Raffaella Afferni*

The experience of Geolat group about the project DAGOClaT – Digital Atlas with Geographical Ontology for Classical Texts

Keywords: Geography; Ontology; digital maps; classical texts

Summary: *DAGOClaT* is a project for the creation of a Digital Atlas with Geographical Ontology for Classical Texts, supported by the Compagnia di San Paolo, and involving institutions and specialist from several fields: geography, philosophy of science, computer science and classical studies. *Geolat* (Geography for Latin literature) is the inter-disciplinary, inter-university and international group founded to develop *DAGOClaT* and to enrich with geographical metadata the *corpus* of texts written in Latin in the classical period. Thanks to Web 2.0 technologies the researchers could be able to deploy geographic information, web cartography and analytical design in thematic mapping. The aim of *Geolat* group is to make this information available in digital map libraries accessible not only to scholars and teachers, but also to the general public.
The project is now in a start-up phase and the main activities are in the perspective of focusing and developing methodologies to manage the complex cartographical and graphic historical information dealing with places named in Latin texts.

DAGOClaT is a Digital Atlas project to enrich with geographical metadata the *corpus* of texts written in Latin in the classical period. This project is possible thanks to the financial support of the Compagnia di San Paolo, is now in a start-up phase and is produced by an international, inter-university and inter-disciplinary work team named "*Geolat*" (Geography for Latin literature). Italy, United Kingdom and Belgium are the countries in which work the members of *Geolat* (Università del Piemonte Orientale, Università di Roma – Tor Vergata, Politecnico di Torino, Durham University and Universite de Liége[1]) whose main interests are in geography, philosophy of science, classical studies and computer science.

The choice of focusing the research on Classical world is linked to the interest that in human studies has the use of two languages, Latin and Greek, to link the European culture. Geography of Classical world describes places which could be said as a whole as "Greek-Roman places". However, the *corpora* of texts written in Latin and Greek are so large that a single project, as *DAGOClaT*, is not able to deal with. So *Geolat* group decided to start with the building of a geographical ontology for one of the two classical languages. Latin was chosen for three main reasons: it was the unifying language of Europe; it allowed relationships and contacts among countries and peoples of the entire Mediterranean area, as well as the Middle East and India; its value is recognized by Council of Europe – Language Policy Division (http://www.coe.int/t/dg4/linguistic/langeduc/le_platform/boxc3-foreign_en.asp).

Regarding the texts written in Latin, the project will build a global *corpus* for digital Latin, even if only a fraction of it will be effectively used. This is because the building activity is a single one, substantially independent from the number of collected texts; while the number of texts to be annotated radically influences the global times and the costs of the project. The result is that the

*Geolat, Dipartimento di Studi Umanistici, Università del Piemonte Orientale [raffaella.afferni@lett.unipmn.it]

[1] In particular, LASLA (Laboratoire d'Analyse Statistique des Langues Anciennes) participate DAGOClaT project. LASLA is a centre of world renown for linguistic analyses and morphological parsing on Latin texts.

global amount of texts to be annotated must be dimensioned on the available resources within the "call" boundaries.

Objectives of DAGOClaT project

The project has three main objectives. The first is the creation of a geographical digital reference "tool" for the classical world. Some of the most famous (active or no more active) works devoted to classical world geography are for example the Barrington Atlas of the Greek and Roman World (http://www.unc.edu/depts/cl_atlas/) Orbis (http://orbis.stanford.edu), GAPVis (Google Ancient Places Visualization, http://nrabinowitz.github.com/gapvis/index.html#index) and Pleiades (http://pleiades.stoa.org/). Other geographical databases (not related with ancient world geography) are Geonames (http://www.geonames.org/) and Dariah (http://dev2.dariah.eu/e4d/).
DAGOClaT will take the existing geographical knowledge about places of Classical World (in Europe and the Mediterranean area going up to the Middle East and India) contained in these specialized dictionaries, encyclopaedias and atlases and transfer them into a new medium (Stillwell, 1976; Müller, 1855-61; Riese, 1878). As Mark and Smith (2004) state, it will be a transfer of information contained in unsurpassed authoritativeness works into a new container that allows a widespread access and use in the digital world.
The second target of *DAGOClaT* regards the geographical markup. The aim of the project is to provide a geographical semantic markup of the *corpus* of Latin texts from III century B.C. up to about VIIVIII century A.D., which include about 10 million words, 1200 works of 800 authors. The geographical annotation will be done *ex novo* by means of a software environment to associate semantic metadata – expressing information about places – with texts passages. These will contain not only the meaning of the name, the occurrences in ancient works, the ancient name of the geographical region to which the place belongs, the references to recognized itineraries, the people who travelled along it; but also the current name, the geographic coordinates, the State or Nation, the bibliographical references. And *Geolat* group will do this geographical markup using free software, open standards, CC (Creative Commons) licenses and open data, in order to establish a *de facto* standard, which anybody can adopt for his own tagged *corpus*.
The ultimate goal of *DAGOClaT* is ambitious: the creation of a website allowing free access and use of the research results. The website will contain texts and maps enriched with layered info and didactical tools for users (not only scholars but also territorial organization and more generally common people), as well as the reports and studies written by researchers. The texts annotated with geographical metadata will provide knowledge that could be published according to the Open Linked Data best practices, which currently represent an innovative frontier of the Web (Heath and Bizer, 2011). If the project will attract the interest of the users, the system will remain available even after the project funding ends (for browsing, for input of new data as the science always progresses, and for output to computer systems accessing the data).
In addition to these three main objectives, the project aims to gain some significant indirect and intermediate results. The first regards the creation and publication on the Web of a global digital geographical resource for Classical World, containing Latin names of place from the III century B.C. to the VII-VIII century A.D. and connecting them to current geographical information. Such a resource does not currently exist and its publication could be both a short and a medium term goal, to be built in steps, through versions and allowing collaboration with scholars not belonging to the project.

The second implies the creation and publication on the Web of one single *corpus* of Latin language available to the public (also through Europeana http://www.europeana.eu). Such a large and scientifically controlled *corpus* today is nowhere to be found for scholars. To build such an archive, *Geolat* group will link to some existing digital libraries, each one of high scientific level, as digilibLT (Digital library of late-antique Latin texts), Musisque Deoque, BIA (Bibliotheca Iuris Antiqui) and DIGIMED (Digital Medieval Texts).

DigilibLT (http://www.digiliblt.unipmn.it) is a project to provide a free and complete database of late-antique Latin authors and works, as well as an exhaustive canon. Texts can be downloaded freely, which will allow individual scholars to work on their areas of interest with maximum flexibility. All text use both TEI and Beta coding standard to allow users to search the texts with main known software resources (as TLG and PHI databases on CD-ROM). DigilibLT contains pagan prose from the II to the VII century A.D. and, in *DAGOClaT* project, will be responsible also for the management of literary prose texts from III century B.C. up to II century A.D. (about 6 million words), and for texts from *grammatici latini* (1 million words). Musisque Deoque (http://www.mqdq.it) includes in the Web site, continuously updated with new critical apparatus, the whole Latin versification heritage from III century B.C. up to Renaissance (2 ½ million words). BIA (http://bia.lex.unict.it/) includes extant Latin texts of legal argument and is made up of three archives: the electronic text of Digesta, divided into paragraphs and containing almost all Roman juridical sources in full text; a second archive of other juridical sources is made out of that one carried out in Linz University (Austria) and bibliographic archives including the whole scientific production from 1940 to 1998 regarding the Roman law, other ancient laws and the general history of the ancient world. The last digital library is DIGIMED (http://www.tdtc.unisi.it/digimed/) that contains Medieval Latin texts, with particular attention to geo-historical subjects.

Further implications of DAGOClaT project

In the proposal of the project, the principal investigator of *Geolat* group states that *DAGOClaT* has interesting implications in four main directions: science, education, public use, economic and tourism activities (Università del Piemonte Orientale A. Avogadro, 2012).

Regarding science, a Digital Atlas of the *corpus* of Latin texts offers not only new opportunities both for linguistic and literary studies with an interdisciplinary nature, but also for similar studies in other linguistic/textual areas. Moreover a same or similar research project could use tools and data produced by *DAGOClaT* opening new interdisciplinary connections with linguistic and literary studies and adding new GIS tools for the comparison (presence and form) of geographical names in both geographical texts and cartography ranging from Antiquity to Renaissance (Università del Piemonte Orientale A. Avogadro, 2012).

As for education, the development of the project will provide new didactical tools for teachers and scholars of Classics, History and Geography. *DAGOClaT* will offer a formal ontology that, thanks to the usage of well-known standards, will represent a shared and reusable semantic reference for both humans and software tools.

The project could also have a public use. Data included in *DAGOClaT* may be useful to common people that, through a visual interface based on digital maps, are able to display e.g. “the mountains of Italy mentioned in [the Latin writings of] Caesar” or “the towns of Gaul mentioned in the writings of I century A.D.” and so on. This type of information could also be useful to

travellers that plan a trip. In this case maps do not offer only information on distances and journey times, but also for discovering places as were cited, named and known in texts written in Latin in the classical period. *Geolat* group aims to allow a student of classics to study the location of the rivers mentioned in the poetry of Horace and, at the same time, a mayor to collect historical information about his town and its surroundings in order to promote a cultural event.
And the value of the project in economic and tourism activities is the last, but not the least, of the four important implications of *DAGOClaT*. Geographical metadata allow creative activities by the user (e.g. travel agencies or travellers could use the maps to build cultural tourism for example "on the paths of Caesar in Gaul") and also permit to obtain suggestion of texts mentioning other historical characters who have been in the same places at the same time, or even a virtual or real tour (e.g. across those places of Caesar in Gaul). Since the metadata do not describe only urban names, but identify any geographical information (lakes, rivers, mountains, regions, etc.), it will be possible to discover less known or marginalized lands and to show their own cultural importance. The Lisbon Treaty declared tourism an area of interest of European Union which supports, coordinates, supplements actions taken by members (art. 2E), and encourages the creation of a favourable environment for the development of undertakings in this sector (art. 176B).

The first step of the project: the Geographical Ontology

Moving from these assumptions, the first steps of *Geolat* group were made in April 2013 and the project is now in a start-up phase. The main activities of the researchers are in the perspective of focusing and developing methodologies to manage the complex cartographical and graphic historical information dealing with places named in Latin texts. So the group is now dealing with choices about the geographical ontology.
The word ontology is a compound word, composed of *onto-*, from the Greek *ὄν, on* (gen. *ὄντος, ontos*), i.e. "being; that which is", which is the present participle of the verb *εἰμί, eimi*, i.e. "to be, I am", and -λογία, *-logia*, i.e. "science, study, theory" (Online Etymology Dictionary, 2013). Encyclopaedia Britannica states that ontology is «the philosophical study of being in general or of what applies neutrally to everything that is real. It was called "first philosophy" by Aristotle in Book IV of his Metaphysics. The Latin term *ontologia* ("science of being") was felicitously invented by the German philosopher Jacob Lorhard (Lorhardus) and first appeared in his work Ogdoas Scholastica (1st ed.) in 1606. It entered general circulation after being popularized by the German rationalist philosopher Christian Wolff in his Latin writings, especially Philosophia Prima sive Ontologia (1730; "First Philosophy or Ontology")» (Encyclopaedia Britannica, 2013).
Traditionally listed as a part of the major branch of philosophy (known as metaphysics) where the term represents a systematic account of existence (Enciclopedia Treccani, 2013) ontology is also utilized in computer science and information science. In this case it formally represents knowledge as a set of concepts within a domain, and the relationships between pairs of concepts. It can be used to model a domain and support reasoning about concepts. For Gruber (1993) an ontology is a "formal, explicit specification of a shared conceptualisation" and for knowledge-based systems, what "exists" is that which can be represented (Cocchiarella, 1991; Guarino, 1995; Studer, Benjamins and Fensel, 1998). Ontology provides a shared vocabulary, which can be used to model a domain, that is, the type of objects and/or concepts that exist, and their properties and relations (Arvidsson and Flycht-Eriksson, 2002). Thus, we can describe the ontology of a program

by defining a set of representational terms. Gruber (1993) states that in such ontology, definitions associate the names of entities in the universe of discourse (e.g., classes, relations, functions, or other objects) with human-readable text describing what the names are meant to denote, and formal axioms that constrain the interpretation and well-formed use of these terms. For this reason as Gruber (2009) states, ontologies are said to be at the "semantic" level and due to their independence from lower level data models, they are used for integrating heterogeneous databases, enabling interoperability among disparate systems, and specifying interfaces to independent, knowledge-based services. In the technology stack of the Semantic Web standards, ontologies are called out as an explicit layer (Berners-Lee, Hendler and Lassila, 2001). Today's W3C Semantic Web standard suggests in particular a specific formalism for encoding ontologies (OWL), in several variants that vary in expressive power (McGuinness and van Harmelen, 2004). This reflects the intent that ontology is a specification of an abstract data model (the domain conceptualization) that is independent of its particular form.

Gruber (2009) states that the key role of ontologies with respect to database systems is to specify a data modeling representation at a level of abstraction above specific database designs (logical or physical), so that data can be exported, translated, queried, and unified across independently developed systems and services. Successful applications to date include database interoperability, cross database search, and the integration of web services.

Regarding in particular geographical information to assist spatial search, geographical ontologies could be interesting tools for researchers (Jones, Alani and Tudhope, 2001; Riekert, 2002; Guarino, Masolo and Vetere, 1999; Schlieder, Vgele. and Visser, 2001). However these types of ontologies are very rare and studies are frequently based on the use of a gazetteer in which a place is normally represented by point locations. As Fu, Jones and Abdelmoty (2005) state «the limited spatial semantics associated with these approaches narrows the scope of their ability to effectively retrieve useful resources for spatial queries». On the other hand a geo-ontology plays a central role in intelligent spatial search on the Web, because it serves as a shared vocabulary for spatial markup of Web resources. It facilitates disambiguation and interpretation of user queries and it also enables the generation of spatial index to support efficient retrieval of Web resources, as well as relevance ranking of retrieved results.

The first step in the process of building the geo-ontology involves the development of an appropriate underlying conceptual model to support the required functionalities. Secondly it involves populating the ontology with enough detail to realise its full potential. Jones, Alani and Tudhope (2001) state that the later task is a significant challenge because it provides the integration of several different data sources that may differ in their underlying structures, their accuracy and the levels of details on representations of the places.

The challenge of Geolat group regards now (July 2013) the building of the geo-ontology for DAGOClaT project and the problems of utilising multiple data sources for its populations.

When the geo-ontology will be completed, the next step (autumn 2013) will be the construction of the software environment version 1.0 for entering the geographical information that will evolve, in the following steps, to version 2.0 in order to manage new tasks.

References

Arvidsson, F. and Flycht-Eriksson, A. (2002). *Ontologies I* [PDF] Available from: http://www.ida.liu.se/~janma/SemWeb/Slides/ontologies1.pdf [Accessed 30/06/2013].

Barrington Atlas of the Greek and Roman World (2000) [WWW] Available from http://www.unc.edu/depts/cl_atlas/ [Accessed 04/07/2013].

Berners-Lee, T. Hendler, J. and Lassila, O. (2001). The Semantic Web. *Scientific American Magazine* May .

BIA (2002) [WWW] Available from http://bia.lex.unict.it/ [Accessed 01/07/2013].

Cocchiarella, N. B. (1991). Formal Ontology. In H. Burkhardt and B. Smith (eds.). *Handbook of Metaphysics and Ontology*. Munich: Philosophia Verlag. 640-647

Dariah (2013) [WWW] Available from http://dev2.dariah.eu/e4d/ [Accessed 04/07/2013].

DigilibLT (2010) [WWW] Available from http://www.digiliblt.unipmn.it [Accessed 01/07/2013].

DIGIMED (2006) [WWW] Available from http://www.tdtc.unisi.it/digimed/ [Accessed 01/07/2013].

Enciclopedia Treccani (2013). *Ontologia* [WWW] Available from: http:// www.treccani.it [Accessed 30/06/2013].

Encyclopaedia Britannica (2013). *Ontology (metaphysics)* [WWW] Available from: http://www.britannica.com [Accessed 30/06/2013].

Fu, G. Jones, C.B. and Abdelmoty, A.I. (2005). Building a Geographical Ontology for Intelligent Spatial Search on the Web, Databases and Applications, 454.

GAPVis (2009) [WWW] Available from http://nrabinowitz.github.com/gapvis/index.html#index [Accessed 04/07/2013].

GeoNames (2006) [WWW] Available from http://www.geonames.org/ [Accessed 04/07/2013].

Gruber, R. (1993). A translation approach to portable ontologies. *Knowledge Acquisition* 5(2): 199-220

Gruber, T. (2009). Ontology. In Ling Liu and M. Tamer Özsu (eds.), *The Encyclopedia of Database Systems*. Springer-Verlag.

Guarino, N. (1995). Formal Ontology, Conceptual Analysis and Knowledge Representation. *International Journal of Human-Computer Studies*, 43(5-6):625–640 Available from: http:// www.loa.istc.cnr.it/Papers/FormOntKR.pdf [Accessed 30/06/2013]

Guarino, N. Masolo, C. and Vetere, G. (1999). OntoSeek: Content-Based Access to the Web. *IEEE Intelligent Systems*, 14(3): 70-80.

Heath, T. and Bizer, C. (2011). *Linked Data: Evolving the Web into a Global Data Space. Synthesis Lectures on the Semantic Web: Theory and Technology*. Palo Alto: Morgan & Claypool.

Jones, C. B. Alani, H. and Tudhope, D. (2001). Geographical Information Retrieval with Ontologies of Place. In *Proceedings of the International Conference on Spatial Information Theory*. London:Springer-Verlag. 322-335.

Mark, D.M. and Smith, B. (2004). A Science of Topography: From Qualitative Ontology to Digital Representations. In M. P. Bishop and J. F. Shroder (eds.), *Geographic Information Science and Mountain Geomorphology*. Chichester: Springer-Praxis. 75-100.

McGuinness, D. L. and van Harmelen, F. (2004). OWL Web Ontology Language. W3C Recommendation [WWW] Available from: http://www.w3.org/TR/owl-features/ [Accessed 30/06/2013].

Müller, C. (ed.) (1855-61). *Geographi Graeci minors I-II*. Paris.

Musisque Deoque (2007) [WWW] Available from http://www.mqdq.it [Accessed 01/07/2013].

Online Etymology Dictionary (2013). *Ontology* [WWW] Available from: http://www.etymonline.com [Accessed 30/06/2013].

Orbis (2012) [WWW] Available from http://orbis.stanford.edu [Accessed 04/07/2013].

Pleiades (2009) [WWW] Available from http://pleiades.stoa.org/ [Accessed 04/07/2013].

Riekert, W. (2002). Automated Retrieval of Information in the Internet by Using Thesauri and Gazetteers as Knowledge Sources. *Journal of Universal Computer Science*, 8(6):581–590.

Riese, A. (1878). *Geographi Latini minors*. Heilbronn.

Schlieder, C. Vgele, T. and Visser, U. (2001). Qualitative Spatial Representation for Information Retrieval by Gazetteers. In *Proceedings of the International Conference on Spatial Information Theory*. London: Springer-Verlag. 336-351.

Stillwell, R. (ed.) (1976). *The Princeton Encyclopedia of Classical Sites*. Princeton: Princeton University Press.

Studer, R. Benjamins, V. R. and Fensel, D. (1998). Knowledge Engineering: Principles and methods. Data and Knowledge Engineering, 25(1-2): 161-197.

Università del Piemonte Orientale A. Avogadro (2012), Digital Atlas with Geographical Ontology for Classical Texts DAGOClaT [PDF]