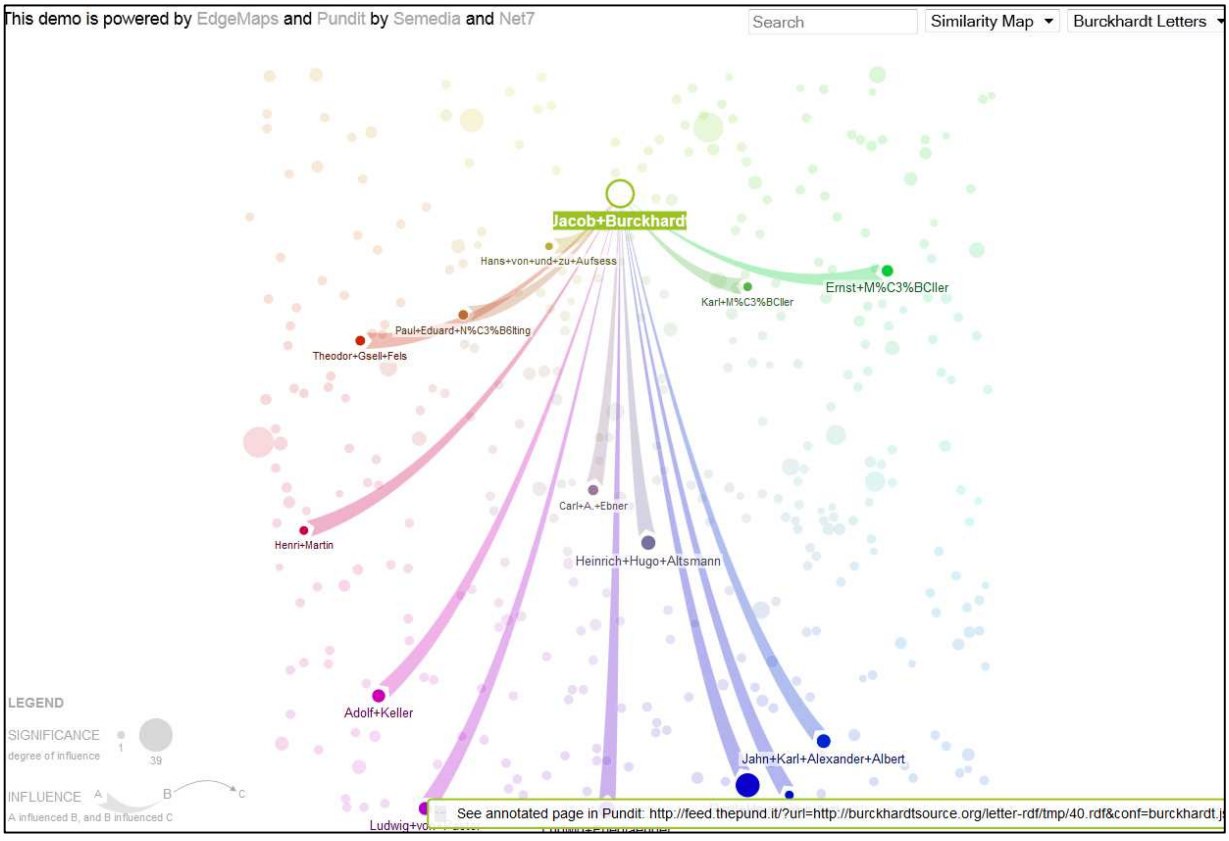

Figure 22:Screenshot of EdgeMaps demo with Jacob Burckhardt as the main node

[50] http://www.burckhardtsource.org/

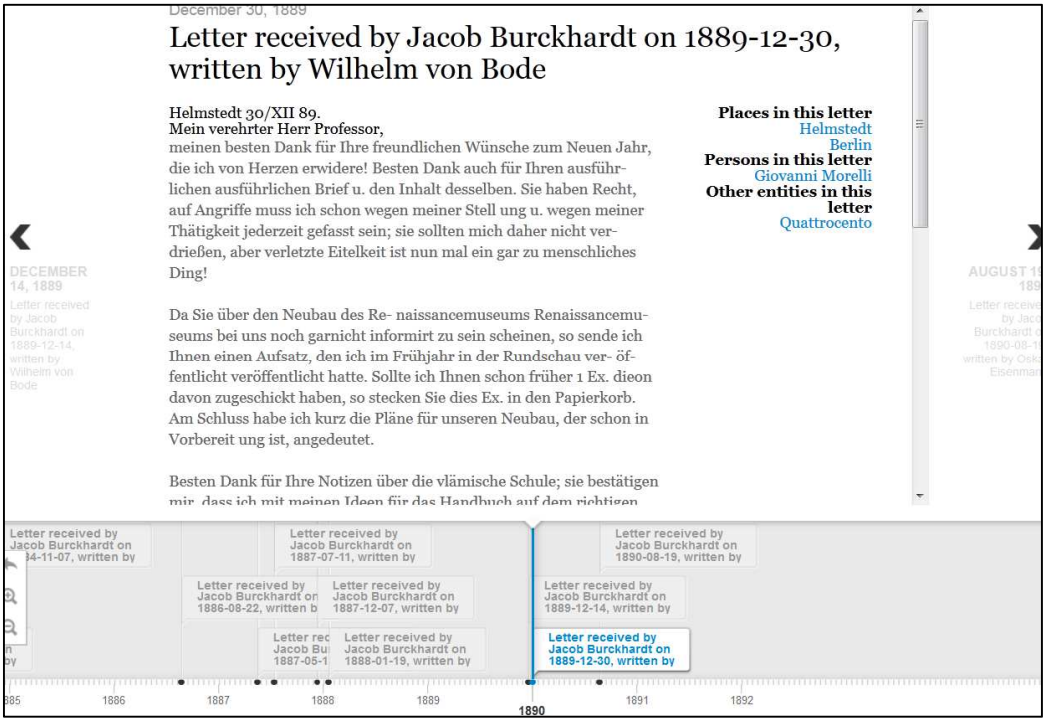

Figure 23: Screenshot of the TimelineJS demo with a letter written by Wilhelm von Bode to Jacob Burckhardt on 30 December 1889

These demos speak volumes when it comes to assessing the (possibilities of the) prototype platform and they provide examples for what the content providers of WP1 will be able to do with their metadata once it has been ingested in the WP2 platform and made annotatable with Pundit. The scholars have created a number of rich document descriptions (semantic annotations), which can be reused by other scholars from any field within the humanities interested in the letters and their contents. The first demo uses EdgeMaps to show the graph of influence between the correspondents and historical intellectual figures mentioned in the letters. The second demo uses TimelineJS to display the letters relating to Wilhelm von Bode in the form of a timeline. This second demo also provides links to DBPedia for places, persons and other entities mentioned in each letter.

Since the Burckhardtsource scholars already had extensive experience using the elements of the prototype platform, a different survey for assessing its use by the digital humanities scholars needed to be created. As a result, the Burckhardtsource questionnaire was geared at ascertaining how the group had used Pundit and what they thought of their experience with it. The survey was sent to the scholars and completed in the second half of November 2013.

6.6.2 Results of the Burckhardtsource Survey

The small Burckhardtsource survey[51] was created by the ONB with advice from Francesca Di Donato of NET7/EUROCORR. It consisted of 23 questions asking about the scholars' research, use of Pundit and their assessment of its use in the digital humanities. It was filled out by two of the three members of the team working on the philological aspects. The scholars ranged in age from 30 to 40; both were post docs with a philological background. Of the two scholars, one was a self-described digital humanist, the other "didn't know" if this title applied.

[51] https://docs.google.com/forms/d/1AdF1m8CfKZrGRBpBEXYZL-qY8Afyp1pb52RU6Oxh4QM/viewform

One of the scholars described their research interest in the following manner: "Through personal and official letters to Jacob Burckhardt we reconstruct a wider cultural panorama in a specific historical period." The answers to the following questions show how they went about reaching this research goal using Pundit. 1) The first set of questions deal with the particular functionalities of Pundit that they are using. 2) A second set of questions deal with the reuse of the rich textual descriptions they have made. 3) The third group of questions relates to any ongoing technical issues they might be having with the prototype platform.

1) Functionalities of Pundit being used by the Burckhardtsource Group

- What is your reason for using Pundt and the other Pundit related tools from NET7 (Ask, Feed, Korbo, etc.)?
  - *Pundit allows new improvement in the field of Digital Humanities on both a horizontal and a vertical axis: horizontally by creating a net linking new content together, vertically thanks to a deepening research tool.*
  - *In our project (digital edition of unpublished letters) we use Pundit for annotating contents such as names, dates, places, institutions, works of art and texts. Thanks to the annotations the reader is facilitated in his understanding of the sources; more experienced users are also able to create "intellectual mappings" of an author (i.e. Jacob Burckhardt) in his relations to the contemporary world; topics and interesting fields can be highlighted.*
- Which functionalities of Pundit and its related tools are important for your Pundit-based research activities and why?
  - *Among the different functionalities of Pundit, I would especially mention the tool to create triple (allowing further research possibilities), free annotations (in order to extend information and fill possible gaps) and customised vocabularies.*
- What types of semantic annotations have you created using Pundit and what was the purpose (i.e. research focus) of the annotations?
  - *Triples to identify single elements with a view to further uses.*
  - *Free annotations to add information.*
  - *Use of customised vocabularies.*
  - *The purpose is to add information in general, but also to prepare future researches.*
  - *We use Pundit for annotations of persons, dates, places, institutions, works of art and texts: all these informations enrich our sources and help the reader to understand the texts.*

2) Reuse of the Work of Burckhardtsource by other Scholars.

- In what ways can your or other (digital humanities) scholars use your annotations for further research?
  - *Annotations are the result of a previous research process, can be reused, but are also liable to improvements and additions.*

- To what extent do you consider Pundit to be a digital humanities research tool?
  - *Its strength in Digital humanities lies in its flexibility.*
  - *I think Pundit can play an important role in popularizing applied research and linking scientific communities in the humanities especially in the field of digital editions. More theoretical researches need reflection and traditional instruments.*

Technical difficulties with the Prototype Platform

- What (recurring) difficulties do you have when working with Pundit?
  - *Through the practical use of Pundit, we sometimes need further tools or predicates, while other functionalities appear not always immediately clear.*
  - *Restrained vocabularies that have to be enriched by the editor; the lack of authority files in the specific field of artworks.*
- What functionalities or features would make using Pundit easier/better?
  - *I would probably lighten some passage and remove some function or better shift them to later steps in the annotation process.*
  - *I have experienced problems with browsing vocabularies that have blocked by Pundit.*
- Is there anything else you would like to say about your experience with Pundit that was not addressed in the previous questions?
  - *Pundit is an interesting and fancy tool, but it still attends to be more functional and reliable.*

6.7 Summary of the results of Task 1.4

The test scenario attempted to assess the prototype platform's ability to facilitate the creation and reuse of rich document descriptions by digital humanities scholars. The test scenario focussed on digital humanities scholars' perceptions of the usability and usefulness platform's tools. Two use cases with digital humanities scholars at DM2E content providing institutions formed the basis of assessment.

The first use case concentrated on scholars of philosophy affiliated with the Wittgenstein Archives at the UIB: the Wittgenstein Pilot Group. This group consisted of fourteen participants who were familiar with both Wittgensteinsource.org and the semantic web concepts of ontologies and semantic annotation, but had not used the prototype platform (Pundit) for their own research purposes. For this group, Alois Pichler (UIB), Christian Morbidoni (NET7) and Simone Fonda (NET7) created a set of exercises for the participants to create meaningful semantic annotations with Pundit using the WAB ontology that realistically simulated possible annotation activities of a group of digital humanities scholars. The second use case dealt with the scholars at ECORR, a newly acquired associate content provider, who are currently using Pundit for the philological preparation of their online edition "Burckhardtsource.org" – a collection of the correspondence of European intellectuals with the Swiss cultural historian Jacob Burckhardt from 1842 to 1897. This group has already created over 3000 semantic annotations using Pundit which they have connected with the linked open data cloud. These philological scholars, two in total, were sent a survey to evaluate their experience with Pundit. Whereas the Wittgenstein scholars

represent a wider target group of potentially interested but uninitiated digital humanities scholars that DM2E in general and NET7 in particular should try to reach out to with dissemination activities, the Burckhadtsource group represents actual digital humanities scholars using the prototype platform for publishable research. As a result, the test scenario gained insight into the possible attitudes and perceptions of scholars who will be encountering Pundit for the first time as well as into the scholarly habit of those already working with the tool.

The results of the study of both use cases were helpful in providing DM2E with feedback about the usability and the usefulness of the prototype platform. The Wittgenstein Pilot Group, who were mostly interested in using the tool to facilitate collaborative scholarly discussion of the Wittgenstein Nachlass and to refine the WAB Ontology, not only provided DM2E with an insight into the scholarly research activities for which they could imagine using Pundit and 'Ask', but the feedback regarding their experience with the prototype platform also uncovered the fundamental theoretical and practical issues of getting scholars to use the linked open data paradigm. Although the scholars were familiar with semantic annotations and ontologies, for most Pundit's functionalities represented a new way of working for them. However, the majority of the scholars were open to using the annotation activities explored in the assignments in their own research in the future. The scholars also collectively came up with a large list of possible uses for Pundit that went beyond merely annotating texts but included propaedeutic reasons, close reading of the material, creating citation networks, collaborative research scenarios, teaching, expanding upon current digital humanities solutions and refining the Wittgenstein ontology. This shows the prototype platform to be a very versatile tool that can be used by scholars for a number of scholarly activities. With regards to the usability of Pundit, the majority of scholars who took each survey found Pundit and 'Ask' to be easy to work with, rating it at four or higher on a scale of 1 (=very difficult) to 6 (=very easy). They also found the user experience to be relatively positive. Those who gave the most negative marks here were the least familiar with or most critical of the digital humanities paradigm or did not find annotations to be relevant to their own research. Many of the bugs and malfunctions were able to be directly dealt with by NET7 at the meeting in Bergen on the 6th of December 2013 and the developers are already working on or have created a number of solutions for the improvements the scholars wanted to see: users can now, for example, edit annotations. Other functionalities requested by the scholars either are outside of the boundaries of the tool (chat function), or can be solved by combining Pundit with other open source tools (translation tool). 'Ask' the Pundit was very well received with most scholars finding the tool to be 'very intuitive'. Scholars also greatly appreciated the 'facets' feature, which helped them to easily navigate the annotations.

One results of the Wittgenstein Pilot Group survey is particularly interesting for the wider context of DM2E. Over the course of the assessment activities it became clear, that Pundit is a great tool for instigating discussion about the wider issues of the linked open data paradigm. In this context, a number of scholars said that they felt Pundit was an excellent tool not only for annotating texts, but also for working on digital humanities solutions in general. Particularly the discussion in Bergen demonstrates this. Here, scholars showed themselves to be less concerned about discussing how they could use Pundit and 'Ask' in their own research and more concerned about discussing the challenges of the linked open data paradigm. Among other issues, the scholars were concerned about the limited resources of in the LoD cloud, the publication status of the annotation, creating credibility in an open medium, the intellectual property rights of the annotation and the fundamental characteristics of ontologies. Discussions of the prototype platform between digital humanities scholars and the developers have indeed helped to create solutions for these issues. For one, NET7 has already created the option of public and private notebooks to deal with the scholars' concerns about creating credibility. Scholars can keep their "guesses" private and make their more publishable annotations public.

The results of the survey with Burckhardtsource are especially important for assessing the prototype platform, because the scholars have used the tool extensively for their own research and have created a number of systematic annotations that they are preparing to publish in the form of meaningful visualisations. Whereas the Wittgenstein Pilot survey reflects attitudes of scholars not yet used to working with Pundit, the Burckhardtsource survey shows scholars attitudes about actual results of research being carried out using the tool.

The philological scholars of Burckhardtsource reported using Pundit for annotating names, dates, places, institutions, works of art and texts in their digital edition. They asses the value of their work with Pundit positively and believe that these annotations help the reader to better understand the sources. Digital humanities scholars with knowledge of the applicable tools can reuse their annotations to create "intellectual mappings" of the author and to highlight interesting fields in the corpus. The scholars also mention Pundit allowing improvements in the field of digital humanities on a horitonal and vertical axis. Horizontally, Pundit allows the researcher to created links between sources. Vertically, the tool assists in the deep analysis of one particular text through the visualisations of data that can be created. According to the scholars, another benefit of using the prototype platform was that it can help to 'popularize applied research' and to 'link scientific communities in the humanities especially in the field of digital editions'.

Although the Wittgenstein and Burckhardtsource groups represent scholars in different stages of their familiarity with Pundit, both mentioned Pundit not being able to replace some traditional aspects of working. Whereas some members of the Wittgenstein group mentioned a workflow of intensive close reading and notetaking that they did not feel to be completely transferable to the semantic annotation paradigm of Pundit, Burckhardtsource mention the need for the use of "traditional" instruments for "more theoretical research". Another similarity can be found in both groups desire to improve on the ontologies they were using (*Restrained vocabularies that have to be enriched by the editor)* and dead ends when it comes to finding certain concepts in the linked open data cloud (*the lack of authority files in the specific field of artworks.*)

In summary, the evaluation of the prototype platform in the context of T1.4 yielded predominantly positive results. The use case of Burckhardtsource shows that Pundit is being used by scholars in the digital humanities domain to create rich document descriptions that are useful for both the scholars working with the material and other scholars who may want to reuse the semantically enriched data that they have created. The use case scenario involving the Wittgenstein Pilot Group shows that, although scholars not intimately familiar with the way of working facilitated by the prototype platform may have some difficulties reconciling it with their current research methods, they can generally see the possibilities of using the platform in digital humanities research. In addition, this use case scenario strongly shows that the prototype platform is an excellent set of tools for helping scholars to come up with further digital humanities solutions regarding the linked open data paradigm.

7 Summary and WP1 Outlook

This document describes the results of DM2E's WP1 Task 1.3 "Testing the User Interface for Creating Mapping, Interlinking Heuristics and for Configuring the Workflow", Task 1.4 "Setup a test scenario for the prototype platform" and Task 1.5 "Organise the final content integration". It represents the summary of the work done in WP1 during project months 7-24.

In Task 1.3, the tools provided by WP2 were evaluated for their suitability for content providers to translate their legacy metadata to the DM2E model. Two questionnaires were circulated amongst the project partners in order to collect user feedback forming the basis of the evaluation. The results of this task provided WP2 with valuable suggestions for the refinement of the tools, some of which were already addressed.

Task 1.4 was dedicated to create a test scenario for the scholarly research platform. In close cooperation with scholars working on digital materials, interesting insight was collected about the way scholars envision digital technologies to augment their everyday work. These findings supported WP3 in tailoring their suite of digital tools further to the needs of such a diverse user base.

Task 1.5 dealt with the final content integration. In close cooperation with WP2 and WP3, additional specifications were drafted for content providers to contribute their data to DM2E in the best possible way. This resulted in a significant amount of content integrated to date.

The next steps for WP1 will be to continue with the final content integration. This includes the continuous documentation of the work done in WP1 on the DM2E Wiki and to resolve open issues in order to finalise the task for the existing content providers. Newly joined partners will be given support in mapping, transforming and integrating their data. Last but not least, effort will be taken to acquire new providers to participate in the project.

8 References

Dröge, E., Iwanowa, J., Hennicke, S., & Eckert, K. (2014). DM2E Model V 1.1 Specification (Draft). Revision 1.1. Retrieved from http://wiki.dm2e.eu/wiki/images/8/8f/DM2E_Model_V1.1_Rev1.3_Final_Specification.pdf (13.02.2014).

Goldfarb, D., Morbidoni, C., & Eckert, K. (2013). DM2E Annotatable Content. Technical Specification. Retrieved from http://wiki.dm2e.eu/File:DM2E_annotatable_content_spec.pdf (05.02.2014).

Goldfarb, D., & Ritze, D. (2013). Recommendations for the RDF representation of DM2E metadata. Technical Specification. Retrieved from http://wiki.dm2e.eu/File:Dm2e_mapping_recommendations.pdf (05.02.2014).

9 Appendix

9.1 DM2E – MINT Questionnaire WP1

Distribution

Version	Date of sending	Name
0.1	23.05.2013	Julia Iwanowa (UBER), Kai Eckert (UMA), Kilian Schmidtner (SBB), Nikolaos Simou (NTUA), Nasos Drosopoulos (NTUA)
0.2	24.05.2013	Julia Iwanowa (UBER), Kai Eckert (UMA), Kilian Schmidtner (SBB), Nikolaos Simou (NTUA), Nasos Drosopoulos (NTUA)
1.0	28.05.2013	WP1 content providers

Revisions

Version	Status	Author	Date	Changes
0.1	Draft	Klaus Thoden (MPIWG) Doron Goldfarb (ONB)	23.05.2013	Initial version
0.2	Draft	Klaus Thoden (MPIWG) Doron Goldfarb (ONB)	24.05.2013	Additions according to Kilian Schmidtner (SBB) and Nasos Drosopoulos (NTUA)
1.0	Final	Doron Goldfarb (ONB)	28.05.2013	Introductory text

Organisation

Name of the organisation	
Address	
Zip/City	
Country	

Overall Purpose of Evaluation

Success criteria for MINT (Requirement from DoW Task 1.3):

Level of completeness of mapping of provider-specific metadata format to the DM2E model(see D1.1 "Requirements Report") achievable through MINT

Task Definition

This questionnaire should document content provider experiences with mapping metadata transformations with MINT http://mint-projects.image.ntua.gr/dm2e/Login.action.

It consists of two main sections:

1) General Information

   Information about the user, experience with metadata mapping, etc.

2) MINT Tasks

   Questions related to the MINT evaluation

In order to maintain a consistent evaluation environment, the evaluation process in section 2 is organised in four tasks:

- Import XML – File
- Create Mapping using Mapping GUI
- Validation
- Export Mapping to XSLT

Each task is introduced by instructions that should be followed as close as possible. In case of questions please contact Klaus Thoden (kthoden@mpiwg-berlin.mpg.de) or Doron Goldfarb (doron.goldfarb@onb.ac.at).

1) General Information

A) Information about User

1) Position of user:

2) Please estimate your XSLT knowledge:

3) Please name data transformation tools/methodologies that you have used besides XSLT:

4) Please name other mapping tools that you have used besides MINT:

5) Please state other aggregation activities/workflows you have participated in:

6) How long have you been using MINT?

B) Type of OS / Browser Environment

1) Please state the operating system of your machine (Windows, Linux, Mac etc.):

2) Please name the Web Browser (+Version) that you use for MINT:

C) Information about Metadata

1) Please describe the format of your metadata (You can also point to Deliverable 1.1):

2) Describe your level of knowledge about the input schema (Data model, field names etc):

3) Do the metadata need any pre-processing before they can be used in MINT?

2) MINT Tasks

Guidelines

- Please answer the questions in the following sections A-D as detailed as possible in the form of free text. Please note any positive / negative experiences within the described steps.
- Refer to the MINT documentation under https://doc.dm2e.hu-berlin.de/redmine/attachments/331/MINT%202%20%E2%80%93%20USER%20MANUALv2.docx and use the online help within MINT.

A) Import XML - File

Instructions: Upload an XML file with multiple item records using the "Local upload" feature in the "Import" pane and follow the required steps in "Define Items" for your dataset.

1) Were you able to successfully upload your XML file?

2) In the "Define Items" step, were you able to define the relevant items by dragging and dropping fields from the source element tree?

3) Was the provided documentation sufficient to perform these steps?

4) Please indicate your general impression of the data preparation functionality:
   a) Ease of use/access
   b) Arrangement of elements

B) Create Mapping

Instructions: Create a new Mapping for your recently uploaded dataset. Use the provided "**DM2E v1.0 Fixed Ranges**" schema. Make sure that the "Enable automatic mappings" tick box is **disabled**. Start editing your mapping with MINT's mapping-GUI and try to transform as many of your metadata fields as possible to the DM2E target schema. Use the DM2E model specification from https://doc.dm2e.hu-berlin.de/redmine/attachments/374/DM2E_Model_V1.0_Specification.pdf as reference.

1) Please note which of your source data fields you were able to successfully map to their respective target and also explain those causing problems. Try to identify if the problems arise from
   a) MINT is lacking functionality / You were not able to find a way to do it in MINT
   b) Usability issues
   c) Problems with the DM2E target schema

2) What mapping functionality is missing in order to map all necessary source data fields?

3) How did you like the general drag & drop concept of MINT?

4) Which of the advanced mapping functionalities did you use - Please provide your experiences:
    a) Functions
    b) Conditional Mapping
    c) Value Mapping

5) Which of the navigational aids did you use - Please provide your experiences:
    a) Bookmarks
    b) Search for input/target element names
    c) Input value statistics

6) Was the provided documentation sufficient to perform these steps?

7) Please indicate your general impression of the visual mapping functionality:
    a) Ease of use/access
    b) Arrangement of elements

C) Validate your Generated Mapping within MINT

Instructions: Use the "Preview" Button in the mapping view for performing a test validation. You will be directed a separate preview pane where you can select the respective item view by clicking in the text field. Choose "Mapped Item" from the list. Then select the item to be validated from the item list below. It will take a while until the transformation will be finished. Once done you can select the "Mapped Item" view by clicking on the respective tab on the right. There you will be able to inspect the result of the metadata transformation. In case of any validation errors, you will see highlighted lines in the result. Try to solve as many of these errors by refining your mapping and note the errors you were not able to solve.

In case of any validation error messages, are you able to understand them?

1) Please describe your progress with iteratively correcting validation errors using the preview:

2) Was the provided documentation sufficient to perform these steps?

Please indicate your general impression of the preview/validation functionality:
    a) Ease of use/access
    b) Arrangement of elements

D) Export Mapping to XSLT

Instructions: Once you are done with your mapping follow the instructions in the user manual to export the mapping to an XSLT file.

1) Was the provided documentation sufficient to perform these steps?

2) If you are familiar with XSLT, inspect the generated code and note any observations - Is there anything that you would have solved differently if you had to code the transformation manually?

9.2 DM2E - OmNom Questionnaire WP1

Distribution

Version	Date of sending	Name
1.0	28.11.2013	WP1 content providers

Revisions

Version	Status	Author	Date	Changes
0.1	Draft	Doron Goldfarb (ONB)	22.11.2013	Initial version
1.0	Final	Doron Goldfarb (ONB)	27.11.2013	

Overall Purpose of Evaluation

Success criteria for OmNom (Requirement from DoW Task 1.3):

This tool should enable content providers to use existing XSLT mappings (created in MINT or otherwise) to transform their legacy metadata to the DM2E model and publish the resulting RDF data on the DM2E triplestore.

Task Definition

This questionnaire is intended to collect content providers' experience with transforming their legacy metadata to the DM2E model using the OmNom data ingestion platform.
http://omnom-worker.dm2e.hu-berlin.de/

In order to maintain a consistent evaluation environment, the evaluation process is organised in six tasks:

- User registration
- Upload & manage files
- Create workflows
- Create configurations
- Start and monitor transformation jobs
- Preview published results in Pubby

A guideline for performing these steps can be found at https://github.com/DM2E/dm2e-gui/raw/master/doc/OmNom_WalkThrough_v2.pdf

In case of questions please contact Doron Goldfarb (doron.goldfarb@onb.ac.at).

A) User Registration

Instructions: When you first start OmNom at http://omnom.hu-berlin.de you will have to register a new user. Please perform the necessary steps as described in the tutorial.

1) Were you successfully able to register a new user?

2) Did you encounter any specific issues with the registration? Please describe them, if applicable.

3) Please give suggestions how the User Registration could be enhanced.

B) Upload File & File Manager

Instructions: Continue with the tutorial to upload your files. Please upload one xml file with legacy metadata and one file with an appropriate XSL transformation.

1) Please describe your experience with the file upload procedure and if you encountered any problems with the interface.

2) In the File Manager, experiment with the various filter options for reducing the amount of displayed items, do also use the „My Stuff" filter on top. Please describe what you did and if you had any problems with the intended operations.

3) If you have created any XSL transformations with MINT, try to access them through the File Manager. Were you successfully able to locate your files?

4) Please give suggestions how the File Upload & the File Manager could be enhanced.

C) Create Workflow / Workflow List

Instructions: Follow the tutorial steps to create a workflow for the XSL transformation and the subsequent publishing of your uploaded legacy metadata to the DM2E RDF data model.

1) Please describe your impression of the GUI architecture with respect to the placement of the various "functional areas" present there.

2) Please describe your experience with the drag & drop functionality for creating workflows.

3) Use the "Visualize" function to get a graphical overview of your workflow.

   a) Do you understand the various modules featured in the diagram in relation to their representation in the workflow creation GUI?

   b) Do you understand the role of the different connectors

4) Were you successfully able to create the intended workflow? If not, please describe your specific problems in detail.

5) Please give suggestions how the GUI for creating workflows could be enhanced.

D) Create Configuration

Instructions: After you have created a workflow, please continue with the tutorial steps to set up an appropriate configuration.

1) Do you understand the relationship between workflows and configurations? If no, please explain further.

2) Use the filter options to narrow down the list of available input files. Did the filters help you with identifying the input files for your workflow configuration?

3) Please describe your experience with the drag & drop functionality in the create config UI

4) Did you encounter any problems with identifying those parameters which expected (drag & drop) filenames as input and those which expected strings as input?

5) Were you successfully able to create a configuration for your workflow? If not, please describe your specific problems in detail.

6) Please give suggestions how the GUI for creating configurations could be enhanced.

E) Job View

Instructions: After finishing a configuration for your workflow, start a job by clicking the "Run" button. As described in the tutorial, you will be directed to the job progress display where you can check how your job proceeds. You will have to constantly click the "Refresh" Button in order to see the job's progress!

1) Please describe your impressions of the "Job progress display". Do you understand the meaning of the different informative sections?

2) Did your job finish successfully or did it fail? If it failed, did you get the information why the job has failed?

3) Please give suggestions how the job progress display could be enhanced.

F) Pubby

Instructions: If your job has finished successfully, please click on the "Show in Pubby" button that should now be visible within the job progress display. You will be redirected to an html view of all the triples that were created by your job. You will first be presented with an entry page showing information about the just ingested dataset, including a list of all the providedCHOs that were created in the process. Start exploring your dataset using this list.

1) For each examined ProvidedCHO, were you able to find all the related classes (Aggregation, WebResources, Agents, Places, TimeSpans etc) ?

2) Do the triples that you created look as expected? If anything differs from "how it should be", please explain it in detail and also try to give an explanation (eg XSLT problem, other problem, etc.)

3) Please express your opinion about using Pubby as a "Preview" for the created triples.

4) Please give suggestions how the Pubby Linked Data preview could be enhanced.

9.3 Annotatable Content Specification

This document can be downloaded from the DM2E Wiki at
http://wiki.dm2e.eu/File:DM2E_annotatable_content_spec.pdf

9.4 Recommendations for the RDF representation of DM2E metadata

This document can be downloaded from the DM2E Wiki at
http://wiki.dm2e.eu/File:Dm2e_mapping_recommendations.pdf