**ECP-2008-DILI-528001**

# EuropeanaConnect

D3.3.3 – Tested prototype ready for deployment in release 1.2–1.6 of Europeana

| | |
|---|---|
| **Deliverable number/name** | *D3.3.3* |
| **Dissemination level** | *Public* |
| **Delivery date** | *19/09/2011* |
| **Status** | |
| **Author(s)** | *Ralf Stockmann* |

**eContent*plus***

This project is funded under the *e*Content*plus* programme,
a multiannual Community programme to make digital content in Europe more accessible, usable and exploitable.

Österreichische Nationalbibliothek EuropeanaConnect is coordinated by the Austrian National Library

# D3.3.3 – Tested prototype ready for deployment in release 1.2–1.6 of Europeana

co-funded by the European Union

The project is co-funded by the European Union, through the ***eContentplus*** programme

http://ec.europa.eu/econtentplus

Österreichische Nationalbibliothek

EuropeanaConnect is coordinated by the Austrian National Library

# Table of Contents

# 1 Introduction

Although the amount and types of data available through public web resources is seemingly-endless and increasing, the main method of finding information is still mainly per-formed by text queries. Because of the typically enormous amount of results for text queries, popular search engines rank their results using a blend of relevance and popularity. The results often include items the user is not interested in, often leading to user query refinement, e.g. by adding additional or using different keywords, and in the worst case also to frustration.

Although the amount of data is increasing, it also becomes more structured. Metadata as well is structural relationships between results can portrait a contextual overview of the data allowing to drill down. Contextual overview can take many forms, topical, geo-spatial and temporal are some of the more popular. Thanks to web-sites such as Twitter and FlickR there exists rich data sources that are annotated with geo-spatial and temporal metadata. Users already familiar with searching in geographic environments like Google Maps, can find results presented directly on a map rather than in a list. There exist also early prototypes for mapping the events and characters of cultural heritage onto a map. These are driven by enthusiasts, using the Google API [GAP] or the INSPIRE framework [INS]. Furthermore, quite sophisti-cated approaches to the mapping of events to an interactive timeline are ready to be implemented. Examples for these are Timeline and Timeplot.

Dörk *et al.* [DCCW08] presented *VisGets:* a web application that shows the result of a query in four linked widgets, each showing a different context. A location *VisGet* shows the location of items in the result set overlaid on a map of earth. A time *VisGet* shows the distribution of items in bins representing years, months and days. A tag *VisGet* shows a tag cloud for the items. A results *VisGet* shows small textual or image thumbnails of items arranged as a table. The location, time and tag *VisGet* allow the user to refine the query spatially, temporally and topically, respectively, updating the other views. These three *VisGets* can be viewed as specialized widgets for dynamic queries [Shn94]. They therefore exploit the synergetic effects of the different representations very well.

We show a web application that enables the synergy of geo-spatial and temporal context exploration of one or more topical queries simultaneously. We aim to enable the user to inspect the results of multiple queries in a geographic and historical context. We present an advanced time widget, that shows a more detailed overview and allows more flexible selection of time ranges. Our advanced map widget prevents shown items from overlapping. Both widgets were extended to show multiple result sets allowing their direct comparison. Our ultimate goal is generating a web application that is used in conjunction with the search engine of Europeana and other search engines for digital libraries alike.

# 2 Features

## 2.1 Map

The central component of Europeana.4D is the map. The current prototype features several well-known concepts in terms of usability. Generally, it works like better-known map services on the web, such as Google Maps. It supports zooming as well as drag and drop for navigation. Additionally, it is possible to display historic maps for better contextualisation of the displayed items.

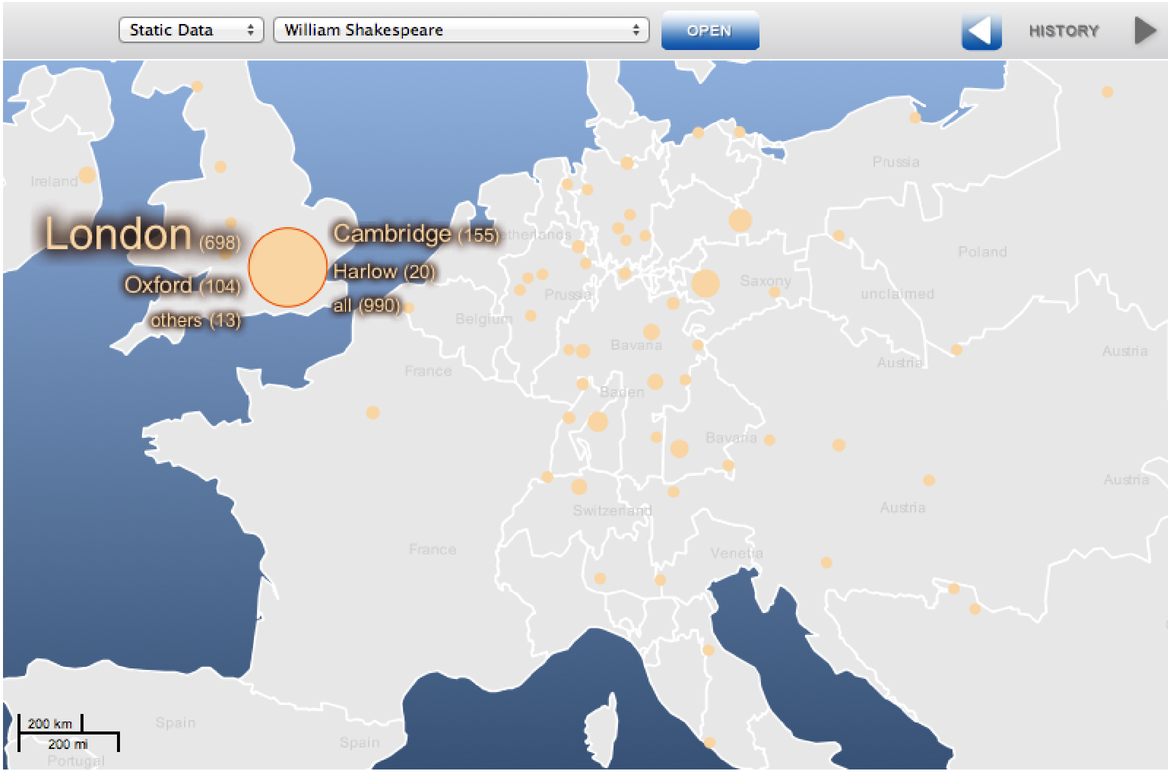



Fig. 1 Historical map of 1783 - showing publications of Shakespeare

The map widget consists of a geographical map, which is overlaid with bubble glyphs. The glyphs are circles, with each of the glyphs representing one or more items of the dataset. We chose a circle representation on the map to display data densities instead of, for instance, a heat map, in order to allow multiple query results to be shown together, and to make every data item individually accessible through its graphical representation.

In addition, to avoid visual clutter, we require that the glyphs do not overlap each other. To achieve this goal, we merge circles based on their size, distances, and the current zoom level in an iterative process.

## 2.2 Historical maps

For datasets with historical content we provide a great variety of historical maps, which show the political situation in specific eras. We offer 23 different historical maps from 2000 BC to 1994 AD provided by Thinkquest. Imagine a dataset that contains elements with time stamps that fall within in the Middle Ages alone. Unfortunately, one map cannot represent all political situations in the time span “Middle Ages,” and a “mean” between country borders would not only be hard to compute, it would also misrepresent the facts. For that reason, we choose the map for the median time stamp of all data objects of the dataset as default. For instance, this could be the historical map of 800. We also allow the user to switch maps.

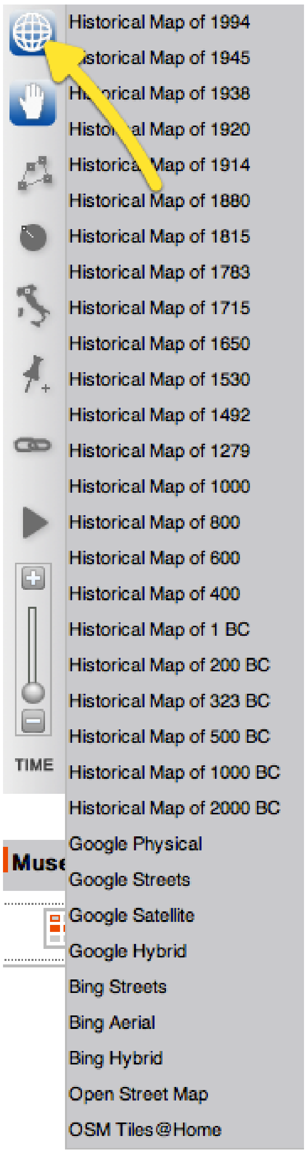
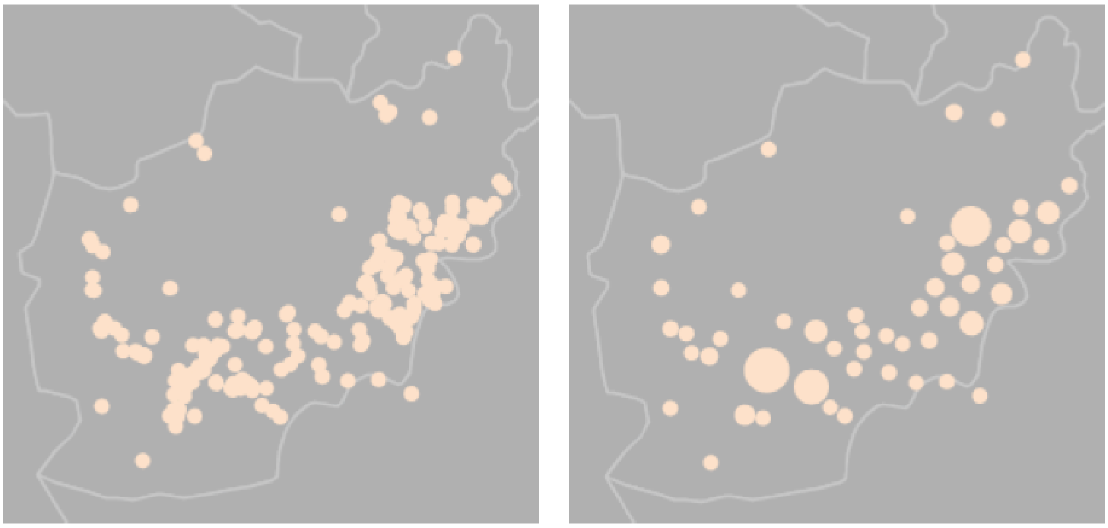



Fig 2 Menu to switch between maps

With this feature, the representation of the elements on the map that shows the political situation of that time can be more informative than visualising the results on a contemporary political map.

## 2.3 Dynamic aggregation of data

To avoid a confusingly large number of dots for large results or datasets, the interface automatically aggregates points depending on the zoom level. To ensure non-overlapping circles, we perform a variant of the agglomerative clustering algorithm.



(a) Individual circles (b) Aggregated circles

Fig. 3 Example of the interface automatically aggregating points depending on the zoom level

Each circle with minimum radius rmin represents one event. The high degree of overplotting makes it hard to access each circle individually (a), and to determine the event-centres. (b) 307 elements are merged to 52 non-overlapping circles. The composition of circles reflects that there were a lot of events taking place in the south.

## 2.4 Placename tag cloud

Each of the map's circles is associated with a tag cloud of the most frequent place names, including their quantity and whether they were provided by the given data. The size of the literals of a place name is proportional to its quantity in the corresponding circle.

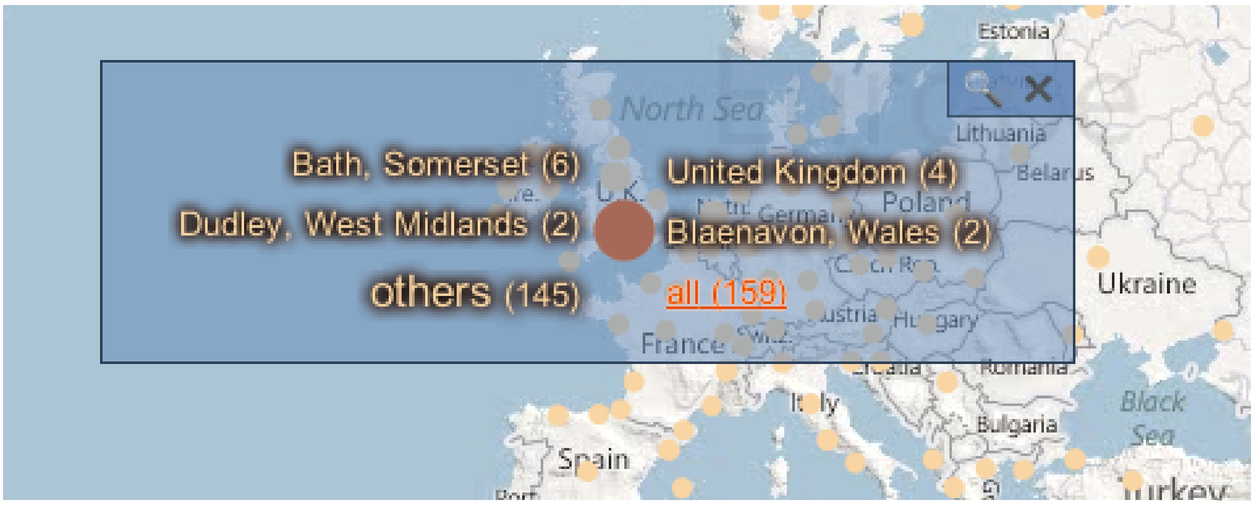



Fig 4 Place name tag cloud

This feature enables on-demand labelling of points and also provides a preview of how a glyph arising from agglomeration would split if zoomed in. If the data offers different levels of detail for a place, we choose the label dependent on the current zoom level. We distinguish four semantic levels: country, region (which can be, for instance, a state or a countryside), city, and borough (which can be a district, specific place or an address of the given city).

## 2.5 Time line

Since the timeline is a rarely used control in web applications, it wasn't possible to reuse existing GUI concepts. The timeline is primarily used for drill-downs in time. Another functionality is the display of individual datasets for a given point in time. This is achieved via a simple mouse over.

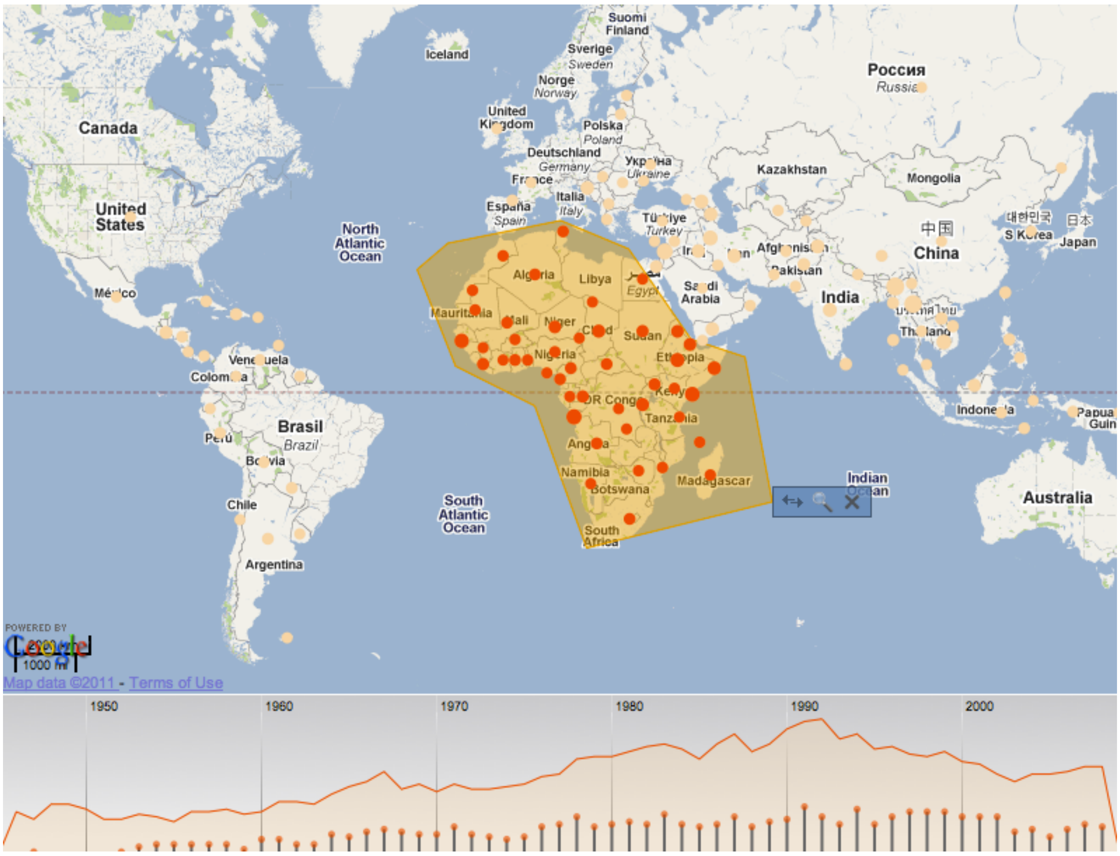



Fig 5 Display of individual datasets for a given point in time

## 2.6 Fixed time selection

The time widget allows both the clicking on one bin and the selection of a time range using a mouse drag gesture. A toolbar is then shown that offers the possibility of modifying the left or right border of the selected time era and adjusting a feather range beyond the selection borders to smooth the transitions between selected and non-selected elements.

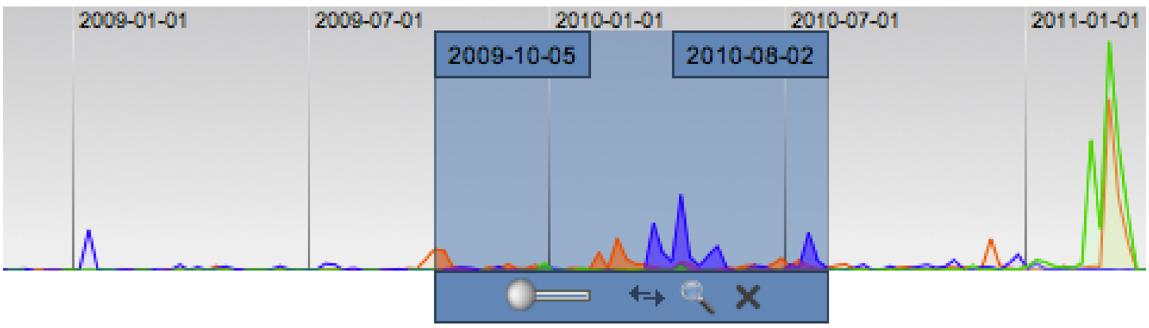



Fig. 6 Fixed time selection

This triggers a weighted colouring of the non-selected circles inside the map. Additionally, a user can display time-dependent connections between data items. For each bin within the selection a minimum spanning tree between the corresponding circles' centres is displayed on the map. This helps to detect geo-spatial dependencies in short time periods. Finally, the user can drag the selected time era manually or by animation. We then update the circles on-demand to reflect how locations change over time. A play button starts the animation and the selected time span moves smoothly over the entire time range. The advantage of this feature is that the user can direct his attention to the changes inside the map. Finally, a fixed selection can be turned into a refined query.

## 2.7 Table

The third method for displaying Europeana - result set is a table. The table also interacts with the timeline. If a duration is selected, the items will be highlighted.

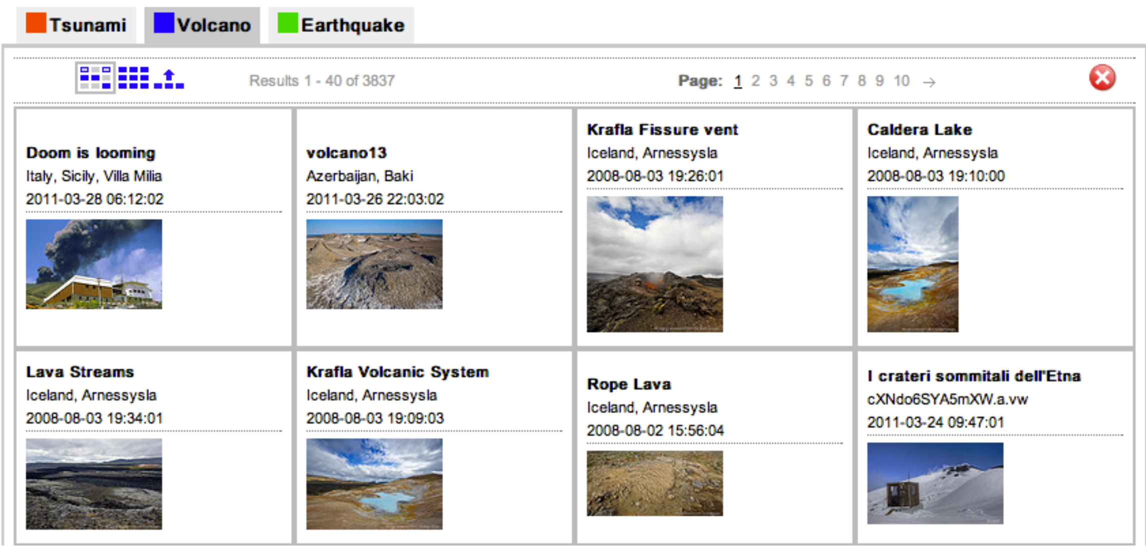



Fig 7 – Displaying Europeana results as table

The boxes of the table widget can also be changed by selecting a map or timeplot. A temporary selection causes highlighting of a box's border and a fixed selection fills the box's area. Additionally, the table can be used for the selection and the de-selection of single elements. Via this feature, the user is able to adapt a selection by adding or deleting specific elements. Finally, the user can switch the display mode of the table, so that only selected items are shown, which simplifies the browsing through the final selection.

## 2.8 Selection tools

Several selection tools on the map and on the timeline allow drill-down in time and space. The map features selectors for polygons, circles, and countries. These also work on the historic maps.

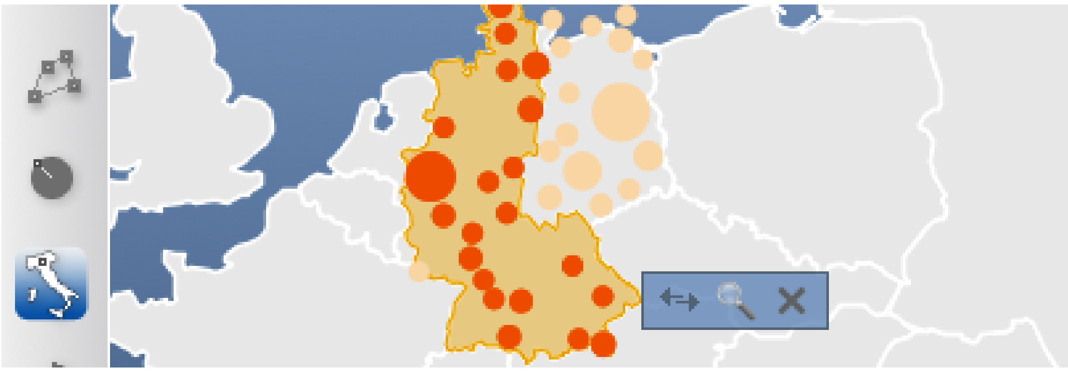


Fig 8 Selection tools

The timeline allows selections of points in time and durations. Durations can be fuzzy. It's also possible to drag a selection on the timeline, during which the dots on the map change in real time.

## 2.9 Concurrent datasets

An advanced feature of this prototype is concurrent searching. Using this mode allows combining up to four (limited for clearness) searches. The user has the full possibilities given for a single search. This includes selection editing and display of result sets.

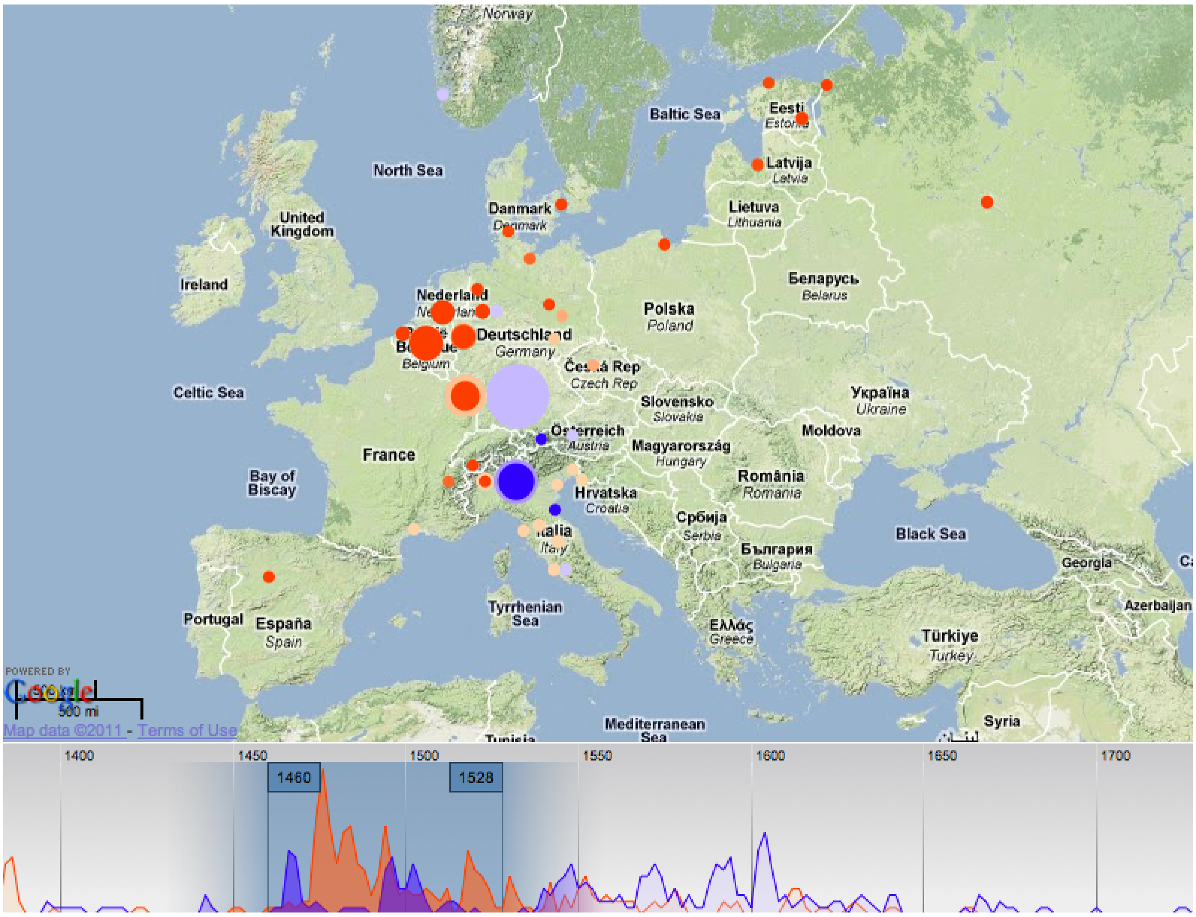



Fig 9 Display of concurrent datasets

## 2.10 Intra dataset comparison

One of the major features of the detail widget is the ability to export elements of a fixed selection from a dataset as a new individual input set.


Fig 10 dataset selection tools

Hence, the temporal comparison of different geographical regions from one dataset is feasible as well as the geographical comparison of different time periods.

## 2.11 Connections

Another advanced feature is the possibility to connect the dots by time. This mode allows one to follow persons or ideas of European cultural heritage in time. By dragging the selected duration the connections can be adjusted.

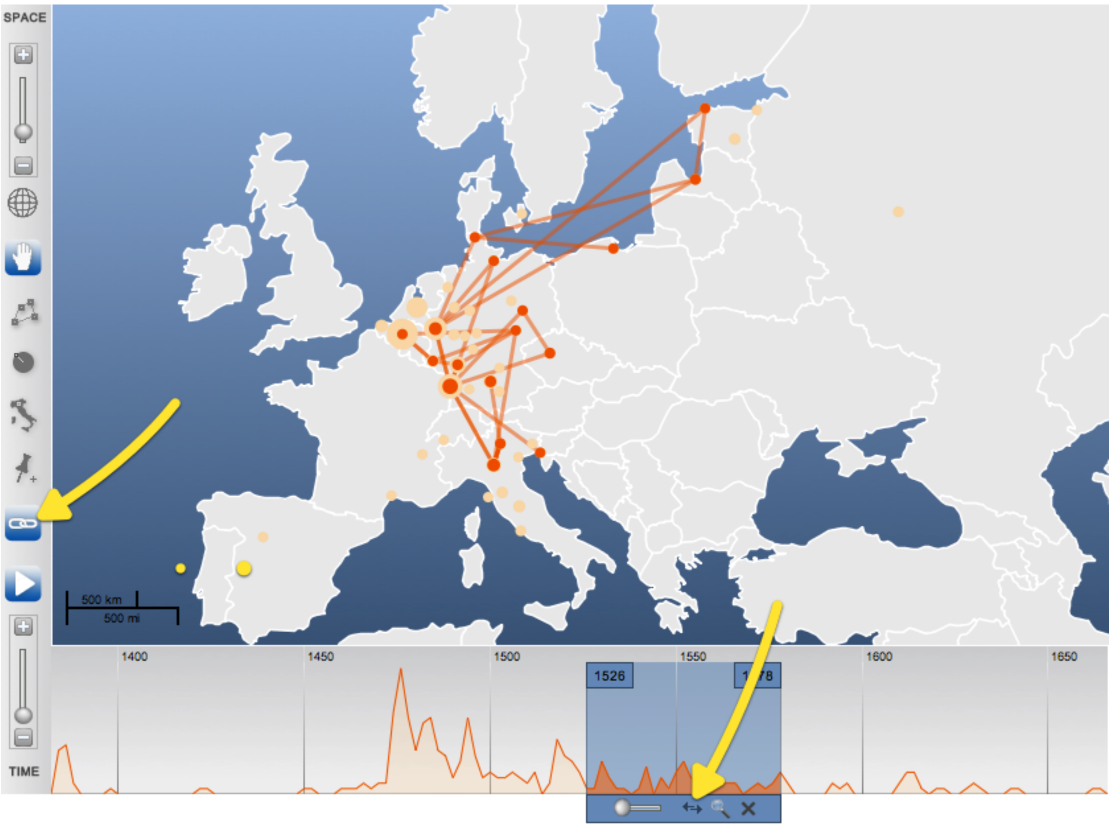



Fig 11 interactive connections tool and controls

# 3 Implementation

We implemented our method in a prototype application in the context of the European project EuropeanaConnect. Our tool is called *europeana4D.* It is a client-server architecture, where we charge the client with the main functionality of the system.

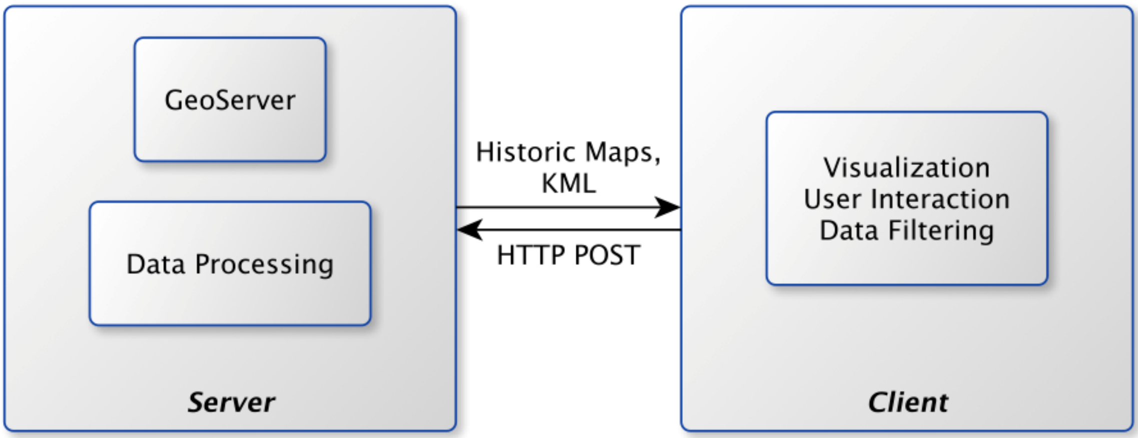



Fig 12 client-server architecture

The Web server is mainly used for the data processing step. Dependent on the requested datasource, the server retrieves the requested data and constructs a KML-file with a specific format. For each received data element, we add a *Placemark* tag to the KML root. Each Placemark has to be filled with a *name,* a *location* (latitude and longitude) and either a valid *timestamp* or *timespan.* Optionally, a place mark node can be enriched with a *location info* (address), that will be used for the placename-tagclouds. A *description* entry can be used to commit a CDDATA section, that includes HTML content. Example placemarks without description are shown in Figure 5. Finally, the KML-file will be sent to the client as a response on the XMLHttpRequest.

```
<Placemark>
    <name>India vs. Pakistan</name>
    <address>Kashmir</address>
    <TimeSpan>
        <begin>1996</begin>
        <end>2003</end>
    </TimeSpan>
    <Point>
        <coordinates>73.9104,33.94718</coordinates>
    </Point>
</Placemark>
<Placemark>
    <name>France vs. MDRM</name>
    <address>Malagasy</address>
    <TimeStamp>
        <when>1947</when>
    </TimeStamp>
    <Point>
        <coordinates>47.0,-20.0</coordinates>
    </Point>
</Placemark>
```

*Placemarks of the Armed Conflict data, one with a timespan, and one with a timestamp*

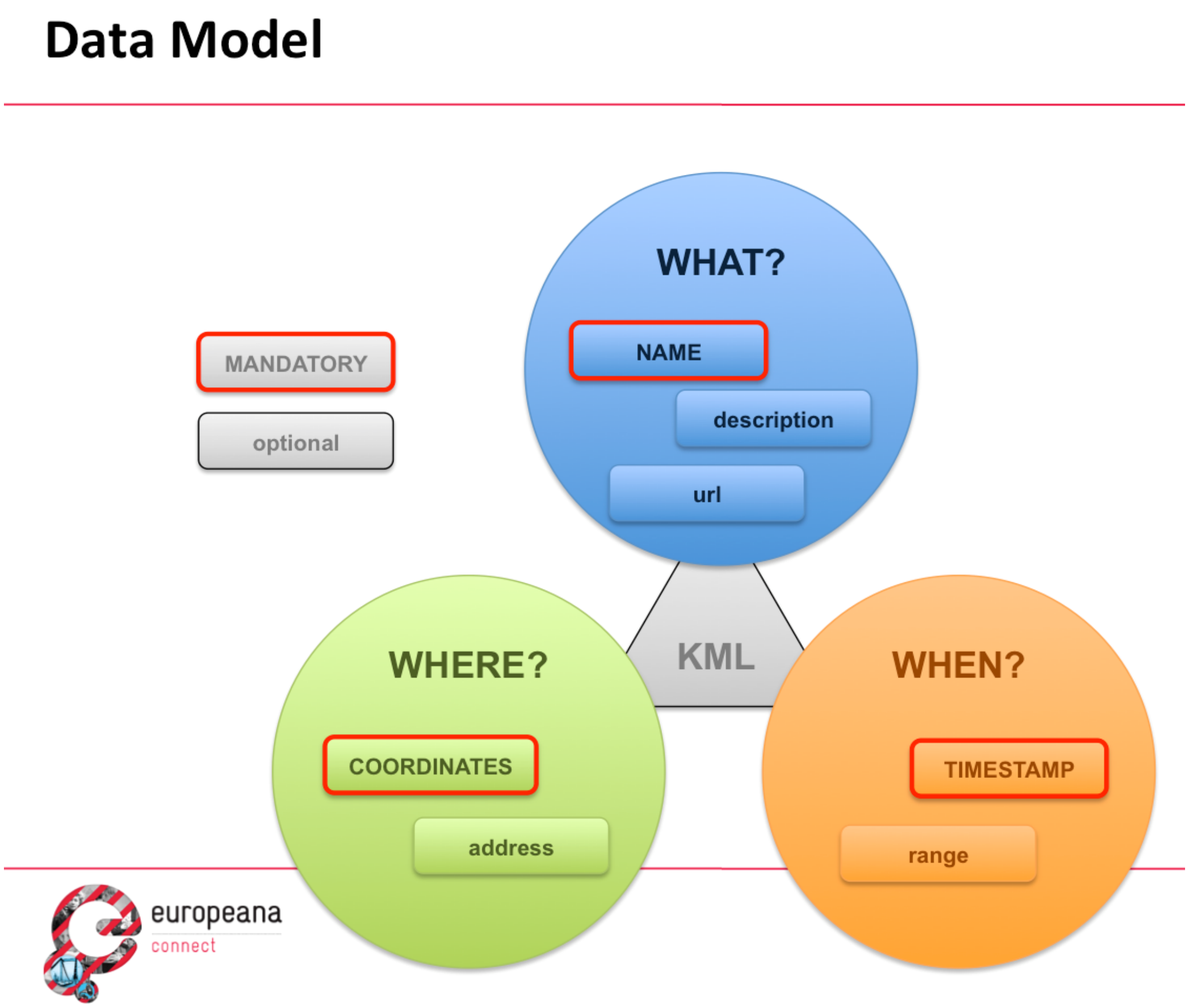



Fig 13 the KML format consists of three components: the event (what), the place (where) and time (when)

The client browser parses the retrieved KML-file, creates JavaScript objects and fills the widgets with content. We decided to construct a pure JavaScript client, because it is supported by every browser without any additional package. All interactions, including data filtering are performed on the client-side. The great advantage of this system structure is, that each modification of a dataset within the client browser just triggers some functions on the client side. This benefits the response time of modifications, since the browser doesn't have to wait on the server.

For the widgets on the client side we make use of two OpenSource JavaScript libraries. A modificated *Simile Widgets Timeplot* instance organizes the segmented area graphs and *OpenLayers* visualizes the circles and arranges the different map layers. We offer some popular maps as well as our historic maps with the political borders dependent on a specific time era. For this, we use a Web-server instance of the *GeoServer* [Geob], that provides the tiles of the historic maps.

Due to the requirements for the integration of Europeana4D (light) into the Europeana portal, the reverse migration from GWT to JavaScript covers the full functionality and interaction of the map and timeplot widget. The header (static datasets,...) and table widget are dropped, since these parts are undesired. As an additional feature, a popup representation for the objects of a circle of the map or a selected time range was included as an alternative to the place name tag cloud. Furthermore, the structure of Europeana4D was revised intensively, and a configuration file was added to manipulate the style of Europeana4D easily. Actually, Europeana4D light is under construction for IE8/9.

# 4 Dissemination and Outreach

## 4.1 Source Code

The source code ist published as open source in the Europeana code repository:

https://www.europeanalabs.eu/svn/contrib/sti/

## 4.2 User Tests

Please find the results of the two user-tests on the EuropeanaConnect website in M3.3.7 – Europeana4D: Testing and Evaluation of the first prototype Report 1 and M3.3.9 - Europeana4D: Testing and Evaluation of the second prototype Report 2

## 4.3 Europeana ThoughtLab

Europeana4D is published in the Europeana ThoughtLab portal :

http://europeana.eu/portal/thoughtlab_semanticsearching.html

## 4.4 Project Website and Demonstrator

A comprehensive project website has been published on:

http://tinyurl.com/e4d-project

Two youtube screencast-clips welcome the visitor to offer a quick glance of the prototypes features. The demonstrator of the expert version is available at

http://tinyurl.com/e4d-demo

offering both a realtime search of the europena repository and some other datasets (Wikipedia et. al.)

# 5 Conclusion

Europeana 4D presents a novel approach and web application to show, compare and explore multiple topical queries in a geograph- ical and temporal context. In comparison with the similar work of Dörk *et al.* [DCCW08] the prototype also uses a linked geo, time, and detail widget but extended them for multiple result sets. We also support brushing and linking as well as query refinement.

Visually comparable datasets in temporal and geo-spatial dimension extends the exploration and analysis abilities of the user in an effective way. It helps to detect equalities and varieties between distinct data contents that unveil their relationships in time and space. Our method is limited to four datasets at the same time mainly to ensure that the colors used for discrimination are properly distinguishable, the splitting of circles does not waste too much screen space, and the overlapping segmented area graphs do not occlude each other too much.

In the user study, It became clear, that the general approach of the interface will be an interesting approach for a broad range of research scenarios and the broad pubic alike. Academic participants were eager to us the tool on their own datasets to explore new correlations. As a result, we can define two major target audiences: a) casual users who are interested in new way to explore interesting data (Europeana data) and b) scientists who already own or are now producing research data with components of both space and time relations (strong focus on the humanities and social sciences).

To offer interesting datasets proved key for the overall acceptance of *europeana4D.* Some suggestions came from the participants: famous travellers (Cook, Humboldt, Polo etc.), Nobel Prize winners, the dispersal of international brands like IKEA or McDonald’s, connecting to real-time datasets like twitter hash tags.

In the future we will direct our attention to the extension of our method to very large datasets by searching for client-server communication where most of the data reains on the server but a working set is transmitted to the client and used there for quick user interaction.

# 6 Description of software developed for Europeana within EuropeanaConnect

This should be one paragraph giving some background information and a short explanation of the software

| | |
|---|---|
| Link to software | Browser: http://europeanalabs.eu/browser/contrib/sti<br>SVN-Repository: https://europeanalabs.eu/svn/contrib/sti<br>GWT (SVN): https://europeanalabs.eu/svn/contrib/sti/branches/sti-gwt |
| Login information | none |
| Development environment | Eclipse with GWT Plugin, Ant |
| Programming language used | Java 1.6.0.26 (GWT 2.2.0), JavaScript 1.8 |
| Application server used | Apache Tomcat/7.0.11 |
| Database requirements | none |
| Operating system requirements | webapp: platform independent<br>development: contemporary OS with Java>1.6 |
| Port requirements / default ports used | Default 8080 |
| Interface | (static) web application |
| Licensing conditions | GPL |

# Appendix A. References

[AA99] ANDRIENKO G., ANDRIENKO N.: Interactive maps for visual data exploration. *International Journal of Geographical Information Science 13,* 4 (1999), 355–374. 2

[AA07] ANDRIENKO G., ANDRIENKO N.: Visual data exploration: Tools, principles, and problems. *Classics from IJGIS: twenty years of the International journal of geographical information science and systems 475* (2007). 2

[AAS 01] ANDRIENKO N., ANDRIENKO G., SAVINOV A., VOSS H., WETTSCHERECK D.: Exploratory analysis of spa- tial data using interactive maps and data mining. *Cartogr. Geogr. Inform. Sci. 28,* 3 (2001), 151–165. 2

[BG03] BRODBECK D., GIRARDIN L.: Design study: Using multiple coordinated views to analyze geo-referenced high dimensional datasets. In *Proceedings of the conference on Coordinated and Multiple Views In Exploratory Visualization* (2003), IEEE Computer Society, pp. 104–111. 2

[DCCW08] DÖRK M., CARPENDALE S., COLLINS C., WILLIAMSON C.: Visgets: Coordinated visualizations for web-based information exploration and discovery. *IEEE Transactions on Visualization and Computer Graphics 14,* 6 (2008), 1205–1212. 1, 2, 8

[GAP] Settings in Shakespeare's plays using google maps. http: //tinyurl.com/2v3usy (Retrieved 2010-12-02). 1

[Geoa] Geonames. http://www.geonames.org/ (Retrieved 2010-12-02). 6

[Geob] Geoserver. http://geoserver.org/display/ GEOS/Welcome (Retrieved 2010-12-02). 6

[GUA] Wikileaks iraq: data journalism maps every death.

http://www.guardian.co.uk/news/datablog/ 2010/oct/23/wikileaks- iraq- data- journalism (Retrieved 2010-12-02). 6

[GWE∗01] GLEDITSCH N., WALLENSTEEN P., ERIKSSON M., SOLLENBERG M., STRAND H.: Armed conflict 1946-2001: A new dataset. *Journal of Peace Research 39,* 5 (2001), 615–637. 6

[Har99] HARRIS R. L.: *Information Graphics: A Comprehensive Illustrated Reference.* Oxford University Press, 1999. 3, 4

[INS] European comission inspire geoportal. http://www.inspire-geoportal.eu (Retrieved 2010-12-02). 1

[Jer09] JERN M.: Collaborative web-enabled geoanalytics applied to oecd regional data. In *Cooperative Design, Visualiza- tion, and Engineering,* Luo Y., (Ed.), vol. 5738 of *Lecture Notes in Computer Science.* Springer, 2009, pp. 32–43. 2

[Kei02] KEIM D. A.: Information visualization and visual data mining. *IEEE Transactions on Visualization and Computer Graphics 8,* 1 (2002), 1–8. 2

[Kra67] KRAVITZ S.: Packing cylinders into cylindrical containers. *Mathematics Magazine 40,* 2 (1967), 65–71. 3

[Mar06] MARCHIONINI G.: Exploratory search: from finding to understanding. *Commun. ACM 49,* 4 (April 2006), 41–46. 2

[NA99] NATALIA G. A., ANDRIENKO N. V.: Making a gis intelligent: Commongis project view. In *AISB Quarterly* (1999), University Press, pp. 15–17. 2

[OL] Openlayers. http://openlayers.org/ (Retrieved 2010-12-02). 6

[PD07] PASCUAL V., DURSTELER J. C.: Wet: a prototype of an exploratory search system for web mining to assess usability. In *Proceedings of Information Visualisation, International Conference on Information Visualisation* (2007), pp. 211–215. 2

[SBdOF03] SHIMABUKURO M. H., BRANCO V. M. A., DE OLIVIERA M. C. F., FLORES E. F.: Visual exploration of spatio-temporal databases. In *Proceedings of the Geoinfo 2003* (2003). 2

[Shn94] SHNEIDERMAN B.: Dynamic queries for visual information seeking. *IEEE Software 11,* 6 (1994), 70–77. 1

[Shn96] SHNEIDERMAN B.: The eyes have it: A task by data type taxonomy for information visualizations. In *Proceedings of the 1996 IEEE Symposium on Visual Languages* (1996), pp. 336–. 2

[SIM] Simile project. http://simile.mit.edu (Retrieved 2010-12-02). 1

[TP] Simile widgets timeplot. http://www.simile-widgets.org/timeplot/ (Retrieved 2010- 12-02). 5

[Tuk77] TUKEY J. W.: *Exploratory Data Analysis.* Addison- Wesley, 1977. 2

[UCD09] Ucdp/prio armed conflict dataset codebook (version 4- 2009), 2009. 6

[War04] WARE C.: *Information Visualization: Perception for Design.* Morgan Kaufmann, 2004. 3

[WBWK00] WANG BALDONADO M. Q., WOODRUFF A., KUCHINSKY A.: Guidelines for using multiple views in information visualization. In *Proceedings of the working conference on Advanced visual interfaces* (2000), AVI ’00, ACM, pp. 110–119. 2