Exploring Pelagios: A Visual Browser for Geo-Tagged Datasets

Rainer Simon[1], Elton Barker[2], and Leif Isaksen[3]

[1] AIT - Austrian Institute of Technology, Donau-City-Str. 1,
1220 Vienna, Austria
rainer.simon@ait.ac.at
[2] The Open University,
Milton Keynes, United Kingdom
e.t.e.barker@open.ac.uk
[3] University of Southampton,
Southampton, United Kingdom
l.isaksen@soton.ac.uk

Abstract. When dealing with networked information spaces, there are two challenges a data navigation system must address: first, it needs to provide the user with a quick high-level overview of the nature and structure of the data. Second, it must offer mechanisms to easily *drill down* and navigate to specific detail information. In this demonstration, we present the *Pelagios Graph Explorer*, a prototype visual exploration tool for place metadata which is derived from ancient world research datasets brought together by the Pelagios Project. Furthermore, we report on ongoing work on the Pelagios API. The API enables external re-use of the Pelagios place metadata, and provides a basis for future development of - and experimentation with - alternative navigation tools for the networked place metadata from the Pelagios project.

Keywords: Linked Data, Visual Browsing, Information Exploration, Digital Libaries, Digital Humanities, Geoinformation

1 Introduction

On-line resources that reference ancient places are multiplying rapidly, bringing huge potential for the researcher - provided that they can be found. But users currently have no way of easily navigating between such resources, or of comparing their contents. The Pelagios[4] consortium is a growing international collective of ancient world projects that aims to address the problems of discovery and reuse with the twin aims of helping digital humanists to make their data more discoverable, and of empowering real-world users (scholars and the general public) to find information about particular ancient places and visualize it in meaningful ways. While the project focuses on the ancient world, the methodology and tools developed will be of interest to anyone working with data

---

[4] http://pelagios-project.blogspot.co.uk

containing spatial references. The Pelagios collaboration intentionally includes partners maintaining a wide range of different document types, including texts, maps and databases[5]. In doing so we take some of the first steps required to build a Geospatial Semantic Web for the Humanities [4].

2   Connecting Data Through Place

Pelagios aims to create links between data entities on the Web based on the places they refer to. It does not impose any restrictions on the type and format of data entities themselves; nor on the metadata schemas or media types used. Instead, it provides a lightweight framework of principles and requirements with which to express place references as additional metadata.

The first, and essential, requirement is that when partners refer to a place, they do so according to a shared reference system: in our case, this reference system is the *Pleiades Gazetteer of the Ancient World*[6], which provides Uniform Resource Identifiers (URIs) for more than 34.000 places in the ancient world. By expressing each place reference in the form of a Pleiades URI, the otherwise disjointed datasets are joined up to an interconnected graph, with Pleiades as the central backbone.

The second requirement is a syntactical requirement: partners need to publish their resulting place metadata online according to a common data representation format, and common vocabulary terms. The representation format used in Pelagios is RDF[7]; the vocabulary used to model place references is the *Open Annotation* vocabulary [5]. The idea behind considering a place reference an "annotation" is that the reference should not be considered certain, but rather that a statement is being made about some kind of relation between the data and the place. Additional descriptive metadata about the dataset as a whole should be published using the *Vocabulary of Interlinked Datasets (VoID)*[8]. This information will typically include things such as textual descriptions of the data, or publisher and license information.

It is important to note that the additional place metadata required by Pelagios can be completely separate from the origin data itself. This approach is sometimes referred to as "standoff markup" [8], and helps to avoid data management problems which would arise if participants were required to adapt or extend their internal metadata schema for Pelagios.

---

[5] At time of writing, a growing list of Pelagios partners included: Arachne, The British Museum, CLAROS, Fasti Online, Google Ancient Places, Inscriptions of Israel/Palestine, Meketre, Nomisma, OCRE, Open Context, Oracc, Papyri.info, Perseus, Ports Antiques, Ptolemy Machine, Regnum Francorum Online, SPQR, The Ure museum. For an up-to-date list, see: `http://pelagios-project.blogspot.co.uk/p/about-pelagios.html`

[6] `http://pleiades.stoa.org`

[7] `http://www.w3.org/RDF/`

[8] `http://www.w3.org/TR/void/`

3   Exploring the Pelagios Network

From a conceptual point of view, the Pelagios network consists of so-called *datasets*. Datasets are collections of *place references*: statements that express the fact that a particular data object (e.g. a record in an archaeological database, a segment of text in an online book archive, an image in a collection of digitzed photographs, etc.) bears some kind of relation to a particular place. Datasets may also be structured into *subsets*, and carry additional metadata such as a title, a description, or license information. In implementation terms, the datasets are defined through the *VoID* metadata.

It is important to stress that Pelagios datasets do not hold the data itself (or copies of it). They only consist of the place references plus a minimum of metadata about the data related to the place. Implementation-wise, this information is expressed via the annotations.

3.1   Pelagios Graph Explorer

The *Pelagios Graph Explorer* [7] is a research prototype that allows users to explore the Pelagios network of place references using visual metaphors. Simple specific queries the Graph Explorer can answer are, for instance:

– Which datasets reference a particular place?
– Which places are referenced in a particular dataset?
– What is the geographical footprint of a particular dataset?

Results are presented in combinations of map- and graph-based views. Views are synchronized and reveal more information as the user interacts with them. For example, hovering the mouse over a dataset in the graph view will show additional details about it, and overlay its geographical footprint on the map.

The key goal of the Graph Explorer, however, is not only to provide answers to questions that relate to single specific places or datasets. Instead, the Graph Explorer aims at, first, providing users with a sense of the overall "connectivity" within the Pelagios network of data and places and, then, enabling them to drill down into specific details. The user interface separates this into two different modes of exploration: one mode is dedicated to exploring *connections between places, as established through data*. The other mode is dedicated to the "inverse" task of exploring *connections between data, as established through places*.

Fig. 1 shows a screenshot of the first exploration mode. In this mode, the user enters names of places in the search field. The Graph Explorer will produce an interactive network visualization that shows how these places are connected to each other through different datasets. In the graph view, datasets are represented as bubbles. Bubbles are colour-coded according to their root dataset: i.e. datasets that are subsets (or sub-subsets, etc.) of the same parent set share the same colour. Bubble size represents the total amount of place references contained in the dataset: the bigger the bubble, the larger the dataset. Likewise, the size of the connecting lines between places and datasets is scaled according to the number
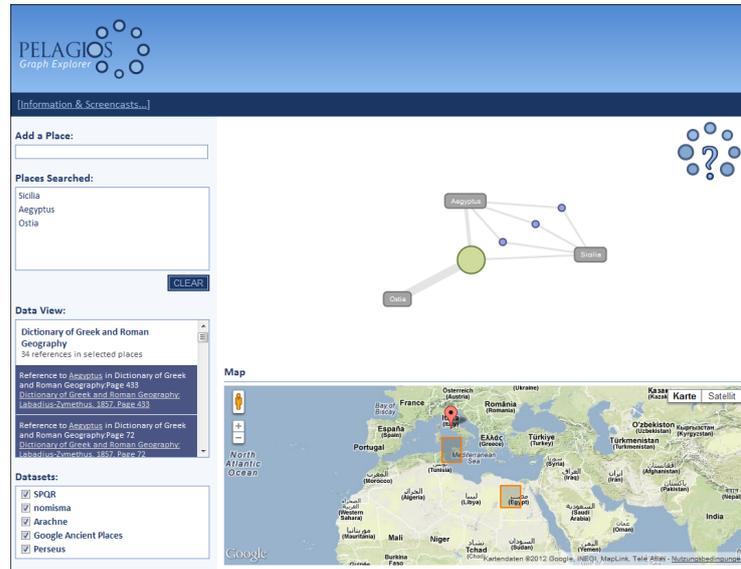
Fig. 1. Exploring Places Related Through Data.

of references to that particular place, in that specific dataset. As a consequence, a connecting line that has a width close (or equal) to the size of the dataset bubble will indicate a case where most (or all) of the references in the dataset are to this one place, and only few (or no) other places are referenced in this dataset. This way, users can quickly identify datasets that are primarily about this place, versus datasets where the place is referenced in isolated cases only. If, in addition to a relatively wide line, the bubble is small (representing a dataset with a low number of place references total), users will furthermore know that they are looking at a very specific, small set of data – such as a single page in a book – primarily about this place; and that this dataset could be worth looking at in more detail. A click on any dataset will bring up a list of the place references and accompanying metadata (including links to the actual origin data) in the *Data View* side panel on the left.

Fig. 2 shows a screenshot of the second exploration mode. In this mode, users can explore all datasets in the graph, using a similar representation. Datasets are again visualized as colour- and size-coded bubbles. Double-clicking on a dataset reveals its subsets, allowing the user to navigate through the dataset hierarchy. Similar to the first exploration mode, hovering over a dataset shows additional information about it, as well as its geographical footprint on the map. The user can select a dataset with a single click to get detail information about the places referenced in it. To provide the user with a sense of how different datasets correspond to (or overlap with) each other in terms of the places they reference, it is possible to select multiple datasets in the graph view. The view will generate connecting lines between the selected sets which are scaled according

to the number of places that co-occur in both datasets. This way, users can quickly grasp the amount of place-wise "correlation" that exists between different datasets, based on the thickness of the lines between them. Hovering over the connecting line brings up further details, namely the number of places occuring in both datasets, and their geographical footprint. A click on the connecting line brings up the actual place references and source data links in the *Data View*.
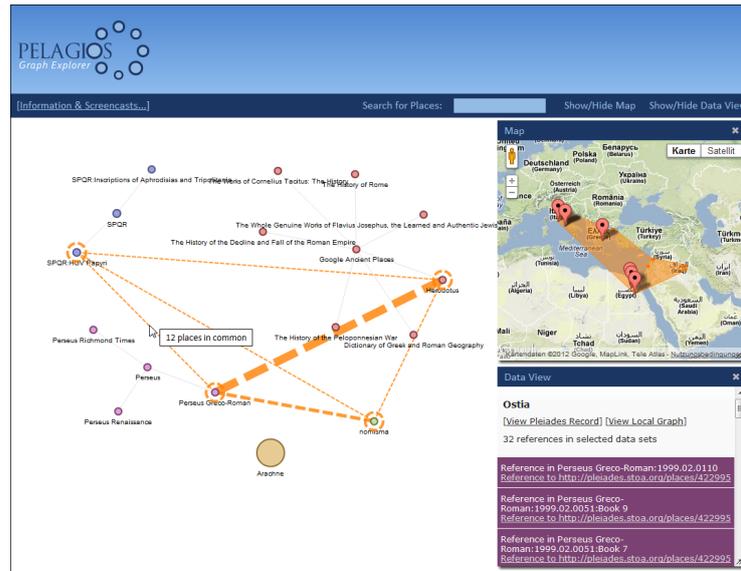

Fig. 2. Exploring Data Related Through Places.

3.2 Pelagios API

Pelagios also provides an HTTP *Application Programming Interface (API)*. The API enables 3rd party applications to explore the raw data behind the visualizations of the Graph Explorer, and to build alternative user interfaces and mashups using common Web technologies such as AJAX and JSON. Functionality offered by the API includes: step-by-step traversal of the Pelagios dataset hierarchy; search for places by name or geographical area; querying for places referenced in a particular dataset, or datasets referencing a specific place; retrieval of place references with configurable pagination; etc.

To foster re-use, the API also exposes Pelagios data as *Linked Data* [2]. It provides URIs for datasets and places, and supports different RDF serialization formats. Using *Content Negotiation* [1], the API also serves HTML representations for human users.

The Pelagios API is currently undergoing a major upgrade to offer increased performance and enhanced scalability. A publicly accessible beta version is available at `http://pelagios.dme.ait.ac.at/api`.

4   Future Work

In its present state, the Pelagios Graph Explorer is a research prototype. It has undergone a first round of user evaluation [6]. Further work has been carried out with regard to tailoring user interfaces built on top of the Graph Explorer or the API towards the needs of specific audiences. An important criterion in this regard has been the distinction between *super users* – Website owners who build, own or maintain place-related web pages, web sites or blogs in the ancient history domain; and *end users* – e.g. ancient history researchers, teachers and students with expertise in the ancient history domain [3]. Ongoing and future work on the user-facing aspects of Pelagios will focus on refinements in this regard.

Furthermore, Pelagios partners and third parties (such as the *Open Context*[9] portal for primary field research data from archaeology and related disciplines, or the *Ancient World Linked Data* JavaScript library project *awld.js*[10]) are starting to make use of the Pelagios API in their own applications. This real-world exposure will be crucial to gather valuable feedback on the use and utility of our toolset, and to guide its future development.

References

1. Bizer, C., Cyganiak, R., Heath, T.: How to publish Linked Data on the Web (2007), `http://www4.wiwiss.fu-berlin.de/bizer/pub/LinkedDataTutorial/`
2. Bizer, C., Heath, T., Berners-Lee, T.: Linked Data - The Story So Far. International Journal on Semantic Web and Information Systems 5(3), 1–22 (2009)
3. Culver, J.: Pelagios project blog: Developing widgets for pelagios (March 2012), `http://pelagios-project.blogspot.co.at/2012/03/developing-widgets-for-pelagios.html`
4. Harris, T.M., Rouse, L.J., Bergeron, S.: The Geospatial Semantic Web, Pareto GIS, and the Humanities. In: Bodenhamer, D.J., Corrigan, J., Harris, T.M. (eds.) The Spatial Humanities: GIS and the Future of Humanities Scholarship, pp. 124–142. Indiana University Press, Bloomington, IN (2012)
5. Haslhofer, B., Sanderson, R., Simon, R., Sompel, H.: Open annotations on multimedia web resources. Multimedia Tools and Applications pp. 1–21 (2012)
6. Ridge, M.: Pelagios project blog: Pelagios usability testing results (September 2011), `http://pelagios-project.blogspot.co.at/2011/09/draft-results.html`
7. Simon, R.: Pelagios project blog: The pelagios graph explorer: A first look (August 2011), `http://pelagios-project.blogspot.co.at/2011/08/pelagios-graph-explorer-first-look.html`
8. Thompson, H.S., McKelvie, D.: Hyperlink semantics for standoff markup of read-only documents. In: Proceedings of SGML Europe 97: The next decade Pushing the Envelope. p. 227229 (1997)

---

[9] `http://opencontext.org`
[10] `http://isawnyu.github.com/awld-js/`