Atλας, a Gazetteer Linking Archaeological Collections

Jean-Sébastien Gros

British School at Athens, IT department, Souedias 53
GR10676 Athens, Greece
itofficer@bsa.ac.uk

Abstract. Atλας, is a toponymic database used as reference for projects and ar-chaeological corpora of the British School at Athens. Mainly stored in a MySQL database it is synchronized with a dedicated section in the museum collection management system EMu. External projects and databases are linked with a great variety of tools: APIs in major standard syntaxes for semantic corpora or even included in RDB system. The project includes an autonomous web interface allowing synthetic and geographic displaying as well as collaborative maintenance of the data.

Keywords: Place names, Toponyms, Digital gazetteer, Linked Open Data

1 Introduction

One of the primary goals of the digital policy of the BSA is to empower and fully integrate the current databases with the flexibility of the semantic web. The main section I discuss here is the working package Atλας, devoted to the development of a united toponymic database in its interrelation with the data of the BSA as well as with a wider community projects.

2 Background

In 2007 the European Funding Program found the CIRCE project in order to digitize the cult places of ancient Greece. Among the technical requirements was the geo-localization of the sites. Unfortunately the information source was based on archaeological reports and paper publications that dated from the 19th century until the present day. Naturally, information on the localization in the reports was provided by textual place names and not geographic coordinates. Therefore, to digitize the localizations in machine readable format the place names need to be geocoded. Recently tools for geocoding are widely accessible: i.e. Google Map API. But, this wasn't the case at that time. Moreover, since the sources date from the beginning of systematic archaeology in the 19th century, they refer to place names that may have been altered. In which case, modern geocoding is difficult to implement. The frequent modification of localities or administrative

units since the 19th century has been a significant constraint to the project. To answer this need of massive geo referencing of historical data we first created a reference base included in a RDB model. In 2009 a joint project of the BSA and EfA aimed to catalogue the results of excavation and research work carried out in Greece, by the Archaeological Service and Greek universities as well as by the Foreign Schools active in Greece. This has resulted in Archaeological Reports online `chronique.efa.gr`. The Reports, generated from excavation reports, are linked to the Atλας toponyms dataset to get coordinate information. In 2012 a third archaeological project focus on the Cyclades area (SCC) made the use of this geocoding tool. In the meantime the BSA has intensively worked on migrating all archaeological and corporate data into a global database known as the collection management system EMu. Part of this RDB model was a 'gazetteer' and a 'site' module that were mapped to Atλας and gradually merged in 2016.

Since the last decade gazetteer's projects make accessible their data online. For exemple `pleiades.stoa.org` an extension of the Classical Atlas Project (1988-2000), which produced the Barrington Atlas of the Greek and Roman World [2] had become the reference main reference for the Greco-Roman world. A first thought could be to fully take advantage of the work but although coming with an impressive dataset of nearly 40000 locations Pleiades focus on ancient places and in a very wide geographical frame. This couldn't fulfill our purpose, because recording the archaeological campaigns in Greece required a fine grain geographical repertory that covers the entire Greece including all the modern names. However, because putting into relation excavated places and ancient place names is a major problematic in classical archaeology, we couldn't ignore the ancient gazetteer that offered Open Data Link functionalities. A solution of *'gazetteer alignment'* is supported by the Pelagios project and provides an interconnectivity between the projects but also to a wider corpus of texts and objects [1].

3   Sources

The first source of data was the NGA's GEOnet Names Server with 43673 entries for Greece `geonames.nga.mil/gns/html/namefiles.html`. This open access data source is regularly updated and constitutes the core of most of historical gazetteers. The coverage of these data is global and can therefore allow us to consider the expansion of the area. The coordinates are expressed in decimal degrees in the WGS84 system, the commonest format. The accuracy observed when handling these data on Greece is of the order of +/-500m. The data cover all types of features. Of the 43673 Greek toponyms first imported, about 15000 are populated places, 4000 rivers, 2800 capes, 2200 bays, 1500 mountains, 720 peaks, 700 hills and other less numerous elements. The referencing of the natural features is an asset in the archaeological project.

The second important source constitutes the administrative units and populated places. Originally 7217 entries covering the all territories were provided by Eurostat, but now this information is combined with those available from

the Hellenic Ministry of Interior. Administrative Reconstruction is distributed through the official open access data platform `geodata.gov.gr`. This layer of data was then merged with the former Geonames set and provided a higher data quality by: 1 assuring the entire coverage of the territory, 2 introducing a systematic a hierarchical organization, and 3 refining geolocalisation.

Minor database sources also contributed to Ατλας, among them the data provided by the Institute for Neohellenic Research "Name Changes of Settlements in Greece" `pandektis.ekt.gr/pandektis/handle/10442/4968`. About the 4000 records of official changes are in this list and more than the third have been merged into Ατλας. In 2015 various mapping tables were added to ensure the link with most of the known digital gazetteer projects such as the ID mapping between Geonames and Geonet Names thanks to Marc Wick, or the equivalence between successive official codes for settlements thanks to E. Gadolou. Lastly, I should mention the contribution of substantial human processing of the data for these projects since 2007.

4   Structure

Ατλας could be structured into one main core data set - the toponyms - but if we take into account a help dataset and various LUTs, Ατλας can be described as a classic RDB. First are the 'CODES NUTS and ADM1' base on Eurostat and resolving codes in languages of English, French, Modern Greek as ISO843 transliterated plus a transliteration without diacritic of the 51 nomoi of the Kapodistrian organization. Feature Designation Code Help base on NGA code is augmented with administrative division codes specific to Greece: DEME and DD as well as ANS (archaeological site).

Regarding the TOPONYME table a more detail description of the main fields can be given:

- **ID** shared by the following projects in order to help the localisation of their objects: `www.chronique.efa.gr`; `circe-antique.huma-num.fr`; `cs.ha.uth.gr`.
- **GEONAME_ID** equivalence to Geonames.
- **BSA EMu IRN** equivalence to BSA collection database.
- **NGA GEOnet Names UFI** Unic Feature ID from the GNS with exception 0 to 4999 that are newly created for Demos 'DEME' (communities or municipalities) and their subdivision demos districts 'DD'.
- **NUTS** "Nomenclature of Territorial Units for Statistics" from Eurostat standards.
- **LAU1** Local Administrative Units 1, equivalent to Greek OTA (Οργανισμός Τοπικής Αυτοδιοίκησης) meaning δήμοι and κοινότητες.
- **LAU2** Local Administrative Units 2, for the Greek case equivalent to Δημοτικό Διαμέρισμα et Κοινοτικό Διαμέρισμα (Τοπικά Διαμερίσματα since 2006).
- **LAU3** (not originally in Eurostat dataset) full administrative code for the settlement according to the Kapodistrian organization.
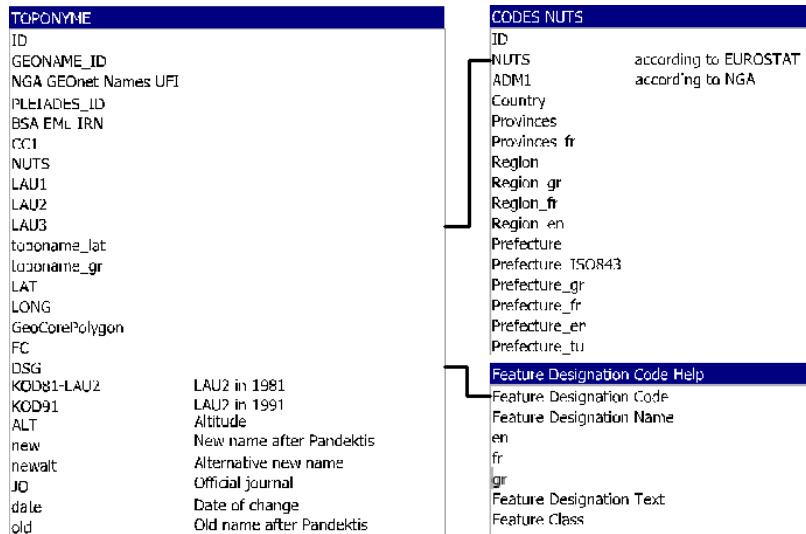
Fig. 1. UML diagram of Atλας

- **Toponame_lat** Names including variant names in Latin characters, separated by comma and the first being the preferred form. Former names also appears and are follow by explanation details in parenthesis eg (anc).
- **Toponame_gr** Names including variant names in Greek characters.
- **LAT** latitude, using WGS-84 system and express in decimal degrees.
- **LONG** longitude.
- **GeoCorePolygon** area of the feature express in array of coordinates.
- **FC** Feature Classification from GNS.
- **DSG** Feature DeSiGnation code from GNS.

5 Accessibility, exchange standards

One of the requirements was the creation of both human and computer readable formats. The web interface allows consultation either through queries or map navigation using the most common open APIs and Javascript libraries as those provided by Googlemap, Mapbox or Leaflet. One of the crucial characteristic of the web interface is to constitute the user friendly front end for managing the data. The tool is developed with the help of Joomla CMS to which we add the Fabrik plugin to manage MySQL database. Thanks to this structure the web interface allows quick and easy editing access to all participating institutions.

With the aim to fully integrate the linked data principles we established URIs base on the original ID of the original database in the following format:

```
at-las.bsa.ac.uk/place/ + ID
```

A dedicated PHP page answers the URI with a formatted web page with basic information of the record as well as a map interface and link to pages or URIs of corresponding data in other related projects. Our goal is also to provide data access in the most common syntaxes and format. Already available is geojson, where a user can open record information using a REST webservices. To obtain a geojson file of all places marked as archaeological site the user has to use URL `atlas.bsa.ac.uk/API/geojson.php?dsg=ANS`. Numerous queries are possible; to look for a single record the ID should simply be included in the URL `atlas.bsa.ac.uk/API/geojson.php?ID=36828`. This API allows online GIS programs or even GIS software to directly use the data set of Atλας. We plan to expand the functionality of the API allowing format selection (XML, CSV etc) from the URL GET property.

6   Conclusion

Atλας was first created as a common georeferencing tool and a standard for restricted projects but has finally been developed as an open data source available to the wider community. Being a highly qualitative and the largest united source for historical geo reference it was therefore natural to provide full open access to scholars and any academic project. The technical development policy is to apply as much as possible the solutions and standards use in other digital gazetteers such as `pleiades.stoa.org`, `gazetteer.dainst.org`, `www.getty.edu/research/tools/vocabularies/tgn`.

In this effort some tasks remain, particularly to increase links, for example, a link in future with the 'archaeological cadaster' being developed by the Hellenic Ministry. An important task is also to enrich data in quantitative manner i.e. entering polygon coordinates, or adding records of archaeological surveys and archaeological sites. This alone would consolidate Atλας as being part of the Linked Open Data paradigm to empower projects by annotating shared heterogeneous resources with the crucial spatial link.

References

1. R. Simon, E. Barker, L. Isaksen, and de Soto C. Linking early geospatial documents, one place at a time: annotation of geographic documents with recogito. *e-Perimetron*, 10(2):49–59, 2016.
2. R. Talbert and R. Bagnall. *Barrington atlas of the Greek and Roman world.* Princeton, N.J: Princeton University Press, 2000.