DELIVERABLE

Project Acronym: **Europeana Cloud**

Grant Agreement number: **325091**

Project Title: **Europeana Cloud: Unlocking Europe's Research via The Cloud**

D3.1 Personas, scenarios and use cases

Revision: 2.20

Authors:

     Erik Duval, KU Leuven
     Gonzalo Parra, KU Leuven
     Marnix van Berchum, DANS-KNAW / Utrecht University
     Hein van den Berg, Stichting VU-VUMC
     Anja Jentzsch, Open Knowledge Foundation Deutschland
     Nikos Manouselis, Ariadne Foundation
     Giannis Stoitsis, Ariadne Foundation
     Akis Tsekouras, Ariadne Foundation
     Sven Charleer, KU Leuven
     Maritina Stavrakaki, Ariadne Foundation

Revision History

Revision	Date	Author	Organisation	Description
0.1	17.07.13	Erik Duval	KU Leuven	First integration of working documents
0.2	19.07.13	H. vd Berg	VU	Adding background, rewriting, editing
0.3	20.07.13	Erik Duval	KU Leuven	Editing, adding context, adding section on next phase, restructuring appendix.
0.4	21.07.13	H. vd Berg	VU	Minor comments, additions.
0.5	22.07.13	N. Manouselis, G. Stoitsis	ARIADNE	Added context and section on process, restructured appendix, reformatted text
0.6	22.07.13	G. Parra	KU Leuven	Minor edits on scenarios 3.5 and 3.6.
0.7	22.07.13	H. vd Berg	VU	Substantial change to scenario "Arianna, Hein,the ARIADNE Finder and Pundit"; suggestions to rewrite certain portions of section 5 (in comments); general comments and minor edits.
0.8	22.07.13	G. Parra	KU Leuven	Final draft for internal review.
0.9	27.07.13	H. vd Berg	VU	Rewritten scenarios 3.9 and 3.10.
1.0	28.07.13	G. Parra	KU Leuven	Editing and rewriting of content based on the external reviewers suggestions.
1.1	29.07.13	Anja Jentzsch	OKF DE	Readability improvements. Spell checking. Scenarios rewritten.
1.2	29.07.13	G. Parra	KU Leuven	Final version.
2.0	14.07.14	G. Parra	KU Leuven	Outline draft to include work from the year 2 cycle.
2.1	27.07.14	H. vd Berg	VU	Added section on methodology for year 2 cycle.
2.2	28.07.14	A. Jentzsch	OKF DE	Reviewed Hein's section. Added year 2 part on conclusion. Reorganized tool section and added some tools.
2.3	28.07.14	G. Parra	KU Leuven	Integration of scenarios and personas. Editing of sections to reflect the work on Year 2.
2.4	30.07.14	H. vd Berg	VU	Rewriting of sections to reflect the work on Year 2. Finished conclusion, and section on methodology and activities for year 2 cycle.
2.5	31.07.14	A. Tsekouras	ARIADNE	Updated appendix with list of tools, small modifications in methodology from Y1.
2.6	31.07.14	G. Parra	KU Leuven	Final draft version.
2.7	05.08.14	E. Duval	KU Leuven	Finalisation.
2.8	14.01.16	A. Jentzsch	OKF DE	Y3 update
2.9	15.01.16	M. Stavrakaki	ARIADNE	Y3 update
2.10	15.01.16	M. Stavrakaki	ARIADNE	Y3 update

2.11	15.01.16	A. Jentzsch	OKF DE	Background update
2.12	25.01.16	M. Stavrakaki	ARIADNE	Y3 update and revision
2.13	25.01.16	S. Charleer	KU Leuven	Y3 appendix tools update and revision
2.14	25.01.16	H. van den Berg	VU	Rewriting, editing
2.15	29.01.16	M. Stavrakaki	ARIADNE	Rewriting, editing
2.16	29.01.16	S. Charleer	KU Leuven	Rewriting, editing
2.17	29.01.16	H. van den Berg	VU	Rewriting, editing
2.18	30.01.16	S. Charleer	KU Leuven	Rewriting, editing
2.19	30.01.16	M. Stavrakaki	ARIADNE	Rewriting, editing
2.20	31.01.16	A. Jentzsch	OKF DE	Final layout

D3.1 Personas, scenarios and use cases

Executive summary

This document describes the personas, scenarios and use cases on which WP3 focused during the first, second, and third year of the Europeana Cloud project (M1-M6, M13-18, M25-32). This work was undertaken in collaboration with the DM2E project (and more specifically the Wittgenstein archives at the University of Bergen), the Axiom philosophy group at the VU University Amsterdam and the University of Amsterdam (in year 1), a research community of musicologists (in year 2) and a research community of social scientists and agricultural researchers (in year 3).

Through the personas and scenarios, we have identified a set of core problems to address, and classes of tools and services that can solve these problems for our target communities of researchers who work with Europeana content.

In the follow-up phases for the first, second, and third year of Europeana Cloud, we configured, integrated and evaluated specific tools from the classes which were identified. The appendix to this deliverable lists the candidate tools for these phases, and this follow-up work is reported in more detail in Deliverables 3.2 and 3.3.

Table of Contents

1. *Background and Context*

The objectives of WP3 include (p.13 of 38 in the Description of Work):

- Development of services and tools that leverage Europeana content for use by researchers.
- Follow an iterative design process to identify typical personas and scenarios of thematic use of content that the Europeana Cloud tools and services as well as the Europeana Research Platform would be expected to support.

In order to reach these objectives, Task 3.1 focuses on Personas, Scenarios and Use Cases:

> In close alignment with the tasks in WP1 on user needs and existing tools, KU Leuven will lead on developing personas (descriptions of typical researchers that we address with this pro£ject), scenarios and use cases that describe in detail what kind of tool a researcher would typically use, and how that use would fit in his typical workflow. This will result in Deliverable 3.1 [M6, M18, M30]

This document is the final version of Deliverable 3.1. It aims to report on personas and scenarios which have been developed, in order to elaborate and consolidate our thinking on the tools and services which can leverage Europeana content for researchers.

A comprehensive overview of the use of personas and scenarios for this kind of purpose can be found in The Encyclopedia of Human-Computer Interaction[1]. From that source:

> Common understanding is that the persona is a description of a fictitious person […] In the design process, we begin to imagine how the product is to work and look before any sketch is made or any features described. If the design team members have a number of persona descriptions in front of them while designing, the personas will help them maintain the perspective of the users. The moment the designers begin to imagine how a possible product is to be used by a persona, ideas will emerge. Thus, I maintain that the actual purpose of the method is not the persona descriptions, but the ability to imagine the product.

A shorter introduction can be found in this Building and Using Personas and Scenarios presentation[2].

The gist of the idea behind personas and scenarios is that they help us to focus on the intended user of the tools and services, rather than on the technical challenges of how to develop those tools and services. This increases the likelihood that we develop tools and services that are actually useful for and usable by the target user.

WP3's main aim is to develop tools and services that allow researchers to make optimal use of Europeana content. In the first year, this meant a focus on developing tools to benefit researchers in

---

[1] Nielsen, Lene (2013): Personas. In: Soegaard, Mads and Dam, Rikke Friis (eds.). "The Encyclopedia of Human-Computer Interaction, 2nd Ed.". Aarhus, Denmark: The Interaction Design Foundation. Available online at http://www.interaction-design.org/encyclopedia/personas.html

[2] http://www.slideshare.net/pennyhagen/introduction-to-building-and-using-personas-and-scenarios-in-design

philosophy and digital humanities. This was achieved by working closely with two communities of researchers interested in using Europeana content:

- The DM2E project3 is also looking at "new tools and services for the re-use of Europeana data in the Digital Humanities". Of particular interest for us is the Pundit4 a 'semantic annotation tool' that enables the sharing of annotations between researchers. We have discussed (online and face-to-face) collaboration with this project and a community of researchers that focus on the Wittgenstein Archives at the University of Bergen 5.
- The Axiom group of history and philosophy of logic, semantics and axiomatics6 at the VU University of Amsterdam studies 'Semantics and Axiomatics from Bolzano to Tarski against the Background of the Classical Model of Science' (the group has recently moved to the University of Amsterdam). Also with this group, represented directly in Europeana Cloud through the participation of Stichting VU-VUMC, we have collaborated in person and online.

On the basis of this early collaboration, we were able to identify three core problems faced by these researchers:
- (i) problems with navigating and identifying relevant (digital) content and with building corpora;
- (ii) a lack of user-friendly computational tools for conducting fine-grained textual research;
- (iii) a lack of appropriate tools and infrastructure that allow members of research groups to work collaboratively rather than individually.

In our work, we have identified
- (a) search tools to find relevant content, and
- (b) visualisation tools (geographical maps, timelines) that assist in overcoming core problem (i).
- (c) annotation tools (see below), incorporating information extraction methods and Linked Data techniques, that help solve core problem (ii),
- (d) social awareness and discovery tools (see below) that help tackle core problem (iii).

The appendix lists the relevant tools that we have considered. Some of these tools have been further developed, linked to each other, and linked to Europeana content. This work has been extensively described in deliverables 3.2 and 3.3. As such, WP3 provides researchers in the Humanities with tools that assist them with various aspects of their work, ranging from the identification of textual resources to the analysis of these resources. This in turn allows them to fruitfully exploit content provided through Europeana.

In the second year of our work (M13-M18), we have focused on developing personas, scenarios, and tools for researchers working within musicology, specifically with Early Music. In the first phase of the second cycle (M13-M15), our aims were to understand the research practices and workflow of musicologists, and to identify digital tools and content employed by these researchers. To this purpose, we conducted desk research and organized numerous exploratory online meetings with Marnix van Berchum, a musicologist working on musical culture of the 15th and 16th century, and Marian Lefferts, a specialist in digital library content and cultural heritage. Both are involved in Europeana Cloud through activities in WP6 and WP4 respectively.

---

[3] http://dm2e.eu

[4] http://www.thepund.it

[5] see http://wab.uib.no/wab_nachlass.page/

[6] http://axiom.vu.nl/. http://www.axiom.humanities.uva.nl/

In the second phase of the second cycle (M16-M18), members of WP3 and Marnix van Berchum have organized three online meetings with a research community of musicologists interested in using digital tools and Europeana content. This community consists of the following researchers:

- Eliane Fankhauser (Utrecht University; working on late medieval polyphonic music from the Low Countries).
- Peter van Kranenburg (Meertens Institute; working on the Tunes & Tales project[7]).
- Laurent Pugin (Swiss office of RISM[8]; works on Optical Music Recognition (OMR)).
- Reinier de Valk (City University, London; working on the computational extraction of polyphonic structures from sixteenth century lute tabulatures)
- Frans Wiering (Utrecht University; works on Music Information Retrieval (MIR)).

In our first and second online meetings (12 and 14 May 2014), we discussed the workflow of the musicologists, as well as computational tools and content they use. After these meetings, we identified a number of challenges these researchers face when conducting their work[9]. These challenges include:

(i) *Problems involving data creation*. Many musicologists require transcriptions of (historical) musical sources to modern scores in order to conduct their research. Although there are computational tools that allow users to create transcriptions and critical editions of musical sources, most transcriptions are done completely by hand. This is a laborious and time-consuming process.

(ii) *Lack of digital corpora of music scores*. Although there exist a number of initiatives aiming to create standardized digital and symbolic representations of music scores (e.g., http://music-encoding.org/home), as well as music libraries offering access to symbolic and encoded representations of music scores, many musicologists still lack appropriate digital corpora of scores.

(iii) *Information exchange and linking of data*. Musicologists often need to match and compare the original (historical) sources of music or music scores to existing transcriptions of these sources. For example, databases such as Europeana provide images of (historical) music sources (typically in PDF format). These databases typically do not, however, provide digital, symbolic, and encoded representations of music sources. Such representations are provided by other online databases (e.g., http://josquin.ccarh.org/). Linking these different types of data and data sources is a non-trivial task.

(iv) *Retrieve and analyse contextual information.* For contextualizing the music that researchers work on, information has to be gathered from multiple other sources, besides the representations of the original sources or transcribed scores. For example, apart from music sources and scores, musicologists require access to general bibliographical and historical information. It is difficult to process, compare, and analyse these different and heterogeneous information sources. These activities are, however, core to the work of many musicologists

On the basis of this interaction, we identified tools that we thought would aid musicologists in their research, and we constructed personas and scenarios that describe how these tools could fit their workflow. The descriptions of these personas and scenarios have been evaluated by the research community of musicologists in a third online meeting (19 June 2014), and have been rewritten and improved on the basis of this evaluation.

---

[7] http://www.ehumanities.nl/computational-humanities/tunes-tales/

[8] http://www.rism.info/

[9] In our work, we have significantly profited from Frans's Wiering's work on user needs and challenges in Digital Musicology. See: http://www.staff.science.uu.nl/~wieri103/presentations/WieringLondonDigitalMusicLabFinal.pdf.

Through our collaboration with the musicologists, we have identified the following kinds of tools as particularly relevant to their work:

(i) *Optical Musical Recognition tools* that allow users to create transcriptions and critical editions of musical sources, and *music notation software* that allows users to create and edit sheet music. These tools assist in overcoming challenges (i) and (ii) above.

(ii) *Search tools, digital repositories and archives, and visualization tools* that allow users to identify and aggregate (metadata of) digital and symbolic representations of music, and that can be used to link these representations to image files of music sources. These tools assist in overcoming challenges (ii) and (iii).

(iii) *Annotation tools, music analysis tools, and classification tools* that allow musicologists to analyse, process, classify, and enrich their data. These tools assist in overcoming challenge (iv).

Several of these tools have been further developed, linked to each other, and linked to Europeana content in the WP3 activities in the second part of the second year of Europeana Cloud (M19-M24).

In the third year of work, our target community of researchers were those working with the Europeana Newspapers Archive. By organizing two workshops with researchers, we identified a large list of problems and requests to better support their workflow. Out of this long list, we prioritised the following issues:

(i) difficulties navigating the online Newspaper Archive
(ii) poor usability
(iii) limited facets for searching
(iv) automatic suggestions of relevant articles to the one the user is accessing
(v) timeline per newspaper

In order to help these researchers, we developed the Newspaper Exploration Environment (see Appendix, 0) , which offers:

- a more visual approach to exploration of the archive
- a recommender system, providing recommendations of newspapers based on the results the researcher is exploring

Furthermore, we developed a data mining tool, eCloudDM, that extracts named entities from the newspaper articles and defines topic tags for the newspaper articles

Additionally, our third year included a focus on developing tools to support agricultural researchers investigating societal or cultural topics, and who are using social science methods and tools. To this purpose, we have worked closely with agricultural researchers at the Agriculture Economics and Policy Research Institute (AGRERI), in order to understand their data requirements and how they could make use of digitised resources and archives available through Europeana to help them answer research questions.

We did this by setting up interviews and an online survey in order to collect demographics on the targeted audience, and aspects on each persona. We also organised meetings with the users to understand their requirements, in order to be able to validate the search tools which were later developed.

On the basis of this collaboration, we identified two core problems faced by these researchers:

(i) problems with navigating and identifying relevant (digital) content from diverse sources;
(ii) lack of tools (like facets and filters) that will facilitate the discovery of the resources needed in specific thematics;

Based on the results of the interviews and online survey, we designed preliminary wireframes for a research information discovery micro-site within the AGRERI website. After the initial feedback from the users, the development of the AGRERI Discovery micro-site began, bringing together content from relevant Europeana resources and other sources. The deployment of the updated discovery page underwent a versioning process, so after the first prototype (alpha version) an internal feedback session took place. Based on this feedback, the beta version of the website[10] was built and sent to the AGRERI team and selected representatives for evaluation. At the same time, a hands-on workshop took place in order to get comments and input from the people, while they interacted with the discovery page. The final version of the discovery micro-site within the operational AGRERI site[11] was then deployed.

---

[10] http://dev.agreri.agroknow.gr
[11] http://www.agreri.gr

2. *Structure of the document*

In Sections 3-6, we list the personas and scenarios that we have developed in dialogue with our target user communities of researchers from the first, second, and third year, respectively. Personas and scenarios are listed in no particular order and there is quite a bit of overlap in the scenarios. This is useful, as it helped us to identify the common issues mentioned in the previous section. Those common issues informed the following work of WP3: the configuration, integration, and evaluation of specific tools. We conclude the sections by presenting the candidate tools used in the final phase of each year. In Section 7, we further analyse the process to follow in order to develop and adapt tools. Finally, Section 8 concludes this document and the appendices list the candidate tools analysed for the different researchers in the first, second, and third year of the Europeana Cloud project.

3. *Year 1: History of Philosophical Ideas*

Personas and Scenarios

Anita and Muse

Persona

Anita is a post-doctoral researcher on Technology-Enhanced Learning in a computer science department of a European university. She is collaborating with colleagues from different European research units, in a Framework 7 EU project. She regularly travels to project meetings, workshops, and conferences throughout the EU. Anita is pretty tech savvy and works in a department that is well equipped with smartphones, laptops, large screens, and even a few multitouch tabletops.

Scenario

Anita is assisting her supervisor with the preparation of a new EU proposal. There are some partners that her team regularly collaborates with and they decide to invite three of those partners to join the proposal preparation. They then turn to the multitouch table and use Muse to explore whom else they have worked with, and, more importantly, whom their colleagues have worked with before. Together with her supervisor, she thus discovers a team that focuses on Computer-Supported Collaborative Learning (CSCL) in Norway. Anita remembers that she met one of members of that team earlier in the year. Through the visualisation, Anita and her supervisor discover that the Norwegian team has a close collaboration with the team in Lausanne that they had already invited for the new project. They thus decide to ask the Lausanne team about the Norwegian partners first. Moreover, her supervisor will travel to Madrid next week for a conference. With the tabletop visualization, they remember that they had previously worked with two research units in Madrid: Anita's supervisor sends an email to both of them to set up a meeting to discuss the new project proposal.

Anna and Annotateit

Persona

Anna is a research associate at the department of philosophy in Cambridge, UK. She is preparing a thesis on Shakespeare's eighteenth-century reception in England and France. She is interested to explore and apply ICT-based methods and tools to enable new research questions and is eager to exchange knowledge, expertise, methodologies, and practices across domains and disciplines. She is an active member of the Open Humanities Working Group of the Open Knowledge Foundation that meets online weekly to discuss Shakespeare's work and to explore new ICT-based research practices and tools for the digital humanities.

Scenario

For an essay on Shakespeare's thoughts on urbanism and the city Anna needs to scan a lot of documents against the topics of interest for her essay and is therefore also interested in the thoughts of like-minded colleagues. She and her colleagues have used the Annotateit tool to share annotations on online documents before, and she now invites her colleagues to help her scan through a series of relevant texts for the topics of interest for her essay and share their annotations, comments and tags with her.

Paul and Pundit

Persona

Paul is a post-doctoral researcher on philosophy at the Wittgenstein Archives at the University of Bergen (WAB) in Norway. The focus of his research is on the linguistic aspects of Wittgenstein's work. WAB has published 5,000 pages of the Wittgenstein Nachlass Open Access on the Web and developed a Wittgenstein domain ontology for these texts. WAB is part of a Framework 7 EU research project, called Digital Manuscripts to Europeana (DM2E), where Paul is involved in testing new technologies for collaborative work with semantic annotations online.

Scenario

Reading Wittgenstein's work using Pundit allows Paul not only to annotate the texts, but also to look up various concepts and persons occurring in the texts.

Pundit allows Paul to add and share simple comments as well as semantic links to the Web of Data (as Freebase or DBpedia) for fine granular cross-references and citations using the WAB's Wittgenstein domain ontologies. Its named-entity recognition features allow for disambiguous linking of real world entities mentioned in texts to their entries in Linked Data sets and thus deriving more information on this entity as well as the possibility to further traverse the Web for more fine-grained information.

Furthermore the annotations help finding other relevant texts on a specific topic or concept in Wittgenstein's work.

Pam and GlamMap

Persona

Pam is a 25 year old PhD-student at the department of philosophy at the University of Utrecht, the Netherlands. She has received her Master's degree at the University of Groningen, where she has written a thesis on Kant's use of the concept 'transcendental logic' in the second edition of the *Kritik der reinen Vernunft* (1787). Pam has enjoyed a classical training in the history of philosophy, which means that she has learned to provide highly detailed analyses of a very small number of texts. At Utrecht, Pam is a member of an NWO Vidi-project in which she is supposed to investigate how the concept of 'logic' is used within German philosophy books published around 1800.

Pam is ambitious, hard working, and desperately wishes to write a good dissertation. Like many of her colleagues at the philosophy department, she is an avid user of technology. She is a member of various social media sites, she owns a smartphone and a laptop and she is constantly online. Pam is eager to use and explore ICT-based methods to push further her research and has recently developed an interest in digital humanities. Nevertheless, her knowledge of ICT-based methods and tools is quite limited. When conducting her research, Pam mostly uses Google Books and Google Scholar.

Scenario

Having just started research for her dissertation, Pam is overwhelmed by the number and complexity of the philosophical and scientific texts that she needs to read and understand. Although her knowledge of the works of Kant is adequate, she lacks an overview of the works on logic published around 1800 that are relevant for her research. In order to identify publications that are important for her research, Pam uses GlamMap. Using this tool, she obtains an interactive geo-spatial visualization of bibliographic metadata of logic books published in Europe between 1795 and 1805. She uses the tool to identify and order various (types of) books that she needs to study for her dissertation. She identifies publications on 'transcendental logic' that have been published in Jena and Leipzig, she

identifies traditional books on term logic published in Halle, and she identifies works on psychology, anthropology, and logic published in Heidelberg. Through exploring the visualization, Pam is also able to quickly distinguish between the relatively small number of books published by famous philosophers in this period, and the relatively large number of textbooks and lecture scripts concerning logic. Since these latter books are easier to read and have also been little studied, Pam decides to start her research by focusing on these books.

John and TiNYARM

Persona

John is a PhD candidate in the doctoral school of human sciences at the Vrije Universiteit Brussel. After finishing his master in arts, he just started to work pursuing a PhD degree in philosophy and history of education. During his master thesis work, he got familiar with tools used by researchers, such as: references managers (Mendeley and BibTeX) and digital libraries.

Scenario

John has a clear idea of his dissertation topic and started with the literature review about the subject. As he is new in the research community and he doesn't know which colleagues from other European universities may be working on the same topic. Colleagues from his research team suggested that he use TiNYARM. His research group employs this tool to share publications read by them, to know what other research groups are reading, and his advisor uses it to suggest articles to his PhD students. As soon as he joined the tool, John saw a recent suggestion from Anna (a researcher from Germany) to one of his colleagues; this article was quite interesting for his research. He also opened the TiNYARM profile page of his advisor and skimmed the different papers his promoter read over the past months. John found some articles that were relevant for his research. Looking at his group activity, John got an idea of what everybody is interested in, based on the different papers they read or skimmed.

John and More!

Persona

John is a PhD candidate in the doctoral school of human sciences in the Vrije Universiteit Brussel. He is doing a PhD in philosophy and history of education. During his master thesis work and his first year in his PhD, he got familiar with tools used by researchers, such as: references managers (Mendeley and Bibtex), digital libraries, SlideShare, and academic social networks (LinkedIn, Academia.edu, and ResearchGate).

Scenario

After his first PhD year, John and his advisor get a paper accepted in an international conference. This will be his first time attending a conference. On the first day, John gets the conference programme and tries to figure out which sessions to attend. While creating the list of sessions, he notices that a QR code is available in the conference programme. Using his smartphone, John scans this code and is redirected to the More! mobile web application showing different links to social networks that the presenter is actively participating in. After seeing the presenter profile and checking his publications and recent tweets, John gets a good idea of what the presenter worked on before and which are his current research interests. John decides to meet the presenter later and sends him a mail using More!. After using this application for a few presentations, he is a bit frustrated with some of the other sessions that do not provide a More! profile of the presenter and that he actually has to 'google' them for some minutes in order to obtain the same information as he gets from More!.

Hein and ARIADNE Finder + Textus

Persona

Hein is a post-doctoral researcher working on the history of philosophy, the history of biology and their interplay. In particular, he is interested in 18th century philosophical ideals of science and their influence on 18th century life sciences. His research has been recently focused on debates in philosophy of history and historiography, where he investigates and needs access to philosophical document and digital media corpora and their visualisations.

Scenario for ARIADNE Finder

Hein participates in a group of post-doctoral researchers and academics interested into the history and philosophy of logic, semantics, and axiomatics. This group explores the concept of semantics and axiomatics in the history of philosophy, exploring ideas coming from philosophers like Bolzano and Tarski against classical models of science. Hein has already been using some interesting visualisation tools for relevant document and literature corpora that are working over a bibliographic metadata aggregator/network that his group has set up. He would like to enhance the web site of this group with a search interface that will allow the group to discover and navigate relevant digital resources that are coming from other large aggregators, such as: material on the history of philosophy collected by the Europeana aggregator,[12] relevant literature that has been indexed and aggregated from various European national libraries through The European Library,[13] as well as presentations, notes, slides, and other relevant educational resources that come from a global network of educational repositories called GLOBE.[14]

Hein would use the ARIADNE Finder technology to carry out the following steps:

- Define the metadata facets that his group would like to use in order to search and browse through the various digital resources on the history of philosophy, and define the desired properties of a local metadata index that will collect periodically relevant material that can be found in the various aggregators.
- Design the information architecture of the content search pages using a simple pre-existing template that will allow him to integrate the search functionalities into the look and feel of the group's web site.
- Run the tool that will set up and create the local metadata index (also with a mechanism for periodic re-harvesting/ingestion and re-indexing from the various aggregators) running over a cloud-hosted virtual machine through a simple web-based interface. The tool will also allow him to select which thematic keywords or specific collections he would like to include when retrieving descriptions of content from the three external aggregators.
- Get as an outcome of the tool a simple HTML code that he can easily embed into his group's web site in order to get the new ARIADNE-powered search pages easily set up and running with no further technical effort.

Scenario for Textus

Hein would like to incorporate in the web site of the group a functionality that would allow the post-doctoral researchers to select digital documents discovered through the ARIADNE Finder and to carry out further discussions and analyses using Web-based open annotation features - in a similar

---

[12] http://www.europeana.eu
[13] http://www.theeuropeanlibrary.org
[14] http://globe-info.org

way that the OpenPhilosophy.org group is working with public domain philosophy texts.[15] For this purpose he incorporates the Textus tool into the web site in a way that allows the group to:

- Select/bookmark some interesting digital resource coming from one of the aggregators like Europeana or The European Library and add it to a shared common space.
- If its license scheme allows it, saving a local copy of the digital resource and opening it in the Textus viewer.
- Providing annotations on various parts within the text under analysis and saving them into the shared space together with the digital resource.
- Generate as a separate new document all the produced annotations in a digital format, including a reference to the original document into its metadata description - saving it in the group's shared document repository.
- Enhance the metadata record ingested from an external source with the generated annotations of the group and saving the new version of the record into the group's shared document repository (either as a new, enriched version of the original authority record or as a new metadata record that stores only annotations and other user actions using some scheme like CAM).

Pam and Textus

Persona

Pam is a 29 years old postdoc working at the department of history and philosophy of science at the University of Cambridge (UK). She has a Master's degree in Computer Science and in Philosophy, both obtained at the University of Edinburgh (Scotland). In her doctoral dissertation Pam studied the philosophy of mathematics of the eighteenth-century mathematician Leonard Euler. At Cambridge, Pam is part of an ERC-Starting Grant Project that aims to provide a comprehensive overview of the history of 18th and 19th century philosophy of mathematics. For her research, it is essential that Pam collaborates intensively with philosophers, historians, and mathematicians from different European research units. She and her partners require easy access to historical documents and wish to collaboratively study such documents. Apart from doing her research, Pam also teaches an annual course on the history of mathematics at Cambridge.

Pam knows a lot about existing ICT tools and methods that can help with her research. Thanks to her efforts, her research group and various European partners use project management software (Basecamp) to coordinate projects, they use reference managers (Zotero) to create shared bibliographies, and they compile large databases using digital resources obtained via Europeana, Google Books, archive.org, and other online repositories.

Scenario

Pam and one of her colleagues working at the University of Duisburg-Essen (Germany) decide to write a joint paper on the reception of Euler's mathematical works at the end of the 18th century, focusing in particular on the work of Immanuel Kant. They decide to use TEXTUS in order to share, study, and collaboratively annotate historical documents that are relevant for their paper. Through their contacts at Duisburg-Essen, they obtain plain-text files of Kant's published writings.[16] After uploading these writings, they use TEXTUS to identify, annotate and discuss passages within the Kantian corpus that concern mathematical topics. Through sharing their annotations with other

---

[15] http://beta.openphilosophy.org
[16] http://www.korpora.org/kant/

project partners and interested researchers, they further obtain valuable information and references that help them in writing their paper.

Apart from using TEXTUS for research purposes, Pam also uses the tool while teaching her undergraduate course. Students following her course are supposed to study historical texts by collaboratively commenting on these texts. They also have to correct and grade each others annotations. Pam notices that by working in this manner her students find it easier to read and understand scientific and philosophical texts.

Jeroen, ARIADNE Finder, GlamMap, Pundit

Persona

Jeroen is a 25 years old PhD student at the department of philosophy at the Vrije Universiteit Amsterdam. He has Master's degrees in philosophy and intellectual history, both obtained at the University of Groningen. In Amsterdam, Jeroen is a member of an NWO VICI Project that aims to rewrite the history of eighteenth-century philosophy of biology. Jeroen and his colleagues need to identify many unknown historical documents and they wish to collaboratively study such documents. The members of Jeroen's research group are pretty tech savvy. They use Whatsapp to communicate, they use Document management & Intranet software (Papyrs) to coordinate projects and write papers, they use reference managers (Zotero) to create shared bibliographies, and they compile large databases using digital resources obtained via Europeana, Google Books, archive.org and other online repositories. Jeroen is keen to use ICT tools while conducting his research and he hopes that they enable him to work in a more efficient manner.

Scenarios

Compiling and organizing primary sources (step 1)

Jeroen and several other PhD students have to write a paper on methodological debates within eighteenth-century natural history. In order to write this paper, they need to compile a database of relevant primary sources i.e., books on natural history published in the eighteenth-century. When compiling his database of primary sources, Jeroen uses the ARIADNE Finder, a search interface that allows him to discover and navigate digital resources that have been collected by the Europeana aggregator, The European Library, digital resources compiled by Google Books, and resources contained within other repositories.

For his paper, Jeroen decides to search for the term "Natural History" from 1700-1800 within the European Library. He obtains over 9000 metadata records of books on natural history published in the eighteenth-century. Using the ARIADNE Finder, Jeroen exports these records to Zotero and compiles a large database of primary sources.

Navigating the primary sources (step 2)

It is difficult to navigate 9000 works. After having obtained his 9000 metadata records, Jeroen downloads these records and feeds them to GlamMap in order to obtain an interactive geo-spatial visualization of these bibliographic metadata. He uses the tool to navigate the metadata and he orders books on the basis of time and place of publication. By doing this he can identify various (types of) books that he needs to study for his paper. After ordering his books, he decides to write a paper on the species concepts of Linnaeus and Buffon.

Identifying secondary sources (step 3)

Having identified his primary sources, Jeroen needs to identify relevant secondary sources, i.e., works on the history of biology published in the twentieth-century. He again uses the ARIADNE Finder in order to aggregate metadata of contemporary (20th century) books discussing Buffon and Linneaus. He collects metadata from the Europeana aggregator, the European Library, Pubmed, Worldcat, Google Scholar, and other possible sources. He selects all the relevant literature and exports his selections to Zotero. He has now created his database of secondary sources.

Navigating the secondary sources (step 4)

Having obtained metadata records of his secondary sources, Jeroen visualizes these records using GlamMap (the procedure is identical to step 2). He obtains an interactive geo-spatial visualization of his bibliographic data. He now has two visualizations: one of his primary sources and of his secondary sources. Jeroen is able to compare both visualizations and is able to find several interesting relations between them.

Studying the relevant literature (step 5)

After identifying the specific literature that is to be studied, Jeroen and his colleagues decide to use Pundit to collaboratively study a number of selected texts. They obtain texts of Buffon and Linnaeus and semantically annotate them by relating sections of these documents to Linked Data resources such as DBpedia and Wikidata. They use named-entity recognition in order to identify names and concepts occurring in the text, and link these items to important online resources. Thus, for example, fragments of Buffon's Natural History are related to relevant Wikipedia articles, to more technical articles on the history of biology in the Stanford Encyclopedia of Philosophy[17], and other educational resources.[18] In this manner, Jeroen and his colleagues can easily navigate and study numerous quite technical texts.

Arianna, Hein, the ARIADNE Finder and Pundit

Persona

Hein is a post-doctoral researcher working on the history of philosophy, the history of biology and their interplay. He is interested in 18th century philosophical ideals of science and their influence on the 18th century life sciences. Arianna is professor of philosophy of language and conducts research on the history of logic, metaphysics and digital humanities. Arianna and Hein are members of a research team investigating the history of logic and science. Both use a variety of digital tools to enhance their research, such as reference mangers (Zotero), WhatsApp, and various other tools.

Scenario

Arianna and Hein have, for research purposes, obtained 7,000 bibliographic records from Wilhelm Risse's Bibliographica Logica.[19] As a result, they have obtained bibliographic metadata of over 7,000 books concerning logic published between 1700 and 1940. Both are currently attempting to obtain a comparable database of logic books by aggregating metadata from Worldcat, Europeana, The European library, and other repositories.

Arianna and Hein wish to write a paper on logic books that have been little studied by researchers on the history of logic. In order to write this paper, they need to compare the historical data from Risse (or similar data obtained via other large aggregators) to bibliographic data and references contained

---

[17] http://plato.stanford.edu/entries/evolution/
[18] http://www.ucmp.berkeley.edu/help/topic/history.html
[19] http://www.olms.de/

in philosophy repositories (e.g., Philpapers),[20], entries on history of logic published within Wikipedia, entries in the Stanford Encyclopedia of Philosophy,[21] the Internet encyclopedia of Philosophy,[22] and possibly results provided by Google Scholar and Worldcat. Arianna and Hein use the ARIADNE Finder to obtain bibliographic metadata of thousands of logic books published from the 18th to the 20th data. Using Pundit, they link this data (e.g., name of author) to resources found within Wikipedia, the Stanford Encyclopedia of Philosophy, the Internet Encyclopedia of Philosophy, Wikidata and Linked Open Data sets. In this manner, they are able to reconstruct which logic books are often mentioned in popular resources and which logic books are rarely mentioned. This allows them to give a reasoned estimate on which logic books have been little studied by contemporary researchers.

John and Researchr

Persona
John is a PhD candidate in the doctoral school of Human Sciences at the Vrije Universiteit Brussel. After finishing his master in arts, he just started to work pursuing a PhD degree in philosophy and history of education. During his master thesis work, he got familiar with tools used by researchers, such as: references managers (Mendeley and Bibtex) and digital libraries.

Scenario
John has to start doing his literature review about the subject. He has downloaded a lot of PDFs (usually with generic or unidentifiable names) and started to take notes using different software and on papers. After realizing the challenges of organizing these, he got really interested in how to improve the management all the PDFs, citation metadata, clippings, notes, and ideas. He started to use Researchr, which was an aggregation of different tools such as: BibDesk (to manage citations), Skim, (to make comments and highlight text), DokuWiki (to manage notes), and scripts; and use it as his personal open publications management system. John uses the uses different scripts of Researchr (via keyboard shortcuts) to: import citations from the browser directly to BibDesk; or send notes, highlights, and images from Skim to DokuWiki.

---

[20] http://philpapers.org/
[21] http://plato.stanford.edu/
[22] http://www.iep.utm.edu/

Selected scenario, tools, and workflow

As mentioned in Section 1, we have identified three core problems for the researchers working on the history of philosophical ideas:

1. problems with navigating and identifying relevant (digital) content and with building corpora;
2. a lack of user-friendly computational tools for conducting fine-grained textual research;
3. a lack of appropriate tools and infrastructure that allow members of research groups to work collaboratively rather than individually.

In the first year of Europeana Cloud we aimed to address these problems through a combination of the following tools (a more technical explanation of these tools is present in the Appendix):

1. Search tools to find relevant content: a specifically configured version of the ARIADNE Finder [23] is one candidate that we have started to experiment with – see illustration below. The Finder lets users search content and browse the results of different repositories, hiding protocols and standards from them.
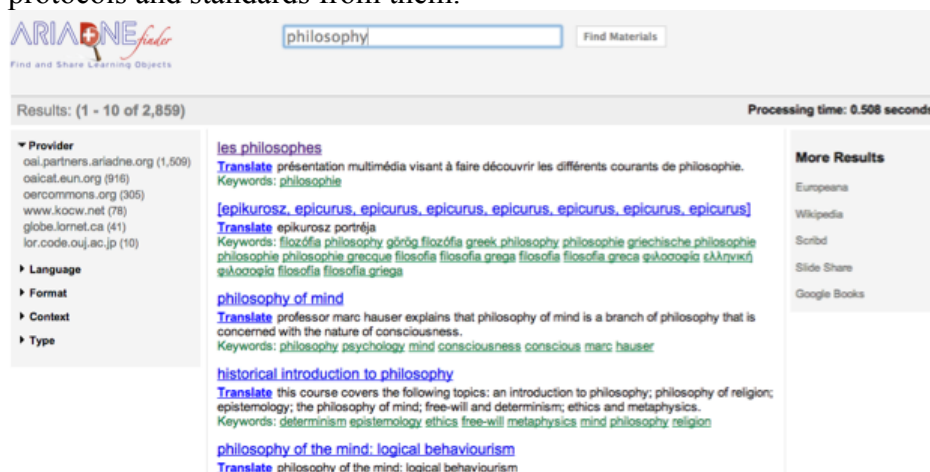
Figure 1: The ARIADNE Finder

2. Visualisation tools (geographical maps, timelines) that assist in exploring the search results. For this purpose, we have experimented with geo-spatial visualisations, such as Muse and GlamMap; which present co-authorship on a multitouch tabletop and bibliographic metadata from books in logic from 1700-1940 on an interactive map, respectively. Also, we experimented with a timeline visualisation of a term or word. Early experimentation screenshots are shown below.

---

[23] Reference?

Figure 2: Muse

Figure 1: 7,100 logic books clustered by place of publication and colour-coded by year of publication.

In the Risse visualisation (Figure 1), documents are assigned a geographic location on the basis of their place of publication. The size of the squares represents the *quantity* of documents published in certain geographic areas or places. The colours of the squares range from light blue to dark purple. Differences in colour represent differences in the *year of publication*: documents published in the early eighteenth century are represented by light shades of blue, whereas shades of purple represent documents published in later periods. The size of the squares and the colour scheme thus reveal different types of data.

GlamMap further provides users with *interactive* visualisations. By clicking on the map, users are able to zoom in on certain geographical areas or places. Zooming in refines the view, splitting the squares into multiple smaller squares, until ultimately each square represents a single city. Users can search, filter and group the metadata by clicking on geographic regions or places. This allows them to determine which books were published when and where (Figure 2). In addition, users can search and filter the metadata by specifying the author(s), title and the year of publication of documents.

Figure 3: GlamMap

Wittgenstein Nachlass Timeline

Welcome to the Wittgenstein Nachlass Timeline.
You can search for a word (e.g. "Denn dies") and see when these words appear in Wittgensteins' work over time.

This app is created by K.U.Leuven HCI Lab and is powered by MongoDB and Node.JS .
The data comes from the Wittgenstein Archives at the University of Bergen (WAB).

russell   [Find Bemerkung]

57 results.

Wunschen Russells Auffassung.

Dec 24 Dec 25 Dec 26 Dec 27 Dec 28 Dec 29 Dec 30 Dec 31 Jan 1 Jan 2 Jan 3 Jan 4 Jan 5 Jan 6 Jan 7 Jan 8 Ja

Aug Sep Oct Nov Dec 1944 Feb Mar Apr May

1941 1942 1943 1944 1945 1946 1947

Words	Start date	end date (if applicable)	Siglum	Url to wittgenstein.org	Show on timeline
Ist die Russellsche	1914-10-21	N/A	Ms-101,55r[2]	N/A	Center timeline on this..
Russells Methode in	1915-01-05	N/A	Ms-102,82r[5]et83r[1]	N/A	Center timeline on this..
Russells Darstellung ist	1918-01-07	N/A	Ms-104,21[10]	N/A	Center timeline on this..
Russells Verdienst ist	1918-01-07	1918-08-31	Ms-104,33[4]	N/A	Center timeline on this..
Russells Definition von	1918-01-07	1918-08-31	Ms-104,34[6]et35[1]	N/A	Center timeline on this..
Die Begriffsschrift Russells	1918-01-07	1918-08-31	Ms-104,54[5]et55[1]	N/A	Center timeline on this..
Russell sagte es	1918-01-07	1918-08-31	Ms-104,56[7]	N/A	Center timeline on this..
Russell, Whitehead, und	1918-01-07	1918-08-31	Ms-104,79[8]	N/A	Center timeline on this..
Russell & Whitehead	1918-01-07	1918-08-31	Ms-104,103[2]	N/A	Center timeline on this..
Die Russellsche Theorie	1929-01-08	1929-11-09	Ms-107,14[5]et15[1]	N/A	Center timeline on this..
Ich glaube Russells	1930-08-02	N/A	Ms-107,290[3]	N/A	Center timeline on this..
In Russells Theorie	1930-02-14	N/A	Ms-107,296[4]	N/A	Center timeline on this..
Russell & Frege	1930-02-16	N/A	Ms-108,64[5]et65[1]	N/A	Center timeline on this..
Russell, Eddington	1930-02-03	N/A	Ms-108,102[2]	N/A	Center timeline on this..
Durch Russell, aber	1931-11-17	N/A	Ms-112,101v[3]et102r[1]	N/A	Center timeline on this..
Russells Erklärung der	1932-02-16	N/A	Ms-113,18v[3]et19r[1]et19v[1]	N/A	Center timeline on this..

Figure 4: Wittgenstein Nachlass Timeline visualization

3. Annotation tools, incorporating information extraction methods and Linked Data techniques, in order to help researchers to share detailed comments and annotations. We have experimented with AnnotateIt and Pundit, which are web applications that allow users to easily annotate web content. As an extra feature, Pundit also allows the creation of semantically structured data that will enrich the Web of Data. Finally, we also evaluated TEXTUS, a web platform that allows the creation of shared and collaborative annotations between researchers. Again, we include some screenshots of our early experimentation in this area.

Annotator

Code / Download / Wiki / Issues

The Annotator Demo

Annotator is a Javascript shim that you can insert into any page, allowing you to select and annotate text, images, or (almost) anything else.

How do I use it? We think you'll get the hang of it pretty quickly. It's running right now on this page — try it out by selecting some text and clicking on the note icon. Enter a note and press "Enter ↵": you should see your annotation as a highlighted piece of text. Hover over the highlight with your mouse to view, edit and delete the annotation.

What happens to my annotations? The annotations on this page are being sent to AnnotateIt an online web service for storing annotations. However it's simple to set up your own store. Check out the Storage and Annotation Format wiki pages for more details.

Figure 5: AnnotateIt: the Javascript annotator

Figure 6: Pundit

Figure 7: TEXTUS and OpenPhilosophy.org

4. Social awareness tools: enable researchers to be more aware of the research activities of their colleagues and peers – see below for an early experimentation screenshot with the use of activity streams listing the behaviour of users.
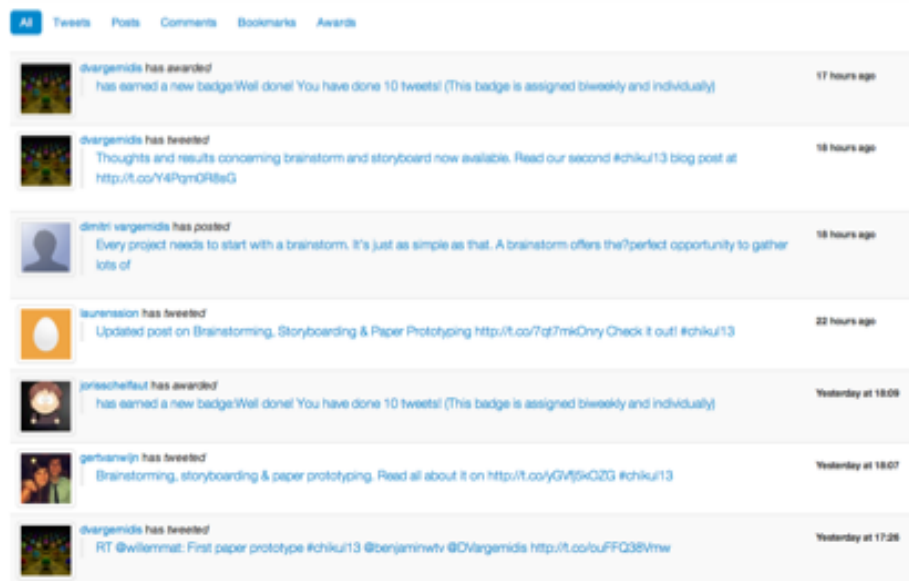
Figure 8: TiNYARMs' Awareness Stream

A generic workflow for these tools could be the following: a Humanities researcher wants to find more about 'Bolzano and Tarski' and uses the ARIADNE Finder to get content to be studied. After that, he tries to visualise them on a timeline or using a interactive map (or other visualisation tools) to further filter the content and get a better overview of the different resources found on Europeana. After selecting the most interesting resources, these could be annotated using Pundit (or other). Finally, all the different actions made during this process (search, visualize, explore, annotate, download) are captured and presented via the Activity Streams tools.

In the second part of the first year of Europeana Cloud (M7-M12), we have linked several of these tools to each other, and to Europeana content. More specifically, we have created a mash-up of the ARIADNE Finder (a search tool), TimeMapper (a visualisation tool), and Activity Stream (a social awareness tool). This mash-up was positively evaluated and was taken to assist researchers with (i) accessing *known* relevant research content or resources; (ii) with discovering *new* research content or data; and (iii) with working *collaboratively* rather than individually. We eventually did not include annotation tools in our mash-up since the philosophers often lacked corpora that allow for annotation by means of tools such as Annotator and Pundit. Work on the mash-up is described in more detail in deliverables 3.2 and 3.3, and in a publication[24].

4. *Year 2: Early Music Tools*

Personas and Scenarios

Mike and Doug working on Orlandus Lassus and Philippus de Monte

Persona

Mike is a musicologist who is working as a post-doc at the University of Leuven. He is particularly interested in Orlandus Lassus and Philippus de Monte. Doug is a doctoral student at the University of

---

[24] H. Van den Berg, G. Parra, A. Jentzch, A. Drakos, and E. Duval. "Studying the History of Philosophical Ideas: Supporting Research Discovery, Navigation, and Awareness". In: Proc. of i- Know '14. Graz, Austria: ACM, 2014.

Utrecht who is focusing on the same two composers. Mike and Doug both participate in a small EU project with additional partners in Spain and Poland.

Scenario

Mike has been working on a manuscript he discovered recently in the library of Bologna. He has a hunch that the manuscript may be an unknown piece by Lassus or De Monte or a contemporary. The manuscript is just a fragment of a large piece. Mike has obtained high-resolution photos of the manuscript while he was in Bologna for a research stay.

This morning, Mike uploads the photos to the Europeana Cloud environment. A first rough transcription into musicXML[25] is created automatically and Mike inspects the transcription through MuseScore[26]. Mike then searches for similar transcriptions in the AriadneFinder. The results are displayed on a map, on a timeline and as a Google-like list of results. Mike filters the results to the time period that interests him and discovers that there are three existing manuscripts that may be of relevance, one of which was created in Munich, where Lassus lived.

Mike uses Aruspix[27] to compare the different original manuscripts as well as their transcriptions. For two manuscripts, a sound recording is available: listening to these recordings helps Mike to get a clearer mental picture of the music represented in the manuscripts.

While he is doing this, he receives a message from Doug, who had discovered what Mike was doing through the Europeana Cloud activity stream. Mike is excited about his findings and sends a message back to Doug about his hypothesis that three of the fragments may be part of the same original piece. In a brief videoconference, they discuss how they could validate this hypothesis.

Mike needs to wrap up as he has a lunch appointment. He uploads the Bologna manuscript to IMSLP[28], so that his peers can also make use of the results of his work.

John and Music-Mapper

Persona

John holds PhDs from the University of Edinburgh in Artificial Intelligence and in Musical Composition. His research career covers computational linguistics, computational logic, computational modeling of music perception and cognition, and computational creativity.

Scenario

During his PhD, John created a personal collection of his transcriptions from 15th century music using CMME-XML, as his favorite tool to work with was the CMME software[29]. Currently, he wants to collaborate with other researchers and provide access to his collection, but they only use the Music Encoding Initiative (MEI) schema[30]. Thus, he decided to try a tool suggested by one of his colleagues, the Music-Mapper, which will transform his CMME-XML files (using the up-to-date XSLT transformations) to the MEI schema. By transforming the XML files, he discovered some

---

[25] http://www.musicxml.com
[26] http://musescore.org/
[27] http://www.aruspix.net
[28] http://imslp.org/
[29] http://imslp.org/
[30] http://music-encoding.org

missing features, and wrote the appropriate XSLT templates and contributed his code to the Music-Mapper.

Johnny and Music21-Online

Persona

Johnny is a PhD candidate in the School of Music from the University of Victoria. In the past month, he finished his master and now he just started to work, pursuing a PhD degree in musicology. During his master thesis work, he completed a thesis on hexachordal segmentation of Italian madrigals. His current research interests include musical segmentation and categorization, form, and schema theory.

Scenario

For his PhD, Johnny is trying to analyze music in symbolic form, obtained from his PhD promotor. While his colleagues use Music21[31] for such task, he does not have a strong background in programming languages, such as Python. One of his colleagues mentioned that there is a Music21 online project that he may want to try. Music21-Online provided him with the possibility to just upload his XML files via the tool and try different Music21 scripts already created by the community of musicologists using the tool. By trying a couple of scripts, he noticed that, in order to find modal and hexachordal patterns within polyphonic structures, he needed a combination of three available scripts, so he decided to merge them and create his own, which he later submits to the database of available scripts, so that colleagues can take advantage of his work.

Karin and Ariadne Finder

Persona

Karin is a 20 year old BA student of musicology at the University of Amsterdam. She is interested in historical musicology, especially in the history and development of baroque music. Like most students, she owns a laptop, a smartphone, and she is a member of various social media sites, such as Facebook and Academia.edu. Nevertheless, Karin has little knowledge of ICT tools and methods that may help her with her studies and research. Her use of ICT tools when studying is often limited to standard word processors (Microsoft Word), reference managers (Zotero), and Google, which she uses to find books and articles.

Scenario

Karin is conducting a research assignment on baroque music. For this assignment, she needs to identify and transcribe a number of eighteenth-century music scores of works by composers such as Bach, Händel, and Scarlatti. Karin needs to identify historical prints of works composed in the eighteenth century, but she also wishes to know whether the works she is studying have been included in modern scholarly or critical editions. Identifying historical prints of works and matching these prints to works contained in critical editions is a time consuming affair. Karin uses the Ariadne Finder, a personalized micro site that can be used to search and discover resources, to help her with these tasks. The Finder contains predefined search categories linking to (i) historical music sheets of works by the baroque composers she is interested in (records contained, e.g., in Europeana or RISM), and (ii) links to critical editions containing works of these composers (records provided, e.g., through Google Books and Oxford Music Online). By comparing the search results of (i) and (ii), Karin can match scholarly editions to historical prints more quickly than she could before.

---

[31] http://web.mit.edu/music21/doc/about/index.html

Maarten and Aruspix

Persona

Maarten is a PhD student at Oxford University and specializes in musical culture of the sixteenth and seventeenth century. He obtained his Master degree in digital musicology at King's College London and has only recently started work on his PhD thesis. In Oxford, Maarten is part of a research group that studies written renaissance music. Maarten is tech savvy. He has expert knowledge on popular tools and toolkits used in musicology, such as Humdrum[32] and Music21, and he is familiar with programming languages such as PHP and Python.

Scenario

In order to apply computational tools to written renaissance music, Maarten requires appropriate digital corpora. Such corpora are, however, hardly available. Historical music sources are provided online as digital images, and these images are often of low quality. Maarten uses Aruspix to build his corpus of renaissance music. Aruspix allows Maarten to upload images of music scores, to clean these images, to apply OMR to the cleaned image, and to quickly correct errors that have arisen during the OMR process. As such, Maarten quickly obtains a digital version of the music scores he needs.

Erik, Aruspix, and the Europeana Virtual Transcription Library

Persona

Erik is a professor in digital musicology at the University of Leiden, The Netherlands. He has just received the "spinozapremie" from NWO, and wants to start a large scale transcription project of masses and motets written within the renaissance period. This will allow him to apply a broad range of computational tools to a large corpus of music scores and to detect musical patterns over large periods of time.

Scenario

For his research project, Erik needs to transcribe more than one thousand written music scores. Erik wants to use Aruspix to build his corpus, which allows him to upload images of music scores, to clean these images, to apply OMR to the cleaned image, and to quickly correct errors that have arisen during the OMR process. Erik does not, however, have the time to transcribe all this work himself. Inspired by the Virtual Transcription Laboratory[33], Erik decides to build a virtual transcription laboratory for music scores, which uses Aruspix for OMR. In this virtual environment, musicologists from around the world can upload images, OCR these images, and edit the resulting transcriptions. All transcriptions, including transcriptions in progress, are saved and stored within this environment, and users can also edit and manually enrich the metadata of projects they have been working on. By using this environment, Erik is able to quickly build a high quality corpus, to easily collaborate with research partners abroad, and to share and disseminate research products.

Klaas and MusicRadar

Persona

---

[32] http://www.humdrum.org/

[33] https://confluence.man.poznan.pl/community/display/WLT/Introduction+to+Virtual+Transcription+Laboratory

Klaas is a Postdoc focusing on early music by Josquin des Prez and others, who tries to classify music, find related work and patterns, and share his insights with his fellow musicologists.

Scenario

Klaas wants to find out if a motet by Josquin des Prez he found has been described by other researchers. He therefor first tries to identify relevant scores on ISMPL. As he doesn't find the score for his motet in that repository, he develops the encoding in MusicXML himself. Once he has finished the transcription of the motet, he uploads it to IMSLP, and then runs a classification job from his own music software library. The classifier has access to not only the motet chosen by Klaas but also his other music, the IMSLP, Europeana and other libraries. Internally, the classifier uses one or more components in Weka[34], which Klaas trained before. He finds works by Franchinus Gaffurius and Serafino dall'Aquila and the classifier suggests him clusters of works. Klaas then tags the groups and chooses to share his findings. He furthermore adds some manual tags and notes to the results. His software converts his annotations into an EDM metadata blob and writes it to the Europeana API. As such, his software, which he calls *MusicRadar,* allows musicologists to share and see metadata with provenance information (author, tool, etc.) from the various tools and research tasks.

Hugo working on Ottaviano Petrucci

Persona

Hugo is a 'traditionally' trained musicologist, teaching Early Music at Utrecht University. He has a special interest in music prints from the first decades of the sixteenth century. Hugo recently discovered the richness of the Internet in doing his research. In the past, he mainly used the physical books of his institution library (e.g. the volumes of the Grove music encyclopedia, printed facsimiles of sources of early music and the editions of the Corpus Mensurabilis Musicae), but now he turns his attention more and more to digital resources and tools. Hugo doesn't consider himself to be very technical, so a friend of his – a bio-informatics professor – helps him in assessing the resources and tools he find.

Scenarios

Metadata mash up (step 1)

A colleague of the media department pointed Hugo at the Europeana portal, where he searches on one of the first printers of music, Ottaviano Petrucci. On the results page, Hugo sees a lesser-known print: the *Motetti de la corona – Libro tertio*[35]. He takes a closer look at the metadata provided and starts combining it with biographical and bibliographical information (titles of pieces/composers, printers surrounding, handbooks) he gathers from the Oxford Music Online[36], the digital edition of Stanley Boorman's Ottaviano Petrucci: Catalogue Raisonné[37], the open data of RISM[38] and several smaller databases on Early Music (like CMME[39], the Josquin Research Project[40] and the Motet

---

[34] http://www.cs.waikato.ac.nz/ml/weka/

[35] http://europeana.eu/portal/record/03486/urn_resolver_pl_urn_urn_nbn_de_bvb_12_bsb00077434_2.html?start=13&query=petrucci+music&startPage=1&rows=24

[36] http://www.oxfordmusiconline.com

[37] http://books.google.nl/books?id=xgW4C5Xp2SgC&printsec=frontcover&dq=boorman+petrucci&hl=nl&sa=X&ei=lwq9U662I-XI0wW-i4CIBQ&redir_esc=y#v=onepage&q&f=false

[38] http://opac.rism.info/index.php?id=8&id=8&L=1

[39] http://www.cmme.org

[40] http://josquin.ccarh.org

Database Online[41]). He gathers, arranges, and saves all the information in a spreadsheet in Google Drive. In a later stage, he plans to make the document available for his students.

Transcription & analysis (step 2)

After finishing his bibliographical research on the Motetti de la corona print, Hugo continues working on the musical content of this source. He downloads the image files of the print[42] and feeds them into Aruspix, recently installed on his computer by his friend. The software produces rough editions of the music, which Hugo analyses making use of the tools of Music21 (which makes a comparative analysis of the styles of the composers involved: how dissonance is treated, distribution voices, the use of pre-existing models). He notices that one of the pieces attributed to Johannes Mouton has features not complying with the style of this composer. He adds his findings to his own spreadsheet, and via the API of Europeana uploads his editions and annotations to the cloud.

Encoded music files matching

A colleague of the music library pointed Hugo at the Europeana portal, where he found a lot of digitizations of early music sources. Hugo recognizes several compositions he saw earlier on the websites of the Josquin Research Project (JRP) and CMME projects. His bio-informatics friends helps him to harvest the files from both projects and advices to contact the informatics people at Utrecht University, who work on Music Information Retrieval (MIR) research. Together with a professor and PhD student of the Interaction Technology department, Hugo is able to extract titles, composer names and incipits from the JRP and CMME files and matches them with the metadata from Europeana, referring to the source images. He uploads the matched encoded files to the Europeana Cloud, linking them with the image files.

---

[41] http://www.arts.ufl.edu/motet/default.asp
[42] http://stimmbuecher.digitale-sammlungen.de/view?id=bsb00077434

5. *Year 3: Exploring the Europeana Newspaper Archive*

These personas and scenarios are results from the Amsterdam Workshop in 2015, further described in D3.3.

Results from breakout group 1

Participants: Gerard Lynch, Simone, Astrid, Suzan

Rada

Persona

Rada is a female Romanian researcher in her fifties, an authority in her own field of female authors in Romanian literature. The research environment at her home university is not state-of-the-art and she is not fully familiar with digital research tools.

Scenario

Collaborating in a H2020 project on the reception of women's literature in a European context, she sees the state of the art in research and tools which are available in her partner research institutes in Finland and UK contexts.

One such tool is a joint database of women writers across Europe which consists of biographical information and a list of results linking to individual national databases, such as the Serbian National Library list.

This database does not connect to any Romanian national data, making her feel left out of the research field as a whole.

She is interested in easy access to the English reception and Romanian reception to English female authors in the press and would be interested in using the Europeana corpus towards this end.

She goes to the computer and searches for the names of female authors in the English newspaper corpus and is interested also in articles on these authors in a Romanian context. One such author is the Romanian Carmen Silva, the pseudonym of Queen Elizabeth of Romania.

She wishes to answer a question regarding female authors writing about children, as there is interest in whether the tendency to write about children was in relation to a personal tragedy or a general expectation of female writers.

Carmen Silvia was particularly well covered in the Dutch press and translated into Dutch, therefore she searches in the Europeana database for the reception in the Dutch press.

Ulrike

Persona

Ulrike is a female Dutch postdoctoral researcher on a digital humanities project on identifying individual female authors who until now were lesser known. She has experience with VRE environments from her current project. She wished to expand upon this data using the Europeana corpus.

Scenario

She wishes to find out about lesser-known female authors in the Europeana corpus. These authoresses may be mentioned in a context of a more famous female author for whom she has more information.

The search would focus on searching for specific names or simply going through a list of review articles from a particular time period.

The search results will suggest female names which occur in a context related to well-known female authors (e.g Countess Dash) and suggest whether these are likely to be authors themselves based on the textual context (X wrote, X's novel, the author X). From the reviews we wish to extract similar authors based on textual comparisons (X is like Y).

These authors names will be imported from the existing VRE lists.The results should be ranked by "plausiblity" of whether this is a person and if possible also an author.

These lists will then be manually validated by her own hand.

She searches for her interested author (Countess Dash) and receives a number of results. These are then added by hand to the VRE database. It would be ideal to have an export function from the Europeana search result.

Jan

Persona

Jan is a postdoctoral researcher in a digital humanities project concerned with antisemitism and the rise of nationalism in a European context. He is familiar with scripting languages such as Python and has experience in visualisation and computer programming.

Scenario

He is interested in documents related to the topic of antisemitism and cultural nationalism which is not simple related to simple keywords.

At the beginning, he searches for keywords such as 'Jew', 'jewish' and famous Jewish people related to anti-Semitism (Dreyfus Affair for example) in the Europeana corpus. The system should suggest related entities to his source terms, related to the Dreyfus affair he would like to find out about the French major Ferdinand Esterhazy who was the real culprit in the affair.

He would like to find the occurrences of these keywords related to a geographical and temporal context. As a first step, he would like a sentiment analysis on these documents as a first step to help him focus on a temporal period with more negative comments for example. This sentiment should be presented using a combination of a heatmap or timeline to enable the user to visualize the sentiment over time.

Jan talks to a historian colleague in the history department and he validates his sentiment results with a summary of the findings extracted from the output, the historian tells him that some later examples of positive sentiment he found in a French context are likely to be inaccurate.

He would be interested in a machine-readable API (preferably in JSON) which would allow him to download and analyse the text locally.

Results from breakout group 2

Participants: Jacco van Ossenbruggen, Christos Mais, Federico Nanni, Michail Moatsos, Lucy Delap

Maria and Eileen

Persona

Maria is a 25 year old PhD student from university of Coibra. She works on comparative anti-colonial struggles, as represented in UK and Portugal through the press in the 1940s to 1970s. Her professor is 62 and non computer literate. He is sceptical about Maria's research methodology. Maria speaks English, Portuguese and Spanish. She has digital access to daily newspapers, but not the illustrated or quarterly press, which are in paper format only, and she works on these in more traditional methods.

Persona

Eileen is a postdoctoral researcher in Ireland; she comes from a Republican family, and her father is a Sinn Fein activist. She studied as a mature student, and is now aged 43. Her PhD was on the Irish civil war in Dublin. She is now interested in the evolution of women's support for the Republican movement, 1922-1970, using prosopographical methodologies as well as media research. Maria's professor was on her PhD committee, due to his expertise in conflict and colonialism. She wants to track how representative her Irish Republican women are in relation to European anti-colonialist, and is contacting other European researchers. Eileen speaks English and Gaelic, but no other European languages.

Scenario

Maria wants to search the newspaper press and gain a full sense of what the newspaper 'feels' like, by being able to page through its full contents. Not all Portuguese newspapers are digitised, and only 3 are on Europeana. There is another subscriber paid platform, Proquest, which hosts newspapers A and B, and she has funding to subscribe for only 1 year, so she has time pressure to download her data from this source into a format that she can work with over the next 3 years. She will download the files in pdf format; these are in a searchable format. Europeana is also used, and provides contextual information on the newspapers charting their price, political affiliation, format and distribution patterns across regions. Europeana also gives her the chance to contact other researchers in her field. She can upload her biographical details and citations to an online account, along the lines of Mendeley[43]. She can see who has searched for the same individuals or themes as her, if they have allowed others to see this information. She herself can opt into being able to allow them to contact her.

She'll start with a look at some key individuals that she knows about, and searches for Ms x, Dr Y and Mr z. How representative are the key individuals that Maria already knows? It turns out that Dr Y was only involved in the Portuguese anticolonial press for 6 months, whereas Ms x was involved for five years. Her search platform can give an indication of how many articles these individuals have authored, over time, and how often they are mentioned in other people's texts. This can be visualised on a timeline.

Mr Z, some searches reveal, was writing in a newspaper which was owned by Mr G, who was strongly invested in Angolan cashew nuts – Maria discovers this by reading a biography of Mr G which was provided as part of the contextualisation of the newspaper. She is able to understand better why Mr Z is very muted in his criticism of the Portuguese in Angola.

How can she convince her professor that her work has value? He questions the quality of the OCR. It turns out that the government-supporting newspapers have a better printing quality, so the OCR makes less mistakes; the results are biased towards them. Maria has enough contextual knowledge to have already noticed this problem, and she reports this back to Europeana through their feedback tool which is embedded in the platform. While waiting for their response, she searches the paper versions of the anti-government papers, so as to counteract the bias.

How can Maria bring together her research insights from her paper based research with her digital work? She is looking for a single representation that spans the digital resources available on different platforms (Europeana plus Proquest Portuguese digitised newspapers). Europeana provides a plugin to integrate multiple datasets, across subscriber-sites, their content, and user generated scans of digital sources. However, due to copyright restrictions, she can't make her scans publically available. She is however able to create her own sole-use account, where she can bring these datasets together to create outputs that include heatmaps, timelines, topic clouds. One of her sources is a pamphlet that is not copyright, and she can share her own scans of this with her research community.

---

[43] http://mendeley.com (Citation Mananger)

Eileen got to know Maria through a social media function on Europeana because they were both looking at anti-colonial republican Ms X. She asks Maria to give her an impression of UK and Portuguese women's involvement in anti-colonial struggle. Maria has not yet considered gender as a major feature of her study. She knows the names of some key women anti-colonialists. How can she create a subset of her data to look at this issue? How can she present her data to Eileen? Maria sends Eileen a database of possibly relevant material from the UK and Portugal, for Eileen to browse, that covers the period 1940-1975.

Despite Eileen not speaking Portuguese, Europeana offers her a rough translation option which allows her to identify which articles seem relevant. Eileen choose to send four of the articles to obtain a fuller and more professional translation via her own university research office. Eileen has now got a better sense of the relevant words to search on in Portuguese, and so she creates her own subset of Portuguese and UK documents to cover the period 1920-1940, which fell outside Maria's expertise.

Both researchers also want to know what other European contexts experienced this kind of activism – Finnish, Greek, Romania, for example. They tried to use Europeana Cloud to create rough maps of where and when anti-colonial activists were being discussed in the press, but realised through reading secondary literature that they didn't have enough knowledge to find the historically adequate keywords. They conclude that only by collaborating with researchers in other national contexts could they do sophisticated searching; they are looking to meet collaborators through their search platform.

Results from breakout group 3

Participants: Eliza Papaki, Gabriela Rus, Christine Watson, Peter Dängeli

Peter

Persona

Peter is a 26-year old, PhD student in Transnational History and is interested in studying Irish Famine in European press. He wants to find to what degree and in what ways different countries reported this event, making a distinction between reports that cover humanistic aspects and reports that are just fact-based. In what journalistic type of documents was this crisis reported and what terms were used to describe this event. He has already explored secondary literature and has a good knowledge of background information and terms and he now wants to move on to primary sources (actual newspapers). Based in Ireland, he has access to Irish press and to whatever he can find online. His supervisor told him about a new platform, Europeana Newspapers platform, on which he can find digital material.

Scenario

He starts searching the terms 'famine', 'Irish emigration', or 'starvation' and their translations (suggested by the platform) in combination with the option "texts about Ireland" and the temporal range 1845 to 1852. His search provides an option to display the frequency of the search terms graphically by means of a heat map of Europe (results are displayed by city or country), by year and over the whole search period. Hovering over heat patches shows a small tooltip giving information about the number of entries and newspapers of that area/city. He then clicks on Paris, Rome and Berlin and a list of articles on Irish Famine appear in a chronological order. He observes that newspapers edited in cities with a larger proportion of Irish immigrants seem to cover the crisis more extensively. The map does however not indicate information about the distribution of e.g. journalistic text categories. Using a tab below the map, Peter switches the view from the heatmap to a timeline that visualizes the hits per newspaper and over time, accompanied by the totals per newspaper over the whole search period in a bar chart.

The heatmap reveals a high concentration of results in Italy, which is geographically remote from Ireland, a fact that draws his interest. So he starts gathering articles of the Italian press. He wants to find out why the interest in the famine is greater there than in countries that are geographically closer, such as Norway. However, as he doesn't have a good knowledge of foreign languages he faces the problem of translating this content. The platform can suggest translations of his search terms into different languages, but he can't evaluate the results he gets or offer translations of the whole article in word.

Now what? He discovers "researcher's chat", a section in the platform where researchers exchange ideas, discuss and validate search results. He then decides to write about his research problem and asks who can help him deal with that. In this way he comes into contact with Verena (see below). Peter creates an account in the platform which allows him to save selected results of his searches in the form of a personal corpus, tag his results in the way that he needs (e.g. with regard to the journalistic textual category) and interact with other researches. Since he is a PhD student, he wants to be able to share his corpus with his supervisor as well as to reference the stable address of his personal corpus in his dissertation, thus giving the reader access to the (digitised) primary sources. He is preparing to give a presentation about his PhD research so far to his supervisor so he wants to elaborate on his search results and present them visually. He goes to the tab Tools of the platform and a section appears with a list of suggested tools for visualisations. He starts playing around with them to see which is more useful to him. He chooses tool X which creates visualisations of the statistics of reference of specific terms in the press. He now uses these visualisations as part of his results, to provide quantitative information while he is interested in creating his own heatmap displaying his own collection of sources in Germany, Italy, France, England and Ireland over 1850-1860. So he opens Tool X, imports his own collection and selects the specific countries that he is interested in for the specific time span. Then, such a visualisation appears which he saves as jpeg on his laptop to be used in his presentation.

Verena

Persona

Verena is a post-doc linguist, based in Germany, with an interest in exploring expressions of empathy and sympathy in German, English, Italian and French newspapers in 17th-19th centuries. She has been using search terms such as 'empathy', 'sympathy', 'tragic', 'poor', and first person pronouns (since texts written in first person may express a subjective point of view), and she wants to know how frequently and in what way empathy and sympathy are expressed having either case studies specific in time and place, or general observations on what are the traditions of specific countries in reporting events. She has excellent knowledge of German (native), English, Italian and French languages which is an asset for her research topic.

Scenario

Verena has created her own collection of texts during her research by gathering relevant texts saved in her personal computer. For analysing this corpus, she uses a personalized set of tags for her results, such as tagging reports as being written in the first person, using strong adverbs like 'extremely', 'terribly' etc. She wants to be able to export the results as text to use other corpus tools, but uses visualisation tools 'TimeMapper' and 'Timeline' as well. What she wants is to be able to see a timeline with a curve of the frequency of expressions for empathy in different times and countries. She wants to be able to visualize how news about particular events spread to different places. One of her colleagues suggests her to use the new platform Europeana Research where she can play with several digital tools listed there.

Verena reads about Peter's problems on 'researchers' chat' and since she has been doing research on the cases in which the press does express sympathy, in a longer time span, but has not looked into the

purely fact-based reports where no sympathy is expressed, she is interested in this aspect of Peter's work. So she replies to his message on the platform and they start exchanging emails and collaborate digitally. They exchange corpus retrieved from the platform on Irish Famine and discuss on the search results and translations provided of the description of sources in a private conversation on 'researcher's chat'. As a German native speaker with knowledge of other languages as well, she evaluates the translations provided and gives feedback to the platform. In order to do this she either posts a message on 'researchers' chat' or she 'likes'/'dislikes' from her account the translations provided.

6. *Year 3: Agricultural Researchers*

Personas and Scenarios

John

Persona

John is a professor of Agricultural Economics and Rural Development at Agricultural University of Athens, Greece. He is interested in having access to economic data for specific crops in Greece and worldwide. He spends a lot of time searching for all the latest developments and publications in his field and he would like to be able to save time when looking for this information by getting the search results in one place if possible.

Scenario

John participates in a European project with partners from all over Europe. One of the tasks that he has to undertake is to present the current status of agriculture for specific crops in Greece, including information about production, market trade (import-export) as well as information on rural inclusion and the significance of farm cooperatives. After gathering this information, he then will have to compare it with equivalent data of other European countries. For the information involving Greece, he knows that he will visit the Agricultural Economics & Policy Research Institute (AGRERI) where all this information is gathered. But he is frustrated as to how find the information related to other European countries. Having enhanced the AGRERI website with a search interface that will allow users to discover and navigate relevant digital resources that are coming from large aggregators, as is the Europeana aggregator, John will be able to browse through resources related to agricultural economics and farm cooperatives, coming from all over Europe simply by using keywords or the filter facets provided and save a lot of time and effort for his task. He will have access to texts and images that will help him present the current status of agricultural economy in Greece compared to that of other European countries.

Michael

Persona

Michael is an agriculture advisor and a farmer in Crete. He takes care of his family vineyard and he produces grapes for his friends and family. His main occupation is to consult and advise accordingly new and experienced farmers on the changes that occur in the agriculture legislation.

Scenario

Michael participates in many seminars held by the Ministry of Agriculture that aim to inform agriculture advisors and farmers about all the developments regarding the market trade, prices and changes in legislation on specific crops. He has a very good idea of the situation in Greece but he wishes to find articles of agriculture legislation from other countries and know what is the status in other countries, so that he can come up with ideas and suggestions on how to improve the situation in Greece and make Greek products more competitive in terms of quality and prices, but above all, to help simplify if possible the bureaucratic procedures for his clients. He visits the Agricultural Economics & Policy Research Institute (AGRERI) where he will find information on managerial and social issues as well as policies and Greek legislation reports. Having enhanced the AGRERI website with a search interface that will allow users to discover and navigate relevant digital resources that are coming from large aggregators, as is the Europeana aggregator, Michael will be able to browse through resources related to agriculture legislation from various countries and get some good practices that could possibly be applied in his country and region.

Maria

Persona

Maria is a MSc student at the Aristotle University of Thessaloniki. Her bachelor degree focused on evaluating some olive tree varieties and their oil, which is of great importance for Greek agriculture and a major export product. She is now preparing her MSc thesis which includes both bibliographical and experimental data, again concerning olive trees. She is now studying the genetic variation of olive tree varieties in Greece, but she also wishes to find historical information of the origin of olive trees not only in Greece but also in Europe.

Scenario

Maria is conducting her research by comparing the olive tree varieties using specific molecular methods. She finds out that some varieties are genetically very similar to other Italian and Spanish varieties found in other experiments and publications and she wants to find historical information on olive trees in Europe so as to make a hypothesis on the origin of olive trees since the ancient years. This information could include details on the market trade in ancient years, and therefore pinpoint how and where olive tree varieties moved around Europe during in the past. She visits the Agricultural Economics & Policy Research Institute (AGRERI) where she will find not only economic data on this specific crop, but she may also find information on social issues concerning the specific crop. Having enhanced the AGRERI website with a search interface that will allow users to discover and navigate relevant digital resources that are coming from large aggregators, as is the Europeana aggregator, Maria will be able to browse through resources with historical significance concerning the olive tree, including drawings from historic books and museums, as well as pictures from olive tree varieties of other countries. She will then add this information to her MSc thesis, in the Introduction chapter where she will be able to present the origin of olive tree varieties not only from a Greek point of view, but also from a European one.

7. *From Personas to Tools*

The main objective of this deliverable is to identify the personas, analyse the scenarios, and present specific use cases for the tools that have been developed or adapted in the context of WP3. To achieve this goal, a specific process is followed in order to connect the work described in this document with the development and adaptation of the tools. This process, also called user-centered design, involves the continuous communication and collaboration of both developers and researchers from the humanities.

The first two steps in this process are described in previous sections. These steps involve the workflow identification of (specific) researchers in the humanities and the preparation of scenarios where the different tools can be used. After writing the scenarios, a process is followed in which developers continuously update the scenarios, create mock-ups and use cases for the modelled tool, while researchers simultaneously provide feedback on their needs and requirements. This continuous communication between the two groups is needed in order to cover the needs of the researchers and final users of the tool. The final steps are the development of a pilot and a demo of the tool. A schematic representation of this process is shown in the figure below.
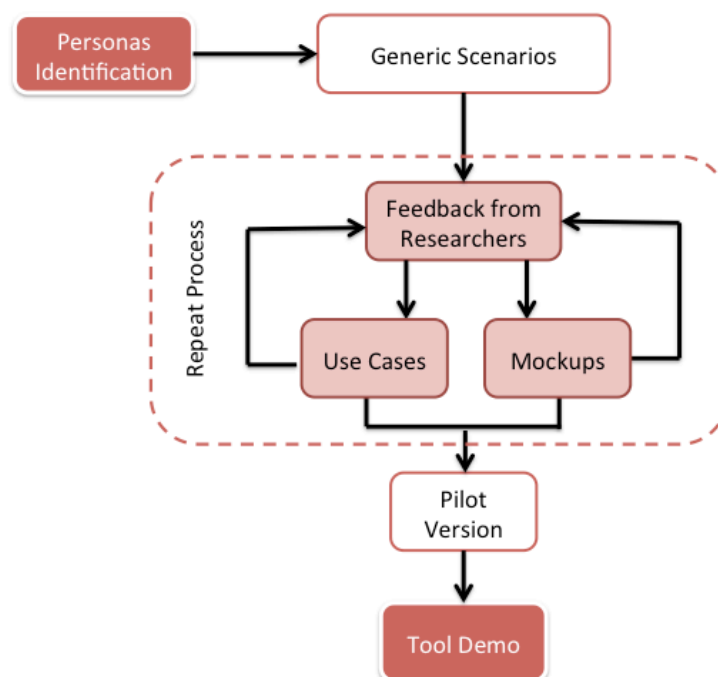
Figure 9: The main workflow followed for the development of the tools

To further analyse this process we will describe the work that has been conducted in the first year of Europeana Cloud (M1-M6) for the development of an ARIADNE Finder for the Axiom group of the VU University Amsterdam. This work is based on the scenario "Hein and the ARIADNE Finder" presented in the previous section.

The first step in our process is persona identification. In our case Hein, a member of the Axiom Group, is a post-doctoral researcher working on the history of philosophy, the history of biology, and their interplay. After analysing the profile of the persona, we proceed with a generic analysis of a scenario for the tool to be developed: Hein could use the ARIADNE Finder technology to embed a

search tool in the site of his research group to quickly search and browse through the various digital resources on the history of philosophy and biology. After the creation of the first generic scenario, feedback from Hein (as the Axiom group representative) was requested to better understand his needs and requirements.

Based on the persona and the described scenario, the following step is the creation of the first use case scenario. In our case, the following use case description was developed.

Use Case Description	
Name	Axiom Group – ARIADNE Finder
Description	The finder helps researchers in history and philosophy of science to find material that they can use in their research.
Legal foundation(s)	The copyright license of the content should be respected during the aggregation of metadata and the use of material.
Pre-condition	Metadata for different type of content should be triplified and provided through a single API.
Flow of Events - Basic Path	
Step 1	A researcher is visiting the axiom group site
Step 2	He is entering a search query or using the browse functionalities
Step 3	A results set is shown to the researcher
Step 4	The researcher can use specific facets to narrow down the results set
Step 5	The researcher clicks on specific results to show more information and to access the resource
Post-condition	All the metadata of resources are aggregated, indexed and provided through a single API
1.1 Actors	
End-users	Researchers
Information provider(s)	Europeana, European Library, GLOBE, YouTube, Flickr
1.2 Expected added value	
Expected value added	- Researchers will be able to search simultaneously in many external sources with different type of content e.g. books, publications, maps - the results will include only relevant resources to the user query - the user will be able to make complex queries - discovery of relevant resources with very good performance that will improve the user experience

Table 1: The use case description for the scenario of "Hein and the ARIADNE Finder "

The following step required the development of the first mock-ups, shown in the figures below. The mock-ups provide an illustration of what the final tool could look like based on the analysis of the scenario. Since the ARIADNE Finder is a technology that is embedded in an existing site, the first version of mock-ups included the following pages:

- A main page, shown in Figure 10a, where users can search the repositories for data. Additional buttons are included to quickly access predefined searches, specific resources or the local repository of the group.

- A second page, shown in Figure 10b, with a number of the philosophers the group investigates, links to a search for their biographies and work, and access to the group's work on a specific philosopher,
- A search result page, shown in Figure 10c, with a faceted search to quickly filter the results with a number of different criteria,
- A view more information on a specific resource page, shown in Figure 10d, where all the metadata of the resource are shown and the user has the option to either view the original resource or save a copy of the resource for annotation (enabled based on the copyright).

(a)

(b)

(c)

(d)
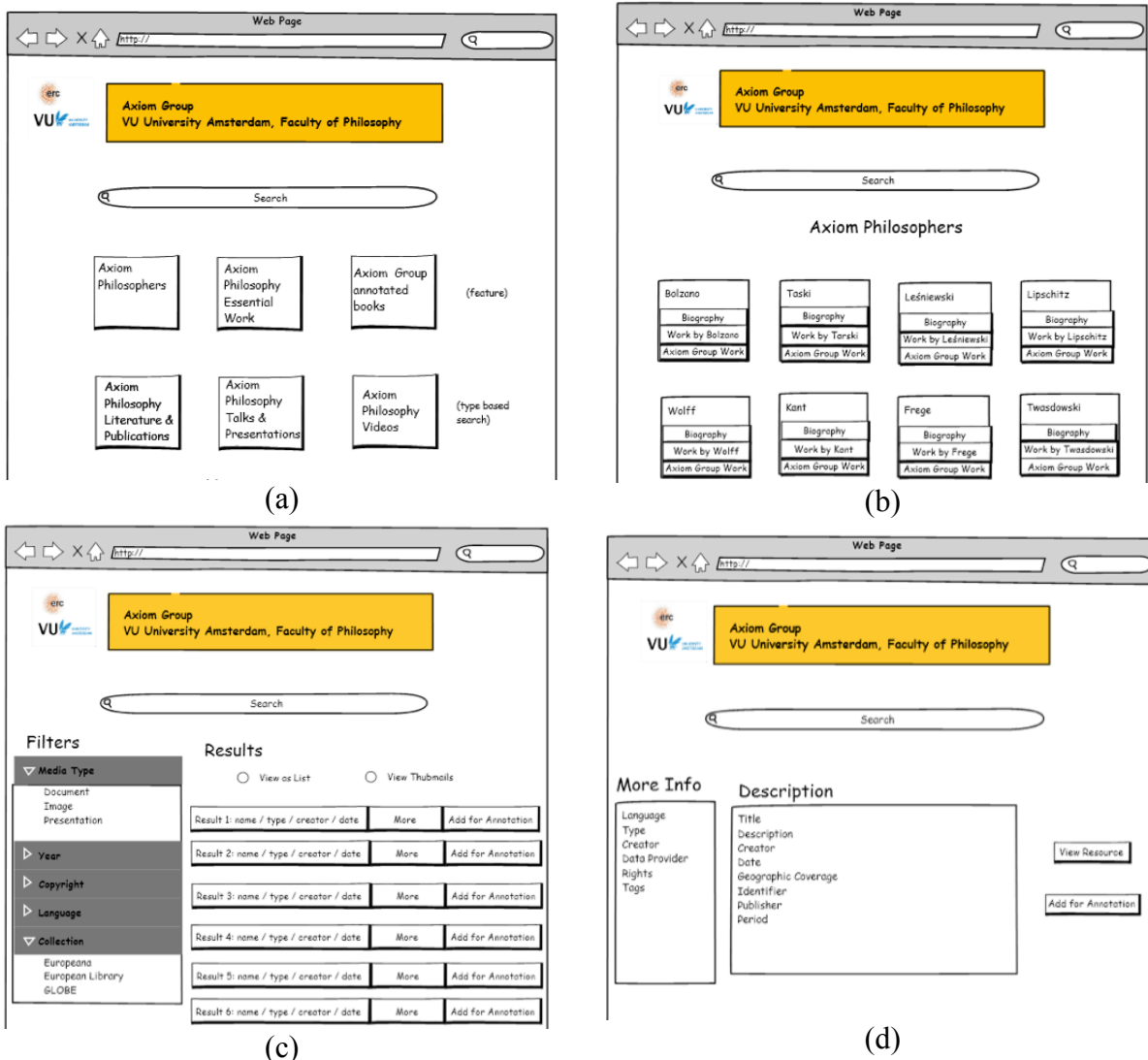
Figure 10: The first version of the 4 pages mock-ups for the Hein and the ARIADNE finder scenario

This set of first mock-ups was distributed among the Europeana Cloud WP3 group and an online meeting, open to all members of the WP3 group, was arranged with Hein, the main persona, to discuss them. The feedback gathered from this meeting led to a second version of the mock-ups and to respective changes in the use case description. Apart from comments made by the user-researcher Hein, feedback was also given by other members of the WP3 group, mainly on the technical side for the tool to be developed. The second version of the mock-ups, shown in Figure 11, had the following additional features:

- A bar to quickly add a time period in any search,

- A button to search for the group's work in Mendeley or Zotero,
- A button to view specific and selected resources (essential work),
- More collections to search,
- When viewing a resource, the options to download, export, or view related resources in Mendeley or Zotero are available.
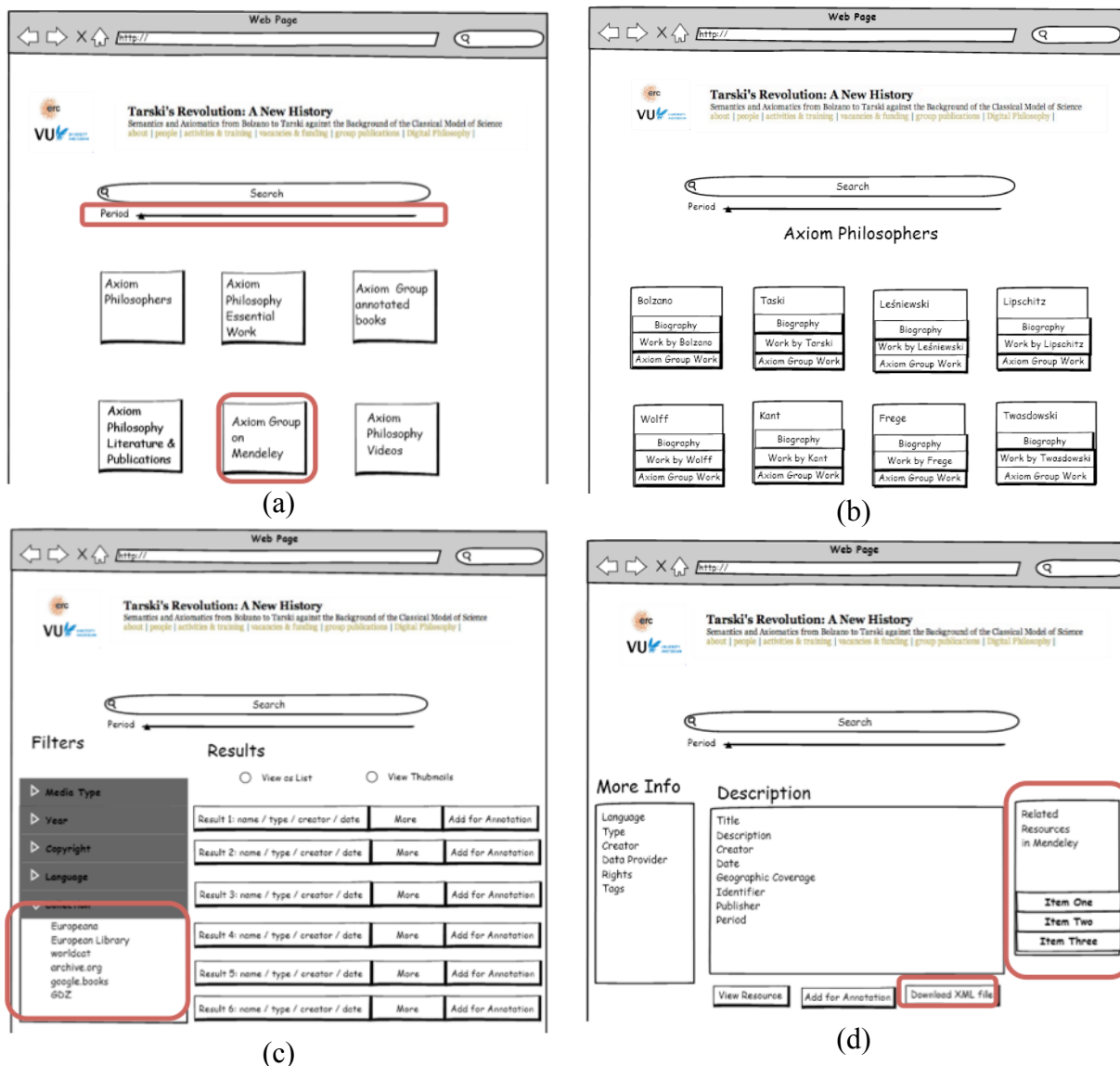
Figure 11: The second version of the 4 pages mock-ups for the Hein and the ARIADNE finder scenario with some of the changes highlighted

As explained earlier, we continuously gather feedback from users in order to update the use case scenarios and mock-ups until we reach a desired outcome and the pilot can be developed. During the first days of July 2013, the Europeana Cloud WP3 group arranged a meeting with Hein and the axiom group to discuss, among others, the ARIADNE Finder demos. The mock-ups were presented to the group along with a number of questions to guide the users in providing feedback for the technical team through a presentation that was later distributed to all the members of the WP3. Hein and the rest of the group members decided to gather additional feedback from other members of the group not present in the meeting before continuing. When all the feedback was gathered, the third and final version of the mock-ups was developed and discussed with members of the Axiom Group.

The following table summarizes all the steps that were followed for the above example in the first seven months of Europeana Cloud.

	Step	Description	Date
	Step 1	Persona identification	28th March 13
	Step 2	Initial scenario development	15th April 13
	Step 3	Feedback from Axiom group representative	27th May 13
	Step 5a	Use case development	13th Jun 13
	Step 5b	Initial mock up	13th Jun 13
	Step 6	Interview with Axiom group representative and WP3 members	19th Jun 13
	Step 7	Revised mock up	5th July 13
	Step 8	Presentation of mock up to the whole Axiom group	8th July 13
	Step 9	Final mock up & specs for implementation	31st July 13
	Step 10	Prototype development – First version	15th September 13
	Step 11	Prototype development – Final version	15th December 13

Table 2: Summary of the work for the case of "Hein and the ARIADNE Finder" in M1-M7

After presenting the first prototype of the ARIADNE FINDER to the Axiom Group, the philosophers provided monthly feedback on the tool. This feedback was considered for the next iterations. The same procedure was adopted for the testing and evaluation of the mash-up of tools (described in section 3). Once the philosophers had access to the tools of the mash-up, they were asked to provide expert feedback on whether and how these tools could support their research and workflow. A more detailed description of the mash-up of tools and its evaluation is provided in deliverables 3.2 and 3.3, and in a publication [44].

---

[44] H. Van den Berg, G. Parra, A. Jentzch, A. Drakos, and E. Duval. "Studying the History of Philosophical Ideas: Supporting Research Discovery, Navigation, and Awareness". In: Proc. of i- Know '14. Graz, Austria: ACM, 2014.

8. *Conclusion*

WP3 has developed services and tools that leverage Europeana content to be used by researchers. In this deliverable, we focus on the first step before the development of the tools, which is to understand and analyse the user needs by developing personas, scenarios and use cases; and to present a number of different existing tools that can be combined, expanded, and integrated to cover the needs of researchers. In addition, we introduced the process to be followed in order to connect these scenarios with tools development. This document is the third and final version of the Deliverable 3.1.

During the first eCloud year, we have identified three core problems for the researchers in philosophy and digital humanities:
- problems with navigating and identifying relevant (digital) content and with building corpora;
- a lack of user-friendly tools for conducting fine-grained textual research;
- and a lack of appropriate tools and infrastructure that allow members of research groups to work collaboratively.

We have addressed these problems through a combination of search, visualization, and awareness tools and the integration of those in the Europeana environment. The results are presented in Deliverables 3.2 and 3.3.

During our second year, we have identified four core problems for musicologists:
- Problems involving data creation.
- Lack of music scores.
- Information exchange and linking of data.
- Retrieval and analysis of contextual information.

We have addressed a number of these problems through a combination of tools that are already in use by musicologists as well as a set of new tools. This work is reflected in the corresponding versions of Deliverables 3.2 and 3.3.

During our third year, we have identified the following requests and problems for our target community of humanities researchers who work with the Europeana Newspaper Archive:
- difficulties navigating the online Newspaper archive
- poor usability
- limited facets for searching
- automatic suggestions of relevant articles to the one the user is accessing
- timeline per newspaper.

Also, during the third year, we have identified the following problems for another target community, that of agricultural researchers who work with the Agriculture Economics and Policy Research Institute (AGRERI):
- problems with navigating and identifying relevant (digital) content from diverse sources
- lack of tools (like specific facets and filters) that will facilitate the discovery of the resources needed in specific thematics

We have addressed a number of these problems through a combination of tools and services. This work is reflected in the corresponding versions of Deliverables 3.2 and 3.3.

9. *Appendix: Tools for researchers in the Digital Humanities*

Search tools

ARIADNE Finder
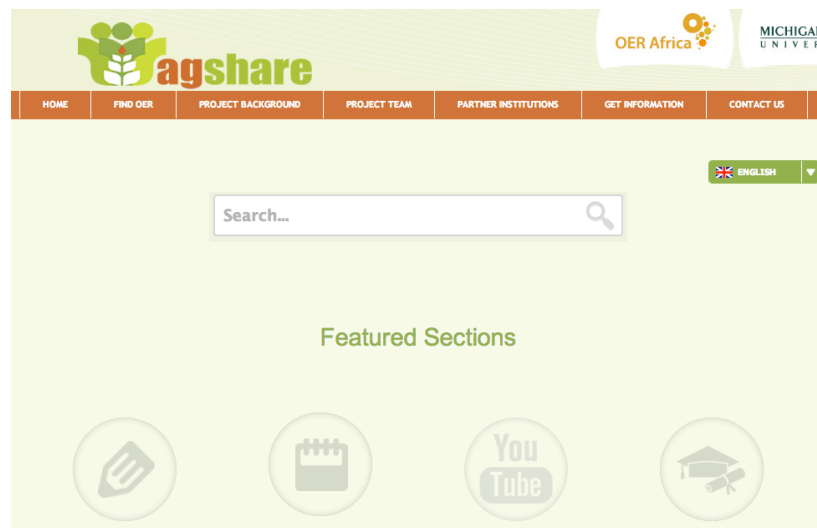
description of functionality	Lightweight search interface (implemented as an HTML page with some Javascript) that works on top of a Solr index. The Solr index is build using a number of metadata facets that can be used to navigate metadata records aggregated through different sources. Ingests metadata records of relevance to the specific users, by a number of ingestion mechanisms/APIs that are creating the Solr index behind the particular search page.
URL	http://ariadne-eu.org/wiki/
contact	nikosm@ieee.org stoitsis@ieee.org
licence	LGPL for older versions, still undefined for current/working ones (but will be a xGPL one)
programming language	HTML and Javascript
current use	at least 5 deployments of beta versions, with estimated usage by >200 users
deployment	http://ariadne.cs.kuleuven.be/finder/ariadne/ [looking into the GLOBE aggregator] http://laflor.laclo.org [looking into the Latin America learning repository aggregator] http://greenlearningnetwork.com/ [looking into the Green Learning Network OER aggregator] http://www.greenlearningnetwork.com/organicedunet/ [looking into the organic collections of the Green Learning Network OER aggregator] http://agsharedemo.agroknow.gr [looking into the Africa-related OER collections of the Green Learning Network]

Visualisation tools

A good overview on visualisation methodologies / tools by functionality:
http://www.visual-literacy.org/periodic_table/periodic_table.html

Data visualisation guidelines, by Gregor Aisch:
http://schoolofdata.org/2013/04/26/data-visualization-guidelines-by-gregor-aisch-international-journalism-festival/

TimeMapper
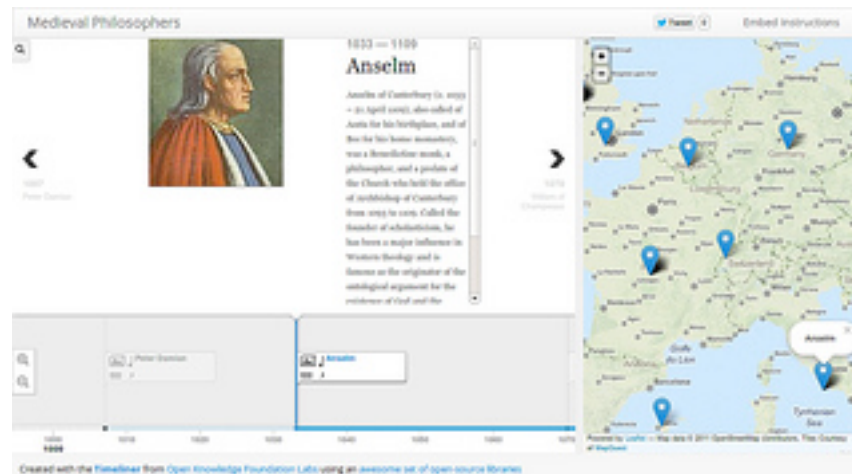
description of functionality	A visualisation tool making timelines and timemaps using Google spreadsheets.
URL	http://timemapper.okfnlabs.org/
contact	info@okfn.org
licence	MIT licence
programming language	Java Script and other open-source components including TimelineJS, ReclineJS, Leaflet, Backbone and Bootstrap
deployment	Example of an implementation on Medieval Philosophers: http://timeliner.okfnlabs.org/view/?url=https://docs.google.com/spreadsheet/ccc?key=0Al6mO9_3Hr2PdGZnRjEwUWxOekhreTNNZEFEMWRZbkE - 2 On the Open Parliament Declaration http://www.openingparliament.org/about
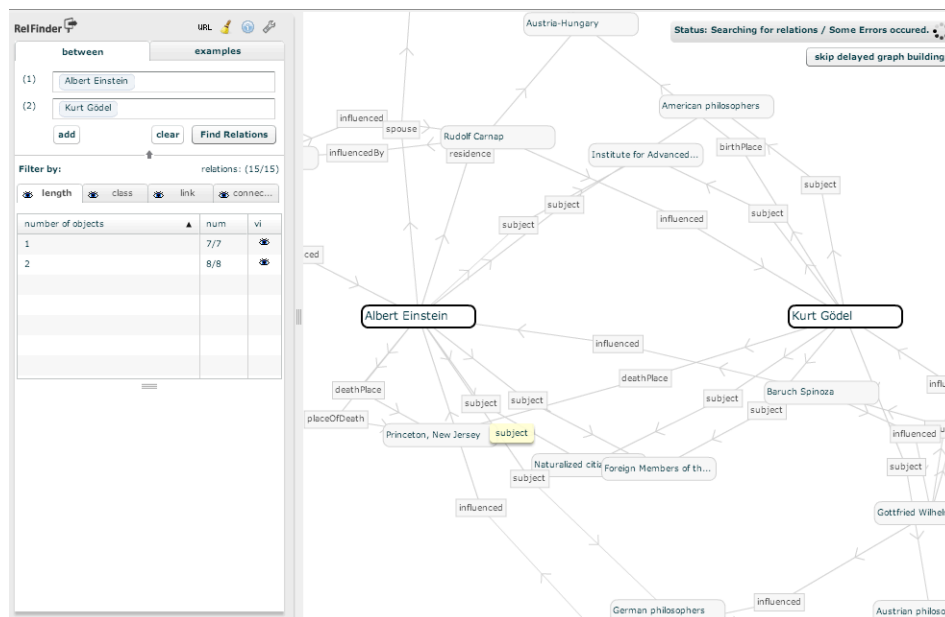
RelFinder

description of functionality	The RelFinder extracts and visualizes relationships between given objects in RDF data and makes these relationships interactively explorable. Highlighting and filtering features support visual analysis both on a global and detailed level. The RelFinder is based on the open source framework Adobe Flex, easy-to-use and works with any RDF dataset that provides standardized SPARQL access.
URL	http://www.visualdataweb.org/relfinder.php
contact	contact@visualdataweb.org
licence	GNU General Public License
programming language	Adobe Flex
deployment	University of Leipzig, http://catalogus-professorum.org/tools/relfinder/RelFinder.swf Ontotext http://linkedlifedata.com/relfinder
bibliography	Philipp Heim, Steffen Lohmann and Timo Stegemann: *Interactive Relationship Discovery via the Semantic Web*, in Proceedings of the 7th Extended Semantic Web Conference (ESWC 2010), volume 6088, series LNCS, pages 303-317. Springer, Berlin/Heidelberg, 2010. Philipp Heim, Sebastian Hellmann, Jens Lehmann, Steffen Lohmann and Timo Stegemann, *RelFinder: Revealing Relationships in RDF Knowledge Bases*. in: Proceedings of the 4th International Conference on Semantic and Digital Media Technologies (SAMT 2009*)*, pages 182-187. Springer, Berlin/Heidelberg, 2009. Steffen Lohmann, Philipp Heim, Timo Stegemann and Jürgen Ziegler.The *RelFinder User Interface: Interactive Exploration of Relationships between Objects of Interest*, in: Proceedings of the 14th International Conference on Intelligent User Interfaces (IUI 2010*)*, pages 421-422. ACM, New York, NY, USA, 2010.

Muse
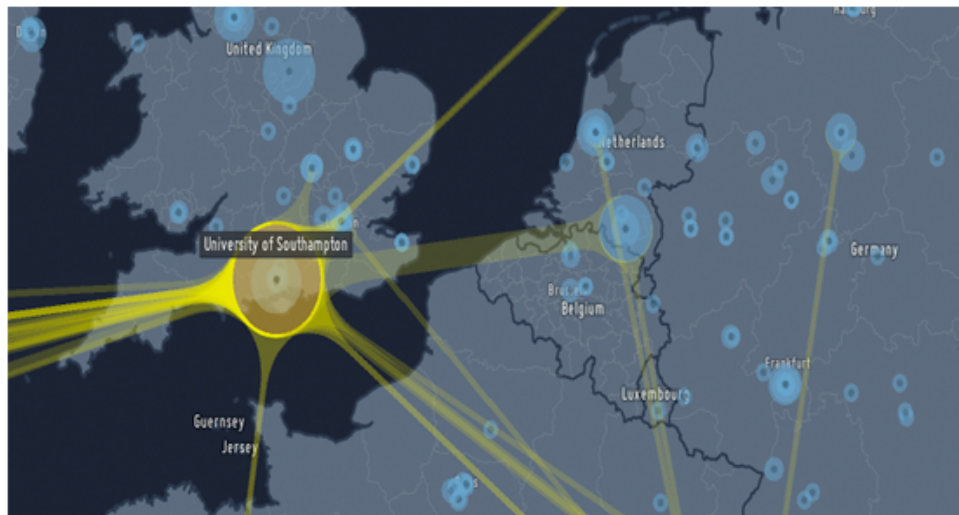
description of functionality	geo-spatial visualization of co-authorship on a multitouch tabletop
URL	http://tillnagel.com/2010/11/muse/, http://tillnagel.com/2011/10/interactive-exploration-of-geospatial-network-visualization/
contact	erik.duval@cs.kuleuven.be
current use	evaluated at ECTEL2010 and Hypertext2011
Bibliography	Nagel, T., Duval, E., Vande Moere, A.: *Interactive Exploration of a Geospatial Network Visualization*. CHI 2012, May 5-10, Austin, USA. Nagel, T., Duval, E.: *Interactive Exploration of a Geospatial Network Visualization*(Poster). VisWeek 2011, October 21-28, Providence, USA. Nagel, T., Duval, E., Heidmann, F.: *Visualizing Geospatial Co-Authorship Data on a Multitouch Table*. Smart Graphics 2011, July 18-20, Bremen, Germany. Nagel, T., Duval, E.: Muse: *Visualizing the origins and connections of institutions based on co-authorship of publications*. Science2.0 for TEL workshop at EC-TEL 2010, Barcelona, Spain.

GlamMap

Figure 1: 7,100 logic books clustered by place of publication and colour-coded by year of publication.

In the Risse visualisation (Figure 1), documents are assigned a geographic location on the basis of their place of publication. The size of the squares represents the *quantity* of documents published in certain geographic areas or places. The colours of the squares range from light blue to dark purple. Differences in colour represent differences in the *year of publication*: documents published in the early eighteenth century are represented by light shades of blue, whereas shades of purple represent documents published in later periods. The size of the squares and the colour scheme thus reveal different types of data.

GlamMap further provides users with *interactive* visualisations. By clicking on the map, users are able to zoom in on certain geographical areas or places. Zooming in refines the view, splitting the squares into multiple smaller squares, until ultimately each square represents a single city. Users can search, filter and group the metadata by clicking on geographic regions or places. This allows them to determine which books were published when and where (Figure 2). In addition, users can search and filter the metadata by specifying the author(s), title and the year of publication of documents.

VALA2014 Conference   3

description of functionality	Geo-spatial visualization of bibliographic metadata
URL	http://glammap.net/
contact	See http://glammap.net/
programming language	See http://glammap.net/
current use	See http://glammap.net/
Deployment	See http://glammap.net/
Bibliography	A. Betti, D.H.P. Gerrits, B. Speckmann & H. van den Berg (2014). GlamMap: visualizing library metadata. In *Proceedings of VALA 2014 – Libraries, Technologies, and the Future*. http://www.vala.org.au/vala2014-proceedings/729-vala2014-session-13-betti .

Awareness tools

TiNYARM

description of functionality	web application to make researchers aware of what their peers are reading
URL	http://atinyarm.appspot.com/

contact	gonzalo.parra@cs.kuleuven.be
licence	Free
programming language	Java (GAE)
current use	HCI group KUL
deployment	http://atinyarm.appspot.com/
bibliography	Parra, G., Klerkx, J., Duval, E.: TiNYARM: This is Not Yet Another Reference Manager. 3rd International Workshop on Motivational and Affective Aspects part of the European Conference on Technology Enhanced Learning (ECTEL) vol: 957 pages:1-4. Parra, G., Klerkx, J., Duval, E.: What should I read next? Awareness of relevant publications through a Community of Practice. In CHI EA '13. ACM, New York, NY, USA, 2375-2376. Parra, G., Klerkx, J., Duval, E.: TiNYARM: Awareness of Research Papers in a Community of Practice. In Proceedings i-KNOW '13). (Accepted).

More!

description of functionality	A social discovery tool for researchers
URL	https://sites.google.com/site/kulmoreapp/
contact	gonzalo.parra@cs.kuleuven.be
licence	Free
programming language	Java, PHP
current use	Currently not used
deployment	http://ariadne.cs.kuleuven.be/more/index.php?id=gonzalo-parra
bibliography	Parra Chico, G., Klerkx, J., Duval, E. (2011). More!: mobile interaction with linked data. In Diaz, P. (Ed.), Hussein, T. (Ed.), Lohmann, S. (Ed.), Ziegler, J. (Ed.), International Workshop on Data-Centric Interactions on the Web in conjunction with the 13th IFIP TC13 Conference on Human-Computer-Interaction (INTERACT 2011). Lisbon, Portugal, 6 September 2011 (pp. 37-47) CEUR-WS. Parra Chico, G., Duval, E. (2010). Filling the gaps to know More! about a researcher. Proceedings of the 2nd International Workshop on Research

2.0. At the 5th European Conference on Technology Enhanced Learning: Sustaining TEL. ECTEL10. Barcelona, Spain, 28 September 2010 (pp. 18-22) CEUR-WS.
Parra Chico, G., Duval, E. (2010). More! a social discovery tool for researchers. Proceedings of EdMedia10: World Conference on Educational Multimedia, Hypermedia and Telecommunications. EdMedia10. Toronto, Canada, 29 June 2010 (pp. 561-569) AACE.

Newspaper Exploration Environment
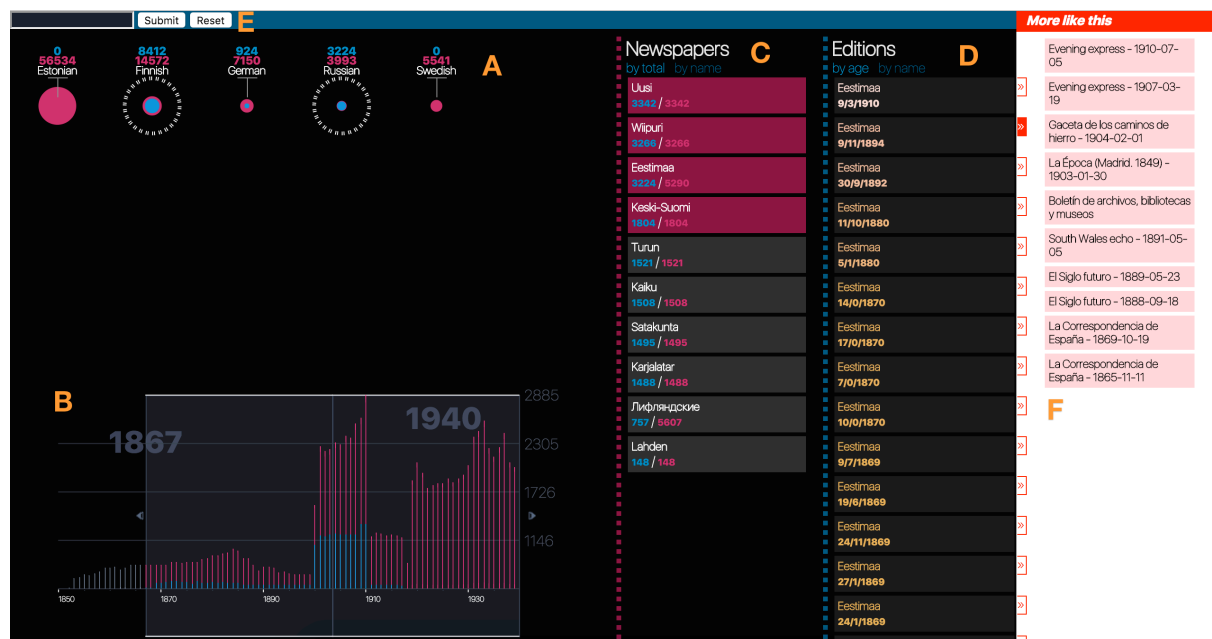
description of functionality	An interactive data visualisation tool to facilitate exploration of digitized newspapers through faceted search
URL	http://daddi.cs.kuleuven.be/ecloud/login
contact	sven.charleer@kuleuven.be
licence	Free
programming language	JavaScript, HTML, CSS, Processing.js, Node.js
current use	Currently not used
deployment	http://daddi.cs.kuleuven.be/ecloud/login

Annotation tools

Annotateit

description of functionality	The Annotator is an open-source JavaScript library and tool that can be added to any webpage to make it annotatable. Annotations can have comments, tags, users and more. Furthermore, the Annotator can be easily extended with new features.
URL	http://annotateit.org/
contact	hello@aroncarroll.com rufus.pollock@okfn.org nick@whiteink.com
licence	GNU GPLv3 or MIT
deployment	http://annotateit.org/

TEXTUS

description of functionality	In a nutshell TEXTUS is an open source platform for working with collections of texts. It enables students, researchers and teachers to share and collaborate around texts by annotating them.
URL	http://textusproject.org/
contact	rufus.pollock@okfn.org
licence	MIT
current use	Open Shakespeare group (http://openliterature.net/shakespeare/)
deployment	http://beta.openphilosophy.org

Pundit

description of functionality	Pundit enables users to create structured data annotating web pages or images, collect annotations and share with others to create collaborative structured knowledge. Furthermore, Pundit already has a built in entity extraction feature which annotates the texts using knowledge bases like DBpedia or Freebase
URL	http://thepund.it/
contact	pundit@netseven.it
current use	Wittgenstein group in Norway (via DM2E project)
deployment	http://release-bot.thepund.it/latest/examples/authors_index.html
bibliography	M.Grassi, C. Morbidoni and M. Nucci. "A Collaborative Video Annotation System Based on Semantic Web Technologies". In press: Cognitive Computation. Springer-Verlag, Berlin. M. Grassi, C. Morbidoni, M. Nucci, S. Fonda, G. Ledda. "Pundit: Semantically Structured Annotations for Web Contents and Digital Libraries". Annett Mitschick, Fernando Loizides, Livia Predoiu, Andreas Nürnberger, Seamus Ross (eds.): Semantic Digital Archives 2012. Proceedings of the Second International Workshop on Semantic Digital Archives (SDA 2012), Paphos, Cyprus, September 27, 2012, CEUR-WS.org/Vol-912, urn:nbn:de:0074-912-6. M. Nucci, M. Grassi, C. Morbidoni, F. Piazza. "Enriching Digital Libraries Contents with SemLib Semantic Annotation System". Proceedings of the Digital Humanities 2012 Conference, Hamburg, Germany, 16th-20th July, 2012. C. Morbidoni, M. Grassi, M. Nucci. "Introducing SemLib Project: Semantic Web Tools for Digital Libraries", International Workshop on "Semantic Digital Archives - sustainable long-term curation perspectives of Cultural Heritage" held as part of the 15th International Conference on Theory and Practice of Digital Libraries (TPDL). 29.09.2011 in Berlin.

OpenAnnotation

A collaboration tool that aims:

- To facilitate the emergence of a Web and Resource-centric interoperable annotation environment that allows leveraging annotations across the boundaries of annotation clients, annotation servers, and content collections. To this end, interoperability specifications will be devised.

description of functionality

- To demonstrate through implementations an interoperable annotation environment enabled by the interoperability specifications in settings characterized by a variety of annotation client/server environments, content collections, and scholarly use cases.

- To seed widespread adoption by deploying robust, production-quality applications conformant with the interoperable annotation environment in ubiquitous and specialized services, tools, and content used by scholars -- e.g.: Zotero, AXE, LORE, Co-Annotea, Pliny; JSTOR, AustLit, MONK.

URL http://www.openannotation.org/

DocumentCloud

description of functionality	DocumentCloud is a tool for annotating documents and sharing those annotations on the web. DocumentCloud runs every document you upload through OpenCalais and extracts entities (people, places and organizations) mentioned in it.
URL	https://www.documentcloud.org/
contact	support@documentcloud.org
deployment	https://www.documentcloud.org/

Researchr

description of functionality	Academic information management workflow, which is at the same time individual and happens on the local computer, but also set up to be easily shared with others. It's really a whole framework with a bunch of applications (the key ones being BibDesk, Skim, DokuWiki, Chrome).
URL	http://reganmian.net/wiki/researchr:start
contact	shaklev@gmail.com
licence	Free
deployment	http://reganmian.net/wiki/researchr:start

10. *Appendix: Additional tools for musicologists*

Search tools

IMSLP Petrucci Music

description of functionality	The Petrucci Music Library (after publisher Ottaviano Petrucci) is a virtual library of public domain music scores. Since its launch on February 16, 2006, over 258.000 scores and 23.000 recordings for over 73.000 works by 7.000 composers have been uploaded (December 2013).
URL	http://imslp.org/wiki/
contact	http://imslpforums.org/ | http://imslp.org/wiki/IMSLP:IRC_Access
licence	(CC BY-SA 4.0)
deployment	http://imslp.org

Digital Image Archive of Medieval Music (DIAMM)

description of functionality	DIAMM (the Digital Image Archive of Medieval Music) is a resource for the study of medieval manuscripts. The project presents images and metadata for thousands of manuscripts. DIAMM also provides a home for scholarly resources and editions, undertakes digital restoration of damaged manuscripts and documents, publishes high-quality facsimiles, and offers expertise through consultanting.
URL	http://www.diamm.ac.uk/ http://imslp.org
contact	http://www.diamm.ac.uk/about/contact-diamm/
license	http://www.diamm.ac.uk/about/copyright/
deployment	http://www.diamm.ac.uk/ https://www.documentcloud.org/

Creation of music sheets

MuseScore
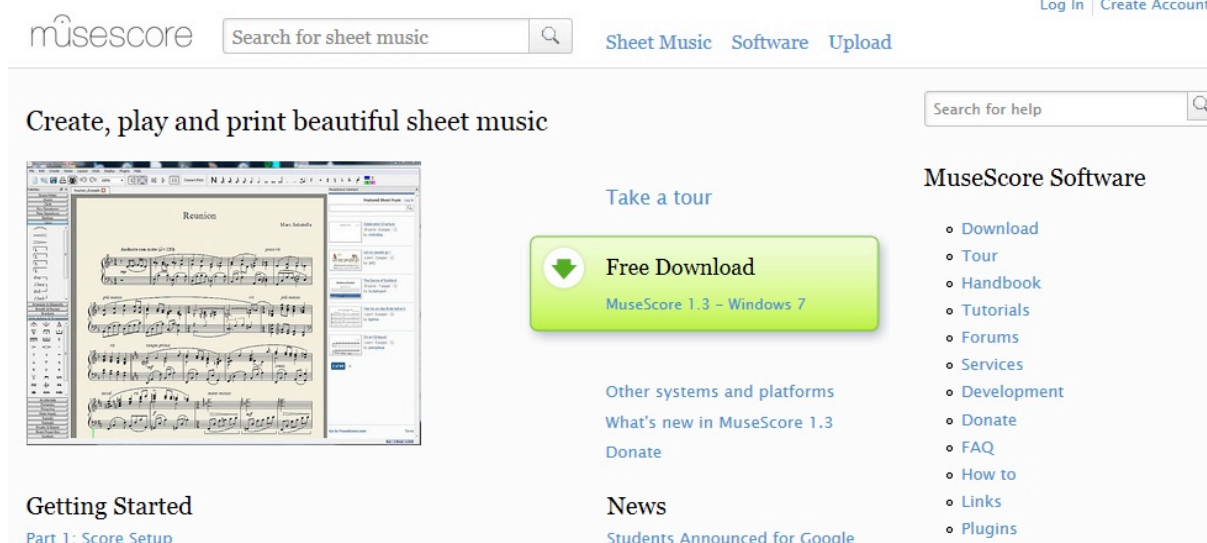
description of functionality	MuseScore is a free and open source music notation software for Windows, Mac and Linux. It is easy to use and makes beautiful sheet music. MuseScore is often praised as a cost effective alternative for Sibelius and Finale.
URL	https://musescore.com/
contact	https://musescore.com/contact
licence	Copyright ©
deployment	https://github.com/musescore/MuseScore
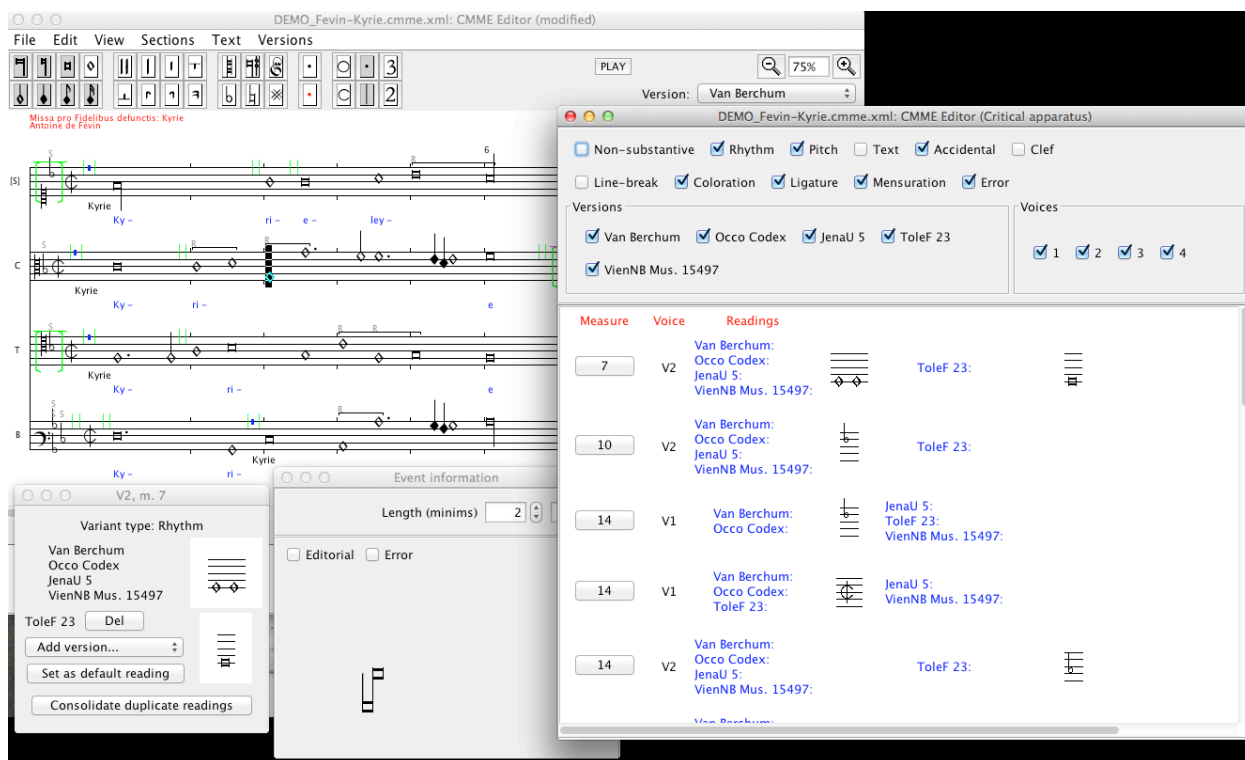
CMME

description of functionality	CMME offers software for creating, editing and viewing encoded score files of mensural notated music (roughly from the 15th and 16th

century). The software is written in Java, and platform independent. The underlying encoding scheme is CMME-XML, a variant of the MusicXML format. On its website, CMME offers a corpus of free scores with additional metadata.

URL	http://www.cmme.org
contact	http://www.cmme.org/contact
license	GNU General Public License
deployment	https://github.com/tdumitrescu/cmme-editor https://www.documentcloud.org/

Optical Music Recognition

Aruspix
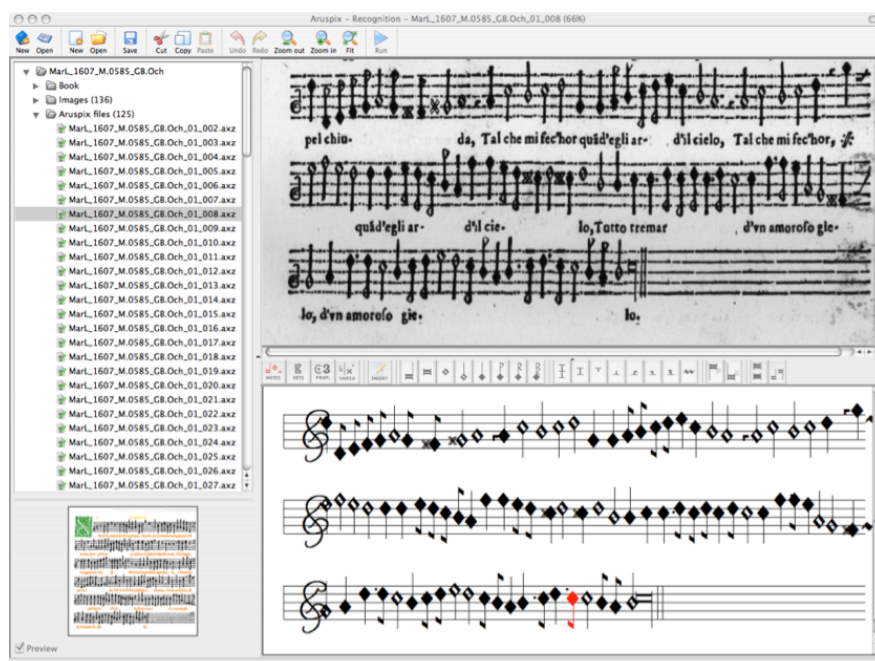
description of functionality	Aruspix is a software application for the optical recognition, the superimposition and the collation of early music prints (16th century). The optical recognition part of Aruspix uses an innovative technique in OMR based on Hidden Markov Models to identify and classify the particular musical characters found in early music prints, after an initial pre-treatment phase performed on scanned images of microfilm and facsimiles. The integrated music editor allows corrections of recognition errors. One original aspect of Aruspix is that it learns and optimizes itself dynamically as soon as a page is processed and corrected by the user.
URL	http://www.aruspix.net/index.html
contact	lxpugin@gmail.com	laurent@music.mcgill.ca

license	N/A
deployment	N/Ahttps://www.documentcloud.org/

Classification tools

WEKA

description of functionality	Weka is a collection of machine learning algorithms for data mining tasks. The algorithms can either be applied directly to a dataset or called from Java code. Weka contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization. It is also well suited for developing new machine learning schemes.
URL	http://www.cs.waikato.ac.nz/ml/weka/
contact	http://list.waikato.ac.nz/mailman/listinfo/wekalist
license	GNU General Public License
deployment	http://weka.wikispaces.com/wiki/changes https://www.documentcloud.org/

Music Analysis

Music21

description of functionality	Music21 is a Python-based toolkit for computer-aided musicology. Applications of this toolkit include computational musicology, music information, musical example extraction and generation, music notation editing and scripting, and a wide variety of approaches to composition, both algorithmic and directly specified.
URL	http://web.mit.edu/music21/doc/about/index.html
contact	http://web.mit.edu/music21/doc/about/about.html
license	GNU Public License
deployment	http://web.mit.edu/music21/doc/developerReference/index.html

The Humdrum Toolkit

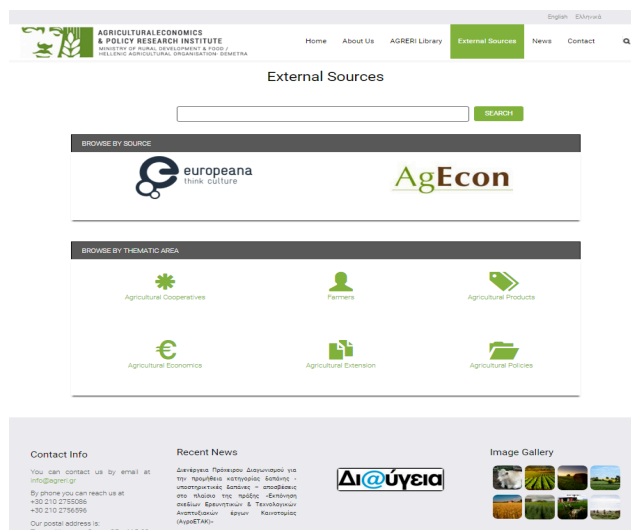| description of functionality	Humdrum is a set of command-line tools that facilitates musical analysis, as well as a generalized syntax for representing sequential streams of data. As a set of command-line tools, it is programming language agnostic. Many have employed Humdrum tools in larger scripts that use PERL, Ruby, Python, Bash, LISP, and C++.
URL	http://www.humdrum.org/
contact	https://groups.google.com/forum/#!forum/starstarhug
license	N/A
deployment	http://web.mit.edu/music21/doc/developerReference/index.html https://www.documentcloud.org/

11. Appendix: Tools for Agricultural Researchers

Search tools

AGRERI Discovery Microsite

description of functionality	Lightweight search interface (implemented as an HTML page with some Javascript) that works on top of a Solr index. The Solr index is build using a number of metadata facets that can be used to navigate metadata records aggregated through different sources. Ingests metadata records of relevance to the specific users, by a number of ingestion mechanisms/APIs that are creating the Solr index behind the particular search page.
URL	http://www.agreri.gr
contact	nikosm@ieee.org stoitsis@ieee.org
licence	N/A
programming language	HTML, CSS, HTTP, Javascript, AJAX
current use	Evaluated with agricultural economics reseachers, with estimated usage by >200 users
deployment	http://www.agreri.gr/external_library/browse [looking into the Europeana and AgEcon aggregators]