Geobiblio – Base de dados bibliográfica georreferenciada

Tiago José Perdigão Fernandes

Abstract

Managing bibliographies is recurrent research activity, these bibliographies can be used privately to support each researcher work or public to disseminate knowledge. With respect to georeferenced bibliography databases, currently existing systems in use have normally some limitation (like being restricted to one operating system or covering only a small region for geo-referencing) leading to a low use of such tools.

Geobiblio is a geo-referenced bibliographic system built using open-source code and technology. Besides using a database management system to store several types of bibliographic references with their associated data, its main feature is the ability to record related locations to some reference. It is even possible to search for references near some location, all with a user-friendly integrated map. We built Geobiblio to fill a lack of a geo-referenced system that is free to use an accessible to everyone and everywhere through the internet.

1 Introduction

Currently, web based bibliographic systems are used by research individuals, groups, or laboratories to disseminate research and publications. Besides the dissemination of it own production, web based bibliographies can be used to gather knowledge about a certain study area, resenting related documents structured and organized in a pre-determined way. The way these bibliographies are maintaining and managed varies greatly between institutions: can be a simple HTML references listings or backed by a database management system and web services. Personal bibliographies help the gathering and organization of references and material relevant to individuals doing research on a certain field.

Currently there are a large number of open source bibliography database management systems that serve these purposes to most research communities [1], providing stand-alone (on multiple operating system architectures) or online (using different programming languages). The available Web based systems serve the two main objectives of bibliography database management systems: i) allow the production of a listing of documents accessible by other users, search engines (Google) and aggregators (Google Scholar), and ii) can act as an internal productivity tool since these database management systems provide a standard mechanism to export bibliographic information (BibTex [2] or RIS) to reuse on documents.

In most research areas the bibliographic data schemas implemented by such systems is enough, but if geo-referencing is required, none of the generic widely available system is useful. In several research fields (Earth, life, and some social sciences) the publications are relevant to a certain study area, and some of the searches for relevant related work is targeted at specific geographic location or area [3]. Existing generic bibliographic systems do not allow such geographic searches. To solve this limitation technologies from various fields have been producing specific tools that allow the storage of geo-referenced bibliographic data [4] [5] [6]. These tools

allow the association of geographic coordinates to the documents' records, along with geographic searches, but are specific to a certain research field, not allowing the ease of reuse by other researchers.

Researchers requiring the storage of geo-referenced bibliographic information are required to build from scratch systems that suit their needs, leading to the developments of several similar systems, custom tailored to the specific needs of the particular authors, and difficult to be used by others. The geo-referencing of bibliography entries also allows the easy export of information using standards such as wfs of geojson, and integration with external services or GIS applications.

Geobiblio address the previous problems (isolation and low generality) providing a generic open geo-referenced bibliographic database management system. Geobiblio extends an existing open-source web-based bibliographic system [7], adding the necessary data models and the user interfaces, so that users can geo-reference bibliographic entries, present maps, and do searches. Geobilitio also export document references as features using the WFS [8] protocol. Geobilio is currently deployed and supports a tsunami related web-bibliography.

Next section presents related work, showing current limitation, next the Geobiblio system is presented. The end of this document presents the tsunamis in Portugal web-bibliography.

2 Related work

The organization and presentation of geo-referenced documents has ben a problem that researches have been trying to solve with different success levels. Doing a fast survey on earth-science laboratories web-pages it is easy to conclude that most of the available bibliographies are limited to a text list of references: some provide links to the pdf documents, but none provide geographical referencing, map presentation, or geographical search.

In order to disseminate specific bibliographies, researchers develop custom-made web portals [4] [6] or resort to online cloud-bases platforms [5] [9] [10]. The development of custom made bibliographic portals on the short run may provide a solution to the dissemination, but incurs high development costs. Furthermore the maintenance and interoperability with other systems becomes a burden necessary to tackle.

The use of infrastructures such as Geocommons [9] eases availability of geo-referenced data or features, but adds an overhead on the maintenance of the data. There is no well define data models for the bibliographic information, the management of the original database management system usually resorts to locally accessible source (spreadsheet or text files) and the update of the information requires a complex workflow (change the original source, definition of the kml file to upload, and upload of data). Furthermore the community effort on the production of a bibliography is hard to manage.

Journalmap [11] offers a centralized cloud infrastructure aggregating references imported from journal publishers and submitted by individuals. The main drawback of this solution is the closed implementation.

Furthermore it does not solve the problem with consistency between the local bibliographic database management systems and the web available data. At the moment it is not possible to integrate the managed data with other external services.

Currently open data initiatives (at the EU or US government levels) have been initiated, but in the area of geo-referenced bibliographies nothing has been done. Work has been done on the digital libraries, for instance europeana (http://www.europeana.eu), but the georeferencig of document records is limited and its integration with Geographical Information System (GIS) is impossible.

3 Geobiblio

Geobiblio solves the previously presented problems providing a bibliographic database management system infrastructure that in a single installation can be used in different ways:

- Personal geo-reference bibliographic database management system
- Public visible collection of bibliographies
- Source of data for other geo data services (for instance Open Street Maps[1]) - or GIS database management systems

Geobiblio is a three tier web system that can be seen on the figure below. It is composed by a web server built using the open-source bibliographic system Aigaion2 which was build using PHP and it was extended to support geo-referencing. A database management system is needed to support the persistence of the data and its fast search for information, it was used the well-known database management system MySQL because it was the database management system of choice by the bibliographic system Aigaion2 and it was also integrated on it. However to support geo-referencing it was necessary to extend the data model, adding new tables only and leaving the original data model unchanged. On the left tier we got the services provided by the web server, which are HTML and JavaScript web pages that can be seen on every web browser and a WFS service which returns WFS data in XML notation and can be used by a third-party program to process that information.

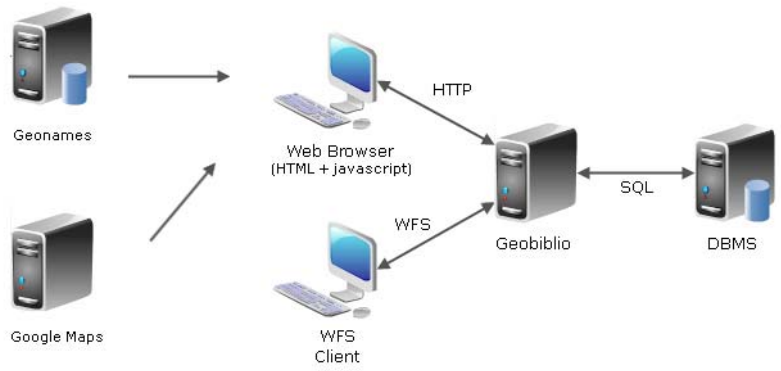

Figure 1 - Geobibio architecture

[1] http://wiki.openstreetmap.org/wiki/Potential_Datasources

3.1 Georeferenced data model

In order to build our geo-referenced system, new tables had to be created in the database to store the georeferenced data. The most important table that was created is the **locations** table where it holds the information about each stored location like its name and coordinates. Another table called **pub_has_locs** is used to link some location to a bibliographic reference and to store other information about that relation like the location rate and some notes. The third and the last created table is called **online_topics** and it is used just to store information about if some topic is meant to be published online or not, this table could be easily eliminated and the information stored in the topics table, however in order to keep the initial data model unchanged. This new data-model is represented on the diagram below.

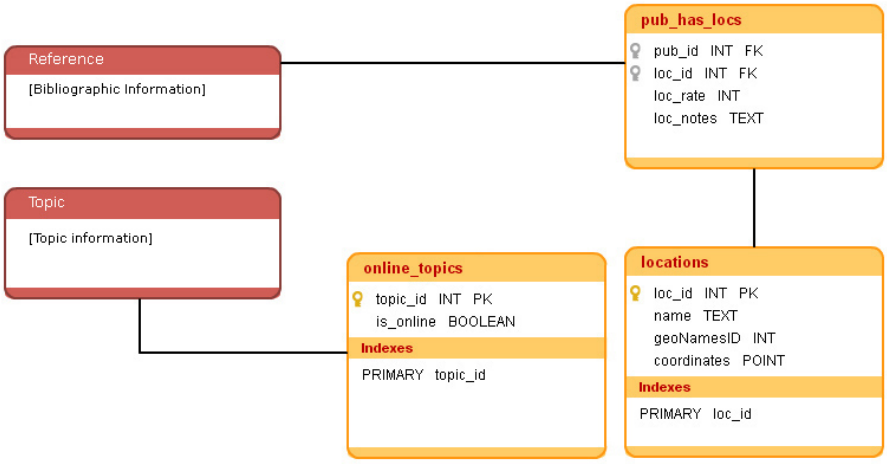

Figure 2 - Geo-referenced data-model

3.2 Main functionalities

Here we describe the complete Geobiblio features.

3.2.1 Geo-referencing documents

The main feature of this system is the possibility to geo-reference the bibliographic references. It has an interface built just to add new references with an integrated map to see the chosen place or to select one just by clicking twice on the map. Besides this method, it is possible to choose a place by entering its coordinates or, more sophistically, by just entering the name of the place and choose one from a list of sugestions.

3.2.2 Browsing

This feature consists in a use-friendly browsing map to look up for references located around the globe. It has some usability characteristics like a spidefier to spread references that to close from one another, it has clustering to join close references in a cluster when a low zoom is applied to the map and finally an info-window to show detailed information about each reference.

3.2.3 Searching

Searching is an important feature in every system that have a considerable amount of data. Geobiblio provides two forms of search for references near a location: by a search term, it is possible to use a search box with an

auto-complete feature to enter some location term or by some location coordinates, the user enters some coordinates and the system returns references that have scientific studies made near that point, ordered by their distance to it. Additionally it is even possible to restrict the maximum search distance from the central point of search by some amount of Kilometers to retrieve only references in that coverage radius.

3.2.4 External service integration

Geobiblio provides two services for sharing the public topics information: by a dedicated web-page or by using the WFS protocol. After a topic is made public, two new links will appear on the topic information page, each one for the respective service.

The first one shows a map with the references associated to a particular public topic in their respective locations on the map. It has some filtering functionalities like show hide or show layers of references by their type and by subtopics. With this service it is possible for an external *website* to embed the map simply by using the **iframe** *tag* with the provided *url*.

The second way consists on the use of the WFS protocol. This method allows any external system to retrieve public data using this protocol. Third-party systems like ArcGIS and QGIS use WFS to retrieve data and its associated locations and in complement with the WMS service to retrieve maps from external providers on the web. The WFS protocol has several capabilities, some to retrieve data and some to upload or update data on an external server, however as only the functionalities to retrieve data were needed, only those were implemented and also some of the filtering functions described in the WFS protocol.

3.3 Used technologies

Geobiblio was built having in consideration the use and integration of open source technologies.

3.3.1 Aigaion2

Aigaion2 [7] is an open-source bibliographic database management system built using PHP and with a MySQL database management system supporting it. It is a complete bibliographic system supporting various types of references and their related bibliographic data. Geobiblio was built by using this bibliographic system as its main core and it was extended to support geo-referencing.

3.3.2 Google Maps API

Google Maps [12] is the map service from google, it is free and it was built to be easily integrated on the code of another system by using javascript. Because of this, Google Maps was the better choice to use on Geobiblio and it supports the browsing and the filtering of references on a map.

3.3.3 MarkerClusterer v3 and OverlappingMarkerSpiderfier

This two technologies where used to improve the usability of browsing the references on google maps. The MarkerClusterer [13] allows the join of several references that are close from one another into a cluster when the map has a low zoom. The OverlappingMarkerSpiderfier [14] scatters references that are too close from one another when the map has a high zoom.

3.3.4 Geonames

Geonames [15] is a system that provides web services to retrieve information about a vast range of types of places around the globe. Those places can be the obvious ones like cities or countries and others like rivers, mountains, etc. It is possible to retrieve from Geonames [15] information like the coordinates of those places, population in the case of cities, names of those places in other languages than English. Besides this a huge advantage of this system is that every place has a unique *id* which makes possible to reference any place without ambiguity especially in cases where two different places have a similar name.

3.3.5 MySQL

MySQL was used because it is one of the most used database management systems worldwide, it is free use and mostly because Aigaion2 supports it. This database management system already supports geo-spatial data in its core without the need to add any extension to it and it has a vast set of functions that can be used on the geo-spatial data.

4 Demonstration

In order to test the new georeferenced system, a database management system with the occurrence of *tsunamis* was created. This database management system consists with a collection of references to literature about *tsunamis* that occurred on the coast of Portugal since the beginning of its history. This database management system is currently available at http://146.193.41.154/aigaion2_rootdir/index.php/onlinetopics. The development of the database management system also allowed the validation of the changes to Aigaion2 described in the remaining of the section

After adding the references to the database management system, it is necessary to add the geo-referenced information. For that we can use the **add locations** button in some reference page to proceed to the page were locations can be added.

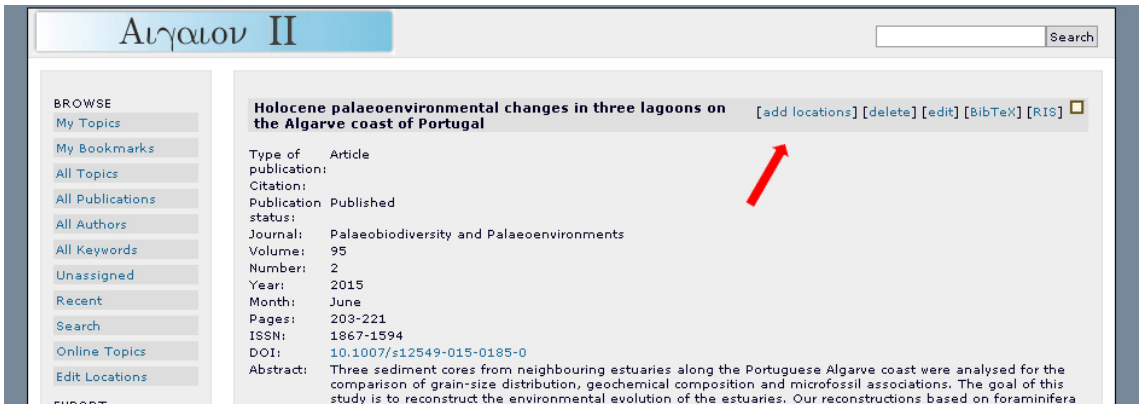

Figure 3 - Adding new locations

Once in the page to add new locations, we can choose a new location by 3 ways: the first consists in simply clicking 2 times on some point in the map, on the second method we can choose a location by entering it's coordinates as seen on the left picture of figure 4 and the last method and the most sophisticated one consists

in just tipping the of the location the search box a choosing one of it's suggestions like on the right picture of figure 4.

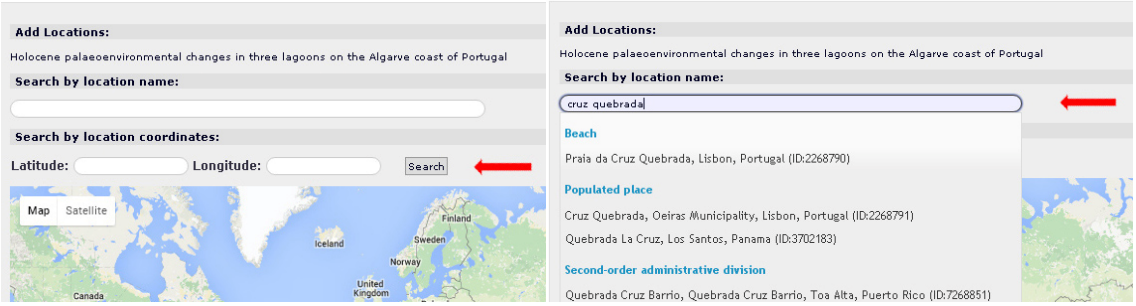

Figure 4 - Choosing a new location

After choosing a location, a box with its information appears on the bottom of the map and then we can change its name before submitting it.

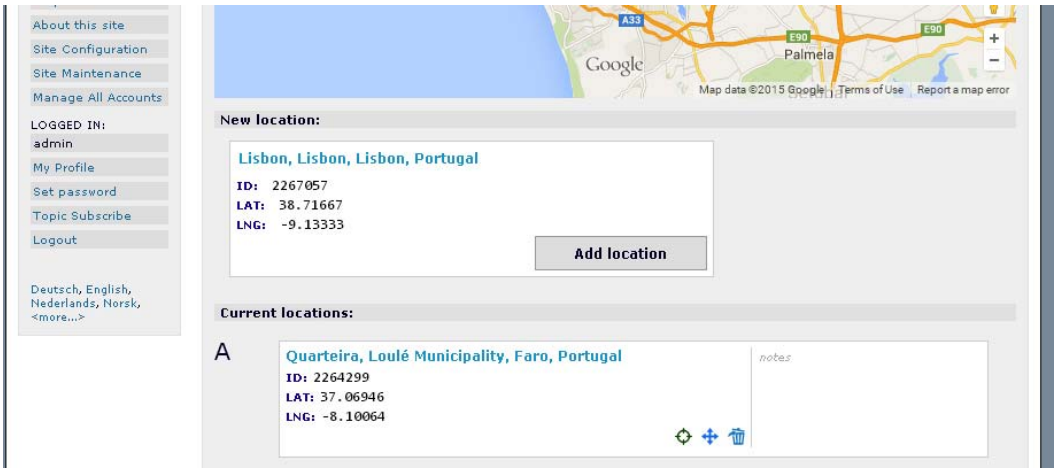

Figure 5 - Submiting a new location

With some georeferenced data on the database management system, we are able to do searches by some specific location. Using the search page it is possible to find bibliographic references that are contained in some circular area. We can use the search box (1) to type the name of a location or we can write the latitude and longitude coordinates on the respective box (2), then we just need to select a specific range from the location point were the system will try to find references that have locations inside that circular area.

Figure 6 - Search by location

For sharing to everyone the georeferenced data about the references of some topic we just need to access the topic edition page and switch on the respective button.

Figure 7 - Topic edition

After that two new links will appear on the topic information page were we can access to an online map with the locations of the references of that topic and also some filters on the left where we can choose references by type or references by belonging to some subtopic. The second link is meant to be used by a third-party system that uses de WFS protocol. It can be used to access to the same data that are available on the online map, like the location coordinates, reference name and type, it's DOI or some URL and also some notes about it.

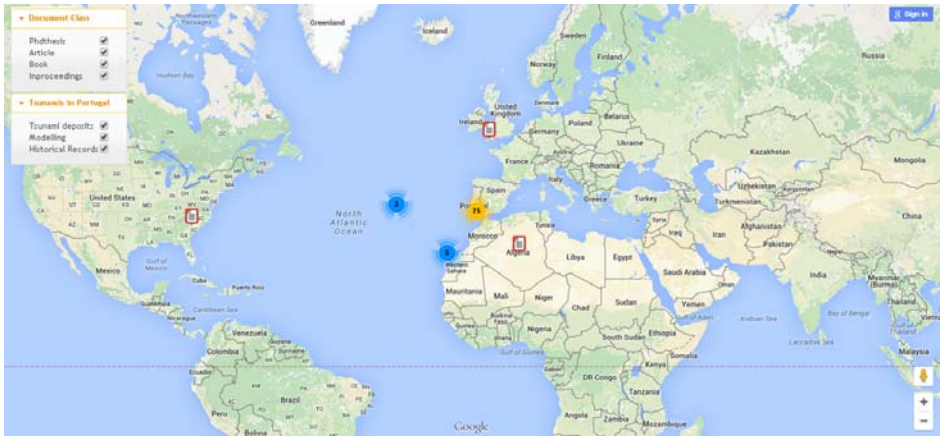

Figure 8 - Map of a shared topic

5 Conclusion and future work

A geo-referenced database management system is an important tool to search and navigate and search through a vast list of references especially when those references are related to some specific scientific areas that have research made in those areas of interest. By studying the actual bibliographic systems we couldn't find a system that fulfils the requirements that a good georeferenced bibliographic system should have. Most of the systems only cover a specific area, are proprietary and run locally on desktop computers. These problems lead to other research institutes to build their georeferenced bibliographic system from scratch. With Geobiblio, research institutes don't have to build their systems from scratch anymore and they have the possibility to georreference any document on any part of the globe.

Geobiblio is based on the Aigaion2 bibliographic system where the new interface with a map and the functionalities to add new locations where successfully implemented. Also, a new search tool, to find references with content related to a specific location, proved to be well implemented and useful. The sharing of information using the WFS protocol was implemented, allowing GIS applications to retrieve and filter data from an Geobiblio online server. This proved to be a nice feature that georeferenced bibliographic systems usually don't have and that allows third-party programs like QGIS to retrieve information.

This was a project that could not be made without some important external technologies, like the Google Maps and the Geonames service. The maps provided by Google and its javascript interface made possible to add a map to the system where to location of references could be seen and also it provided tools to easily manipulate the data shown on the map. The Geonames service was also essential too, it provided an huge catalog of location names and most of all an id to each location that can be used to reference a location without ambiguities. Also important too was the search service provided by Geonames that made possible to build a search box were location names were given as suggestions each time the user types the letters of some location. Other external technologies where useful too like jQuery that made the javascript codding easier and also the libraries MarkerClusterer v3 and OverlappingMarkerSpiderfier despite being less important, made the navigation on the map when there are lots of references agglomerated in a tiny space more easier.

At the end, we can consider that the goals we archived for at the beginning of this project where successfully implemented and also this project proved too that it is possible to build a geo-referenced bibliographic system using only open-source software.

Various types of enhancements could be added to improve this project. Those enhancements could be made in the areas of the information stored about each location, search functionalities or when adding new references or georreferencing them. For these areas we present a list of some improvements that could be added to this project:

- Indexing a location to a reference having an arbitrary shape.
- In the search functionality, there could be a way to use preset shapes like rectangles or circles and use them to define on the map the search area. Also, the possibility to draw an arbitrary shape on the map to restrict the search area for references could be an important feature.
- The possibility to search for references that are restricted to the exact limits of some location.
- The possibility of using a different coordinate system.
- The construction of a parser for scanning the documents and automatically add the bibliographic information and also to georreference it.

Bibliography

[1] Apache OpenOffice Wiki, "Bibliographic/Software and Standards Information," [Online]. Available: https://wiki.openoffice.org/wiki/Bibliographic/Software_and_Standards_Information. [Accessed 12 08

2015].

[2] "BibTex," [Online]. Available: http://www.bibtex.org. [Accessed 08 09 2015].

[3] J. . W. Karl, J. K. Gillan and . J. E. Herrick, "Geographic searching for ecological studies: a new frontier," *Trends in Ecology & Evolution,* vol. 28, no. 7, p. 383–384, July 2013.

[4] D. Yoshida, V. Raghavan, K. Kuwahara and A. Yao, "Implementing Spatially Enabled Bibliographic Database Using Open Source Software," in *Proc. The Regional Conference on Digital GNS*, Bangkok, 2003.

[5] J. Schmitt and B. Butler, "Creating a Geo-Referenced Bibliography with Google Earth and GeoCommons: the Coos Bay Bibliography," *Issues in Science and Technology Librarianship,* no. 71, Fall 2012.

[6] Office of National Marine Sanctuaries, "Papahanaumokuakea Marine National Monument Spatial Bibliography," [Online]. Available: http://www.pmnmims.org/SpatialBibliography/. [Accessed 20 10 2014].

[7] 2008. [Online]. Available: http://sourceforge.net/projects/aigaion/. [Accessed 10 07 2015].

[8] Open Geospatial Consortium Inc., "OpenGIS Web Feature Service 2.0 Interface Standard," 2010.

[9] Esri, 2015. [Online]. Available: http://geocommons.com/. [Accessed 10 08 2015].

[10] USDA Agricultural Research Service Jornada Experimental Range, 2013. [Online]. Available: journalmap.org. [Accessed 10 08 2015].

[11] J. Karl, "JournalMap: map-based discovery of scholarly literature," *Exchanges Our Ideas, Research and Discussion Blog,* 7 September 2014.

[12] Google, "Google Maps JavaScript API v3," [Online]. Available: https://developers.google.com/maps/documentation/javascript/tutorial. [Accessed 25 10 2014].

[13] "js-marker-clusterer," [Online]. Available: https://github.com/googlemaps/js-marker-clusterer. [Accessed 07 09 2015].

[14] "OverlappingMarkerSpiderfier," [Online]. Available: https://github.com/jawj/OverlappingMarkerSpiderfier. [Accessed 08 09 2015].

[15] "GeoNames," [Online]. Available: http://www.geonames.org/. [Accessed 24 09 2015].