



DELIVERABLE

Project Acronym: Europeana Cloud
Grant Agreement number: 325091
Project Title: Europeana Cloud: Unlocking Europe's Research via The Cloud

D3.2 Tools and services: A set of tools and services for researchers that exploit Europeana content

Revision: v2

Authors:

Erik Duval (KU Leuven)
Gonzalo Parra (KU Leuven)
Anja Jentsch (Open Knowledge Foundation Deutschland)
Hein van den Berg (Stichting VU-VUMC)
Giannis Stoitsis (ARIADNE Foundation)
Katerina Galani (ARIADNE Foundation)
Nikos Marianos (ARIADNE Foundation)
Andreas Drakos (ARIADNE Foundation)
Marnix van Berchum (KNAW-DANS / Utrecht University)

Project co-funded by the European Commission within the ICT Policy Support Programme		
Dissemination Level		
P	Public	x
C	Confidential, only for members of the consortium and the Commission Services	

Revision History

Revision	Date	Author	Organisation	Description
V1	19.01.2015	Gonzalo Parra	KU Leuven	Adding the structure to include the work from Year 2
V2	19.01.2015	Andreas Drakos	ARIADNE	Update of Ariadne Finder section
V3	29.01.2015	Anja Jentsch	OKDE	Input for Aruspix service
V4	29.01.2015	Gonzalo Parra	KU Leuven	Input in Music21 service, updates in Activity Stream section and minor edits
V5	30.01.2015	Hein van den Berg	VU	Corrections and minor edits
V6	30.01.2015	Marnix van Berchum	DANS-KNAW / Utrecht University	Minor edits on musicological parts
V7	01.02.2015	Erik Duval	KU Leuven	Consolidation, minor edits
V8	02.02.2015	Anja Jentsch	OKDE	Input for adjusted TimeMapper

Statement of originality:

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.



D3.2 Tools and services

A set of tools and services for researchers that exploit Europeana content

Executive summary

This deliverable consists of software.

The final version of this deliverable is due in M30. The current version is the M24 intermediate version.

This document describes very briefly the software and includes references to where the software can be accessed on the Web.

Table of Contents

Introduction	3
The Ariadne Finder	4
The TimeMapper	8
The Activity Stream	10
Aruspix	12
Music21	14
References	15

List of figures

Figure 1: Screenshots from the Ariadne Finder main page (a), listing results (b), customized categories (c) and view item (d)	5
Figure 2: Screenshots from the Ariadne Finder main page (a), listing results (b), view item (c)	6
Figure 3: Screenshots from the TimeMapper showing the resources matching the search term “Kant” from the ARIADNE Finder	8
Figure 4: Figure 4: Screenshot of the TimeMapper showing resources matching the search term “Gardano” in the ARIADNE Finder	9
Figure 5: The Activity Stream main screen	10
Figure 6: Information sources and outlets of the Activity Stream	11
Figure 7: Aruspix desktop application	12
Figure 8: The Music21 web application	14
Figure 8: The Music21 processed results	14

Introduction

Work Package 3 aims to develop services and tools that leverage Europeana content in the Europeana Cloud for researchers. During the first months, the work of WP3 focused on the development of personas, scenarios and use cases, in order to understand and analyse the user needs. This initial work on personas, scenarios and use cases was reported in D3.1.

As described in the DoW, the work of WP3 proceeds in yearly cycles that target a specific research community of users:

- year 1 focused on the DM2E project (and more specifically the Wittgenstein archives at the University of Bergen) and the Axiom philosophy group at the VU University Amsterdam
- year 2 targeted a research community of musicologists that focus on Early Music.

We developed and evaluated a prototype demonstrator that loosely integrates a variety of tools relevant for the target audience. These tools are:

- the ARIADNE finder: a personalized search micro site to help researchers search and find content related to their work coming from Europeana and other external sources [5];
- the TimeMapper: an integrated tool to visualize search results on a timeline and on an interactive map, so that users can further filter the content and get a better overview of the different resources found on Europeana;
- an Activity Stream, integrated with the other tools, to capture and present the different actions taken in this process (search, visualize, explore, annotate, download) [1],
- Aruspix, an optical music recognition (OMR) tool that scans early music prints and transcribes them using the MEI standard [4],
- Music21, a toolkit for computer-aided musicology to parse, analyse and process encoded scores [3].

In this document, we present a very short description of each of these tools, with a reference to where the software can be accessed on the web.

We have also evaluated the tools: that evaluation is the subject of D3.3.

The ARIADNE Finder

The first tool is the ARIADNE Finder, a personalized micro-site that can be used by users to search and discover resources. The Finder searches predefined collections of datasets, as indicated by user input, and presents the results in a uniform way. It comes as a micro-site, built with lightweight web-technologies (HTML, CSS, HTTP, Javascript, AJAX) in order to be easily embedded in sites and web-applications, without the need to make changes for matching the existing technologies of the application. In the context of the eCloud project and WP3, prototypes of the Finder have been deployed to showcase how researchers can

- search and access resources coming mainly from Europeana,
- have integrated access to all other tools.

First Year Deployment

The first prototype of the Finder in WP3 was designed based on the needs of the Axiom philosophy group.

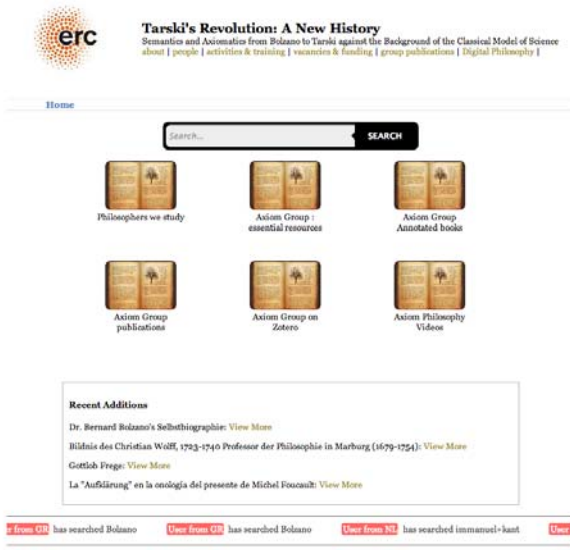
The Finder is a personalized tool in the following two ways: it is graphically designed to be smoothly integrated with the web site of the Axiom group (<http://axiom.vu.nl/>) and it is built on top of collections that have been requested by stakeholders.

The main usage of the ARIADNE Finder is a faceted search interface that allows users to search and quickly filter the results. In addition, predefined categories that allow access over specific content (i.e. philosophers studied by the target audience) are also available.

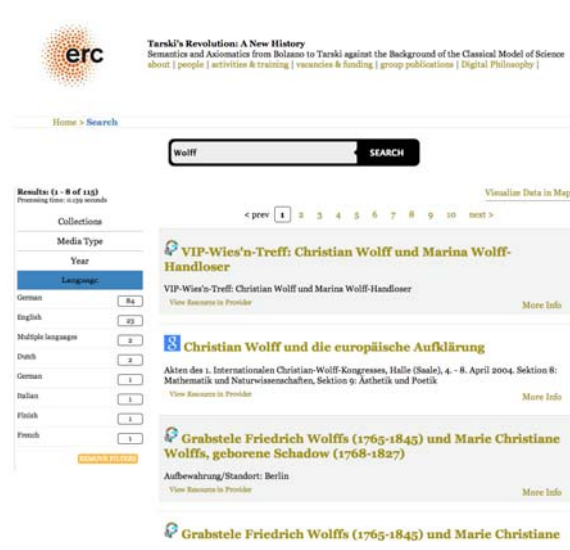
The prototype has been designed and developed with the constant feedback from the Axiom group in order to better catch and cover their needs. During development, a number of discussions were organized between the group and WP3 in order to gather feedback concerning the collections to search, categories to use, and the facets that the stakeholders would like to use.

In order for the Finder to allow faceted search and uniform representation of the metadata from resources coming from different collections, the Finder uses the existing ARIADNE infrastructure to store a repository with all the metadata. In the current development phase, the resources and different collections stored in the repository are limited to two: Europeana and Google Books. In both cases, an API is used to filter thematically resources for the dataset. In order to provide this uniform representation and make the resources available through the Finder a transformation process took place, where all files were transformed from their original scheme to an internal format. During this transformation procedure, metadata has also been enriched.

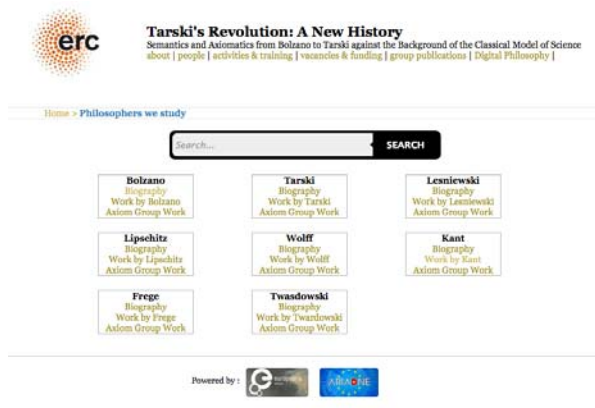
Finally, the Finder is the main tool on top of other tools that have been integrated with it. On the first page of the Finder, the Activity Stream is integrated in the bottom screen as a floating message, while in the listing of the results a link to visualize the results in the TimeMapper is available. More information on these tools is available in the next sections. For the integration of the different tools, a number of technical issues had to be resolved, such as passing the POST search activities in the REST service of the activity stream, passing the JSON file to the TimeMapper, etc. Apart from online meetings and, when needed, bilateral communication, a specific WP3 discussion board was kept to discuss and resolve all technical issues.



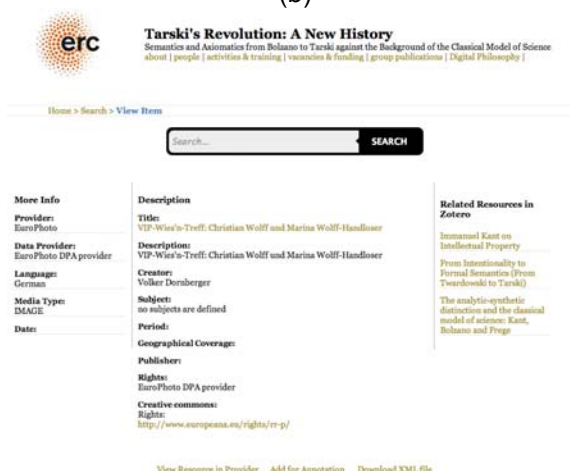
(a)



(b)



(c)



(d)

Figure 1: Screenshots from the Ariadne Finder main page (a), listing results (b), customized categories (c) and view item (d)

The above figures show different screenshots from the Finder. In Figure 1-a the main page of the Finder is shown, with the Activity Stream at the bottom. Figure 1-b shows the listing of the results after a search is executed, with the facets that can be used and the button to visualize the results in the TimeMapper. Figure 1-c shows the menu of the predefined customized categories for quick access to specific results. Finally, figure 1-d presents how a specific result can be seen.

The ARIADNE Finder for the Axiom philosophers group can be accessed at the following url: <http://greenlearningnetwork.com/axiom/>

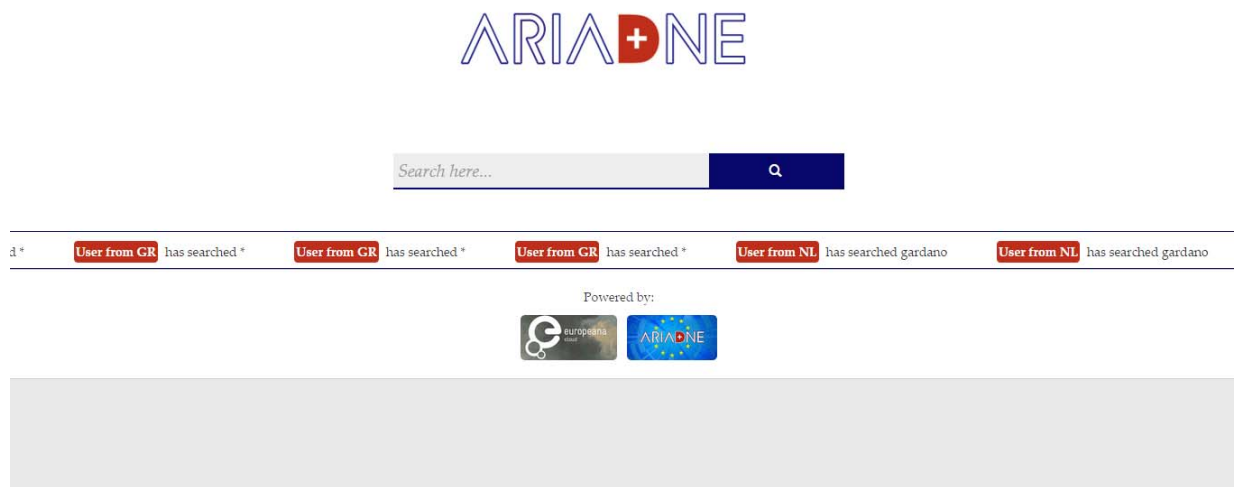
Second Year Deployment

During the second year mash-up, the Finder's prototype was designed based on the needs of a group of researchers studying early music. Following the iterative process of WP3, for the design and deployment of the Finder, a series of meetings with WP3 partners and musicology researchers took place. Design feedback was collected and the appropriate requirements for additional collections of interest to integrate in the Finder were determined.

The second year version of the Finder includes a more simplified user interface of the Finder where there are no predefined categories available on the home screen. Instead, a list of four search facets (i.e. provider, media type, language, and year) is available on the screen where the search results are returned. The Finder includes resources coming from two different collections, Europeana and RISM (<http://www.rism.info/>). The metadata was stored in the existing ARIADNE infrastructure and provided the user with a single search interface.

For the ARIADNE team, the integration of the RISM collection was a great challenge during the second year deployment. The data covered by the RISM collection (mainly scores) are heterogeneous and quite different from the ones provided by Europeana. To allow the integration with the Finder and the visualisation of the results in a uniform way, transformation of the metadata to an ARIADNE internal format was required. A second obstacle of the RISM integration was related to the RISM project itself. With the project having a very different time schedule than eCloud and in the process of providing open access data and linked data, the synchronization of our efforts was at many times difficult. During the development, quite often linking to the actual resource was rather complicated.

As in the first year, the Finder is used as the 'baseline' tool for the integration of the additional tools from WP3. Both the Activity Stream and the TimeMapper are integrated in the Finder to see the past activities (i.e. searches) and to visualise the results. When viewing a search result, the connection to Music21 is also available.



(a)

ARIADNE

Search here...

Active Facets

Search for "16" returned total: 1179 results

Provider

Media Type

IMAGE

TEXT

Language

Year

Cantiones sacrae liber 6 [Musikdruck]
Ausgabe : [Stimmen]

Six duos pour deux violons, oeuvre VIII
Dedicado a: "Dediés a Monsieur le Baron de Bagge"

súbor partov; Missa ex C
prirastková kniha: MUS IX 1-438 Kežmarok Hudobná škola a farský kostol. Notový materiál je evidovaný pod 438 signatúrami, z toho rukopisov 326, tlačí 118 (tlače s príslušným rukopismi tvoria vždy jednu signatúru).-br>prirastková kniha: Lokalita: ...

(b)

ARIADNE

Search here...

Il primo libro de ricercari da cantare a quattro voci. Tenore ([Reprod.] / di Claudio Merulo da Corregio

Description :
Collection : Italian books before 1601 ; 154.2

Organization :
National Library of France

Contributor :
appresso di li figliuoli di Antonio Gardano (In Venetia)

Creator :
Merulo, Claudio (1533-1604)

Rights :
<http://www.europeana.eu/rights/rr-f/>

Language:
Italian

Location:
france

Type:
monographie imprimée, printed monograph

Content Type:

(c)

Figure 2: Screenshots from the Ariadne Finder main page (a), listing results (b), view item (c)

In Figure 2 above, there are different screenshots from the Finder. Figure 2-a shows the main page of the Finder with the Activity Stream at the bottom. In Figure 2-b the listing of the results after a search is shown, with the facets that can be used. Figure 2-c shows how a specific result can be seen.

The ARIADNE Finder for the Musicologists group can be accessed at <http://greenlearningnetwork.com/cmme-finder/>

The TimeMapper

Europeana provides a variety of metadata for its resources. These might include images, geo-coordinates and time information. Having this metadata available, TimeMapper can help to visualize temporal and geographical overlap and dependencies of resources.

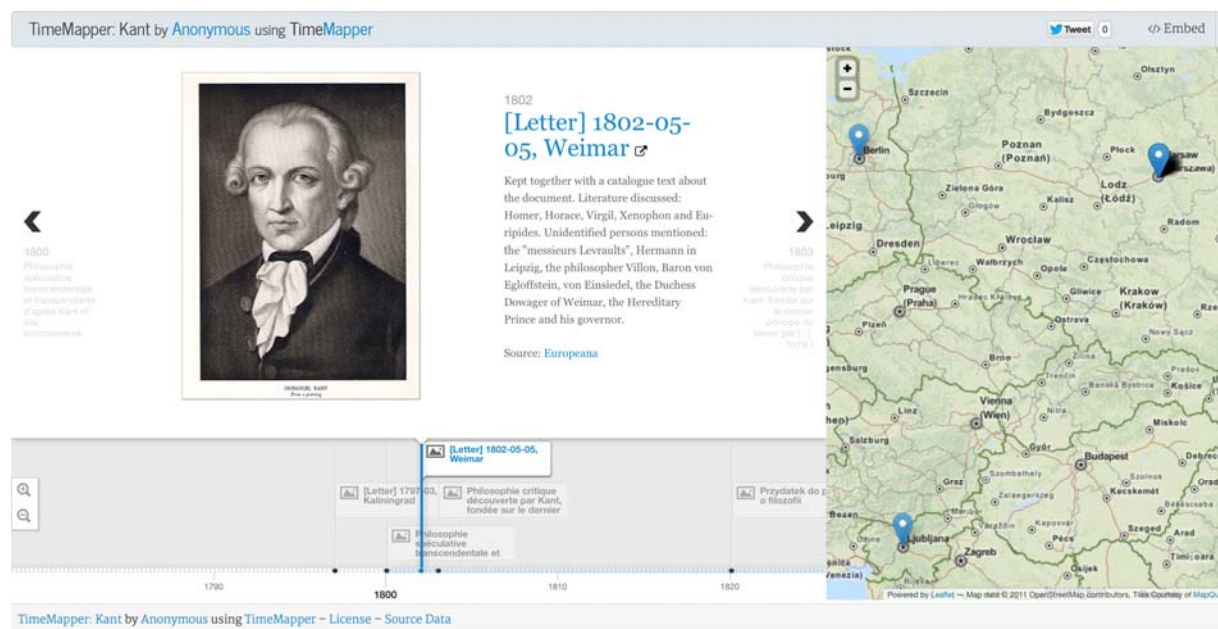
TimeMapper¹ is a data visualization tool that allows for the creation of timelines and timemaps using Google spreadsheets. While the ARIADNE Finder provides the user with a faceted search for Europeana resources, it might still be difficult to navigate through large amounts of records for specific searches. For example, the search for “Kant” returns 158 results.² We integrated TimeMapper in our tool chain to provide an interactive geo-spatial visualization of these bibliographic metadata. This enables users to quickly navigate the metadata and to order resources on the basis of time and place of publication. By doing this they can easily identify various (types of) resources worth studying.

Figure 3 shows the TimeMapper when drilling down into resources that match the keyword “Kant”³. Ultimately, one would also want to be able to visualize different persons and compare their work based on geographical and temporal metadata.

TimeMapper was adjusted for Europeana Cloud to use the JSON format defined by the ARIADNE Finder. It will furthermore be adjusted to meet the specific needs that were pointed out by our user group in a recent study (cf. D3.3).

TimeMapper is available under the MIT licence⁴. The tool can be accessed via the ARIADNE Finder’s button labelled “Visualize data on map” for the year 1 version and via the “View in TimeMapper” button for the year 2 version.

For the second year deployment, more metadata was included into the TimeMapper which allows for documents by several authors being displayed in one timemap⁵ (cf. Figure 4).



¹ <http://timemapper.okfnlabs.org>

² <http://greenlearningnetwork.com/axiom/listing.html?query=kant#>

³ <http://ecloud.okfn.de/timemapper/index.php?search=kant>

⁴ <http://opensource.org/licenses/MIT>

⁵ <http://timemapper.okfnlabs.org/anon/12s4kd-early-music>

Figure 3: Screenshot of the TimeMapper showing resources matching the search term “Kant” in the ARIADNE Finder

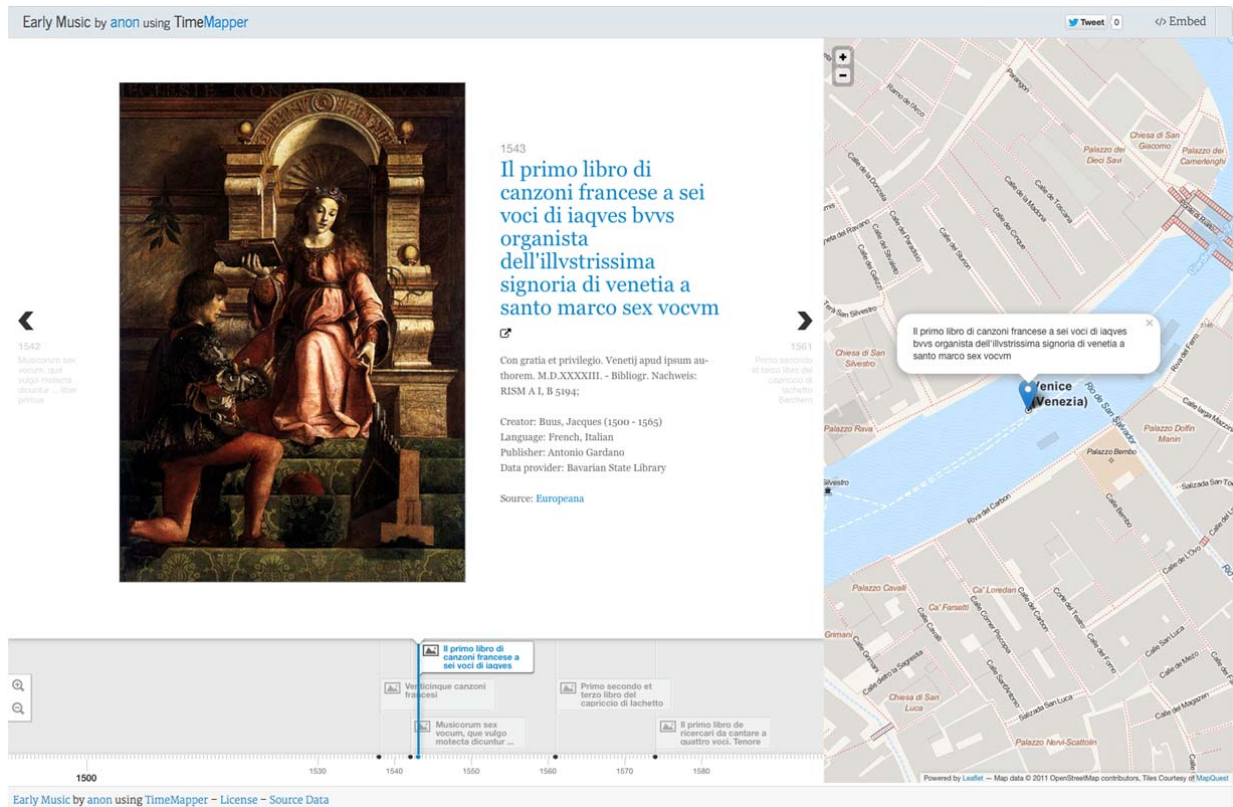


Figure 4: Screenshot of the TimeMapper showing resources that are published by Gardano based on the search results in the ARIADNE Finder

The Activity Stream

Based on the community reading awareness provided by TiNYARM [1] and supporting the Science 2.0 idea of enhancing collaboration among researchers [2], we have deployed a web application called the “Activity Stream (AS)”; enabling researchers to share their work related activities with a community. Specifically, the application aims to aggregate “search” and “visualize” activities, and make researchers aware of what their peers are currently working on.

In the first prototype, the AS presents “searches” that have been performed using the ARIADNE Finder and terms that have been “visualized” using the TimeMapper, as seen in Figure 5. The activities in the stream are structured as: Actor | verb | (Object). For example, User from GR | has searched | Bolzano. For the second year, 2 activities were added to the activity stream: interpretation and processing. These represent the usage of the Aruspix and Music21 components.

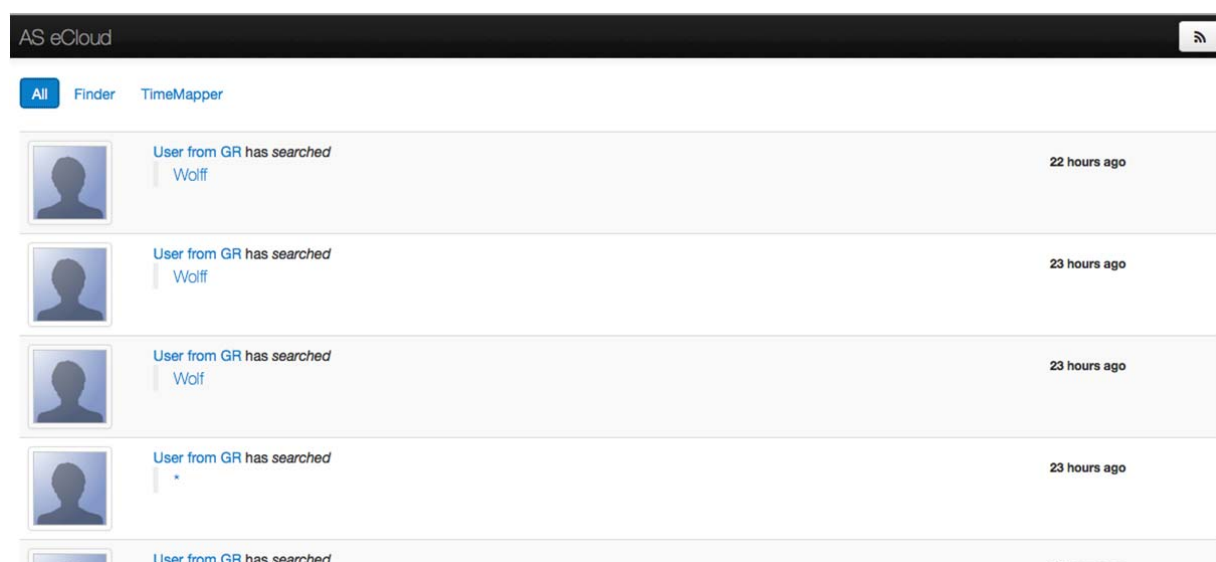


Figure 5: The Activity Stream main screen

The Activity Stream is designed as a web application (using HTML and JavaScript) and deployed using the Google App Engine (GAE). Together with the terms used to perform a search or visualization, a link to the tool showing the outcome of that action is provided. Also, to be able to provide users the flexibility to filter activities, the feature of tool grouping was added to the application. For instance, by clicking on the tool’s name (e.g.: Finder or TimeMapper) you can get the stream of activities from that tool only.

The Activity Stream allows us to digest different events sent from different tools (via REST services) used by researchers, but also provides the possibility to embed these on other software components. For example, the application supports RSS syndication as a passive form of notification system. Figure 6 illustrates the current activity sources and outlets.

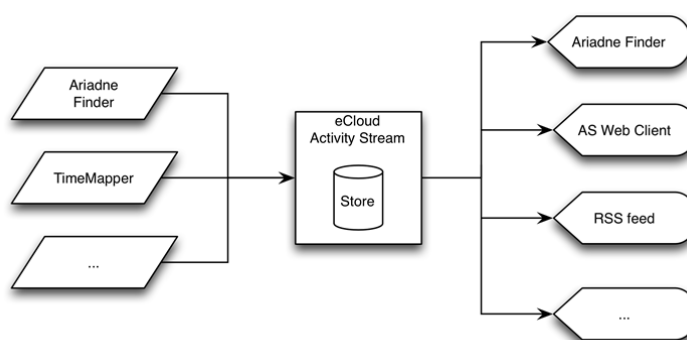


Figure 6: Information sources and outlets of the Activity Stream

The Activity Stream for the Axiom philosophers group can be accessed at <http://as-ecloud.appspot.com/> , and the Activity Stream of the second year can be found at <http://as-ecm.appspot.com/>

Aruspix

Aruspix⁶ is an optical music recognition (OMR) tool that scans early music prints, transcribes them and encodes them into the MEI standard⁷.

While there are other OMR tools for new music available, Aruspix is the only tool to our knowledge being able to handle scores printed in the 16th and 17th centuries with movable typefaces. Such scores are often difficult to examine with existing superimposition and optical recognition software, as they present a number of specific layout and format problems and are quite often in a deteriorated state because of their age.

The printing techniques of that time mean that differences can exist between copies produced in the same print run, and comparison of these copies by superimposition can enable more accurate critical editions to be prepared. Digitising the scores through optical recognition can enable us to collate different editions regardless of layout, and is also useful in the preparation of digital music libraries, for example.

Aruspix is available under the GNU Public Licence.

Figure 7 shows the desktop application for Mac OS which allows for an interactive score transcription.

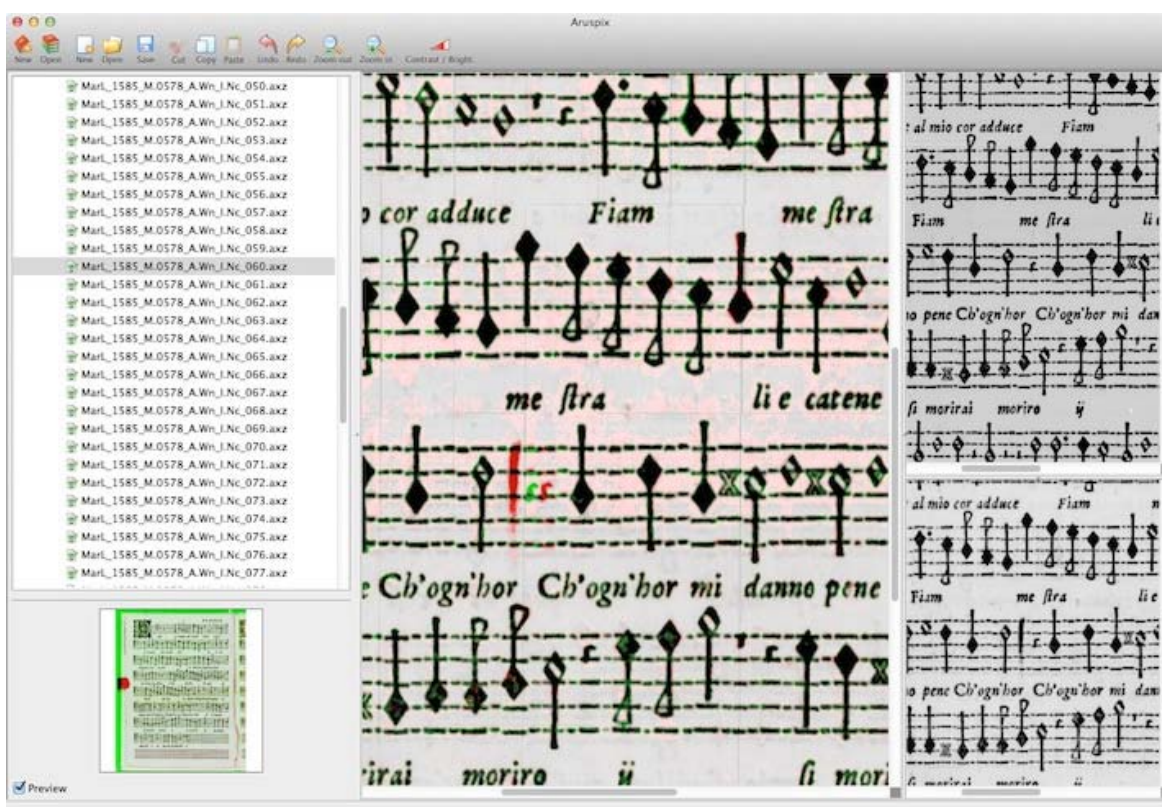


Figure 7: Aruspix desktop application

For eCloud, we use the command line version that automatically converts scores to MEI files in a page-wise fashion. We then need to combine the pages into a single score again. Moreover, the MEI version being used by Aruspix is a new and not yet standardized one.

Since Music21 (see next section) needs MEI as of the 2012 or 2013 specification, we wrote an XSLT program to transform the MEI files.

⁶ <http://www.aruspix.net>

⁷ <http://www.music-encoding.org/exist/apps/mei/home>

The command line version sends requested score transcriptions to the Music21 service for further analysis. Furthermore, it sends activity on transcribed scores to the Activity Stream.

Music21

Music21⁸ is a Python object-oriented toolkit for computer-aided musicology that allows music information, extraction and generation, together with music notation editing and scripting in symbolic (score-based) forms. The toolkit is able to import different formats, such as the MusicXML and Music Encoding Initiative (MEI) standards.

In the Year 2 of the eCloud project, we extended the Music21 web application module in order to provide parsing and processing requests to a Music21 installation running on a server. In the workflow, Music21 is used after the Aruspix service has created an MEI version of a score. With an MEI file, a specific set of actions becomes available to the musicologists: calculation of 'Parts and Measures', calculation of the 'Pitch ranges' and requesting the 'legal melodic intervals' of a score. The main screen of the Music21 web interface is shown in Figure 8.

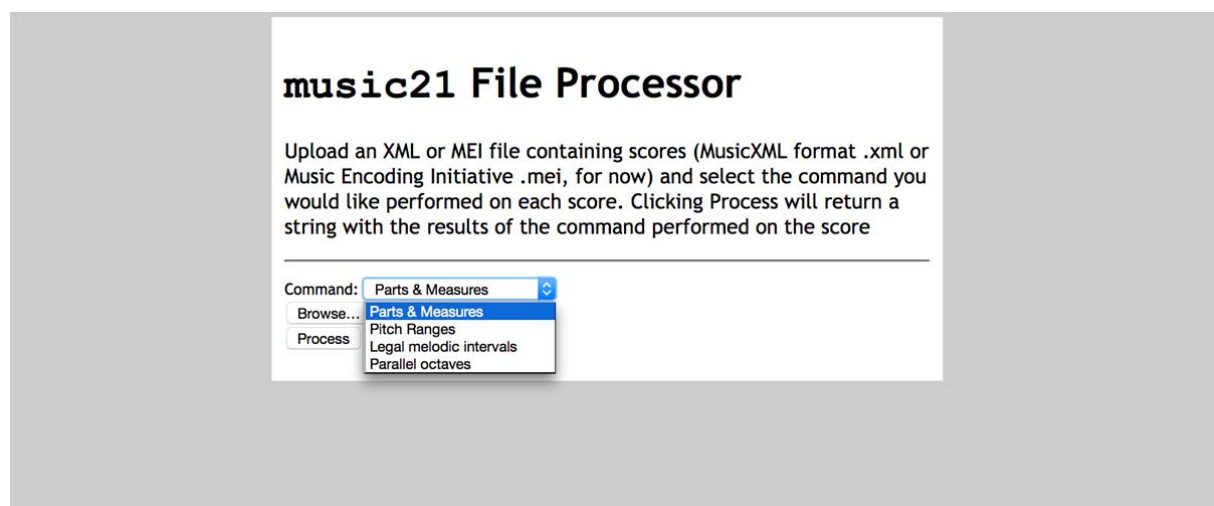


Figure 8: The Music21 web application

Figure 9 presents the different results obtained after processing the encoding of 'Altenburg - Ein feste Burg' encoding: parts, measures and notes and the legal melodic intervals of the measures inside the MEI.

Altenburg_Ein_feste_Burg.mei	
<p>Altenburg_Ein_feste_Burg.mei</p> <p>Parts: 6</p> <p>Part 1: 40 Measures</p> <ul style="list-style-type: none"> • Measure 1 (6 Note(s)) - Time Signature(s): 1 (4/4) Key Signature: 1(, 'major') • Measure 2 (6 Note(s)) • Measure 3 (7 Note(s)) • Measure 4 (4 Note(s)) • Measure 5 (7 Note(s)) • Measure 6 (7 Note(s)) • Measure 7 (7 Note(s)) • Measure 8 (4 Note(s)) • Measure 9 (3 Note(s)) 	<p>Altenburg_Ein_feste_Burg.mei</p> <p>Part 1: 40 Measures</p> <ul style="list-style-type: none"> Measure 1 (6 Note(s)) <ul style="list-style-type: none"> ◦ False : Note 3 (E) and Note 4 (E) Measure 2 (6 Note(s)) Measure 3 (7 Note(s)) <ul style="list-style-type: none"> ◦ False : Note 3 (C) and Note 4 (C) Measure 4 (4 Note(s)) Measure 5 (7 Note(s)) Measure 6 (7 Note(s)) Measure 7 (7 Note(s)) Measure 8 (4 Note(s)) Measure 9 (3 Note(s)) Measure 10 (3 Note(s)) <ul style="list-style-type: none"> ◦ False : Note 2 (F) and Note 3 (F) Measure 11 (3 Note(s)) Measure 12 (8 Note(s)) Measure 13 (7 Note(s)) Measure 14 (3 Note(s)) <ul style="list-style-type: none"> ◦ False : Note 1 (C) and Note 2 (C) ◦ False : Note 2 (C) and Note 3 (C)

Figure 9: The Music21 processed results

⁸ <http://web.mit.edu/music21/>

References

1. G. Parra, J. Klerkx, and E. Duval. Tinyarm: Awareness of research papers in a community of practice. In Proceedings of the 13th International Conference on Knowledge Management and Knowledge Technologies, page 21. ACM, 2013.
2. Shneiderman, B. 2008. Science 2.0. Science. Vol. 319, No. 5868, 1349-1350.
3. Cuthbert, M. S., Ariza, C. music21: A Toolkit for Computer-Aided Musicology and Symbolic Music Data. In 11th International Society for Music Information Retrieval Conference (ISMIR 2010), August 9-13, 2010, Utrecht, Netherlands. pp. 637-642.
4. G. Haus and M. Longari. A multi-layered, timebased music description approach based on xml. Computer Music Journal, 29(1):70–85, 2005.
5. L. Pugin, J. A. Burgoyne, and I. Fujinaga. Goal-directed evaluation for the improvement of optical music recognition on early music prints. In Proceedings of the 7th ACM/IEEE-CS joint conference on Digital libraries, pages 303–304. ACM, 2007.
6. K. Makris, G. Skevakis, V. Kalokyri, P. Arapi, S. Christodoulakis, J. Stoitsis, N. Manolis, and S. L. Rojas. Federating natural history museums in natural europe. In Metadata and Semantics Research, pages 361–372. Springer, 2013.