Department of
Information Engineering
and Computer Science
DISI

UNIVERSITY
OF TRENTO - Italy

DISI - Via Sommarive 14 - 38123 Povo - Trento (Italy)
http://www.disi.unitn.it

A SEMANTIC GEO-CATALOGUE IMPLEMENTATION FOR A REGIONAL SDI

P. Shvaiko, A. Ivanyukovich,
L. Vaccari, V. Maltese, and F. Farazi

May 2010

Technical Report # DISI-10-033

Also: in proceedings of the INPIRE conference 2010

A semantic geo-catalogue implementation for a regional SDI

P. Shvaiko[1], A. Ivanyukovich[2], L. Vaccari[3], V. Maltese[4], F. Farazi[4]
[1]*TasLab, Informatica Trentina S.p.A., Trento, Italy*
[2]*Trient Consulting Group S.r.l., Trento, Italy*
[3]*European Commission - DG JRC, Ispra, Italy*
[4]*DISI, University of Trento, Trento, Italy*

Abstract. In this paper we report our work on the implementation of a semantic geo-catalogue for the spatial data infrastructure (SDI) of the Trentino region, Italy.

Introduction. Within the INSPIRE directive [4], the recently approved regulation on *network services* [3] establishes minimum criteria for the *discovery service* to support search within the INSPIRE metadata elements. However, discovery capabilities are often limited by only syntactically matching the user terminology to the details published by the metadata compiler. The needs to overcome this limitation, both considering multilingual and semantically heterogeneous SDI catalogues, has been identified as one of the key issues for the future of the INSPIRE implementation, especially for the INSPIRE geo-portal[1] [12,1] and architecture [15,17].

This paper focuses on a discovery service implemented by means of the Catalogue Service for the Web (CSW)[2] standard. Several implementations have been already provided for the CSW-based geo-catalogue, including GeoNetwork[3] and Deegree[4]. Specifically, we discuss our practical experience with the implementation of the CSW-based geo-catalogue, starting from the GeoNetwork open source software[2] within the SDI of Trentino. Besides bringing to the actual exploitation by the final users of a basic version of the geo-catalogue, we also experiment, with a semantically enhanced prototype version of the geo-catalogue.

Geo-catalogue implementation. The overall system architecture follows the standard three-tier paradigm with front-end, business logic and back-end layers. The geo-catalogue is being plugged into the existing geo-cartographic portal[5] of Trentino as one of its services. Following the best practices for the integration of the third-party software into BEA ALUI framework[6] (the current engine of the geo-portal of Trentino), all external services are brought together using a portlet[7]-based scheme, where GeoNetwork is used as a back-end. Figure 1 provides the integration view of the system architecture. Specifically, at the front-end, the functionalities are realized as three portlets: *(i)* for metadata management, including harvesting, search and catalogue navigation functionalities; *(ii)* user/group management, which is designed to meet the security requirements of the geo-portal; *(iii)* system configuration, which corresponds to the functionalities of the GAST[8] tool of GeoNetwork. In turn, these functionalities are mapped *1-to-1* to the back-end services of GeoNetwork. Notice that external

[1] http://www.inspire-geoportal.eu/
[2] http://www.opengeospatial.org/standards/cat
[3] http://geonetwork-opensource.org/
[4] http://deegree.org/
[5] http://www.territorio.provincia.tn.it/
[6] http://download.oracle.com/docs/cd/E13174_01/alui/
[7] http://jcp.org/en/jsr/detail?id=168
[8] GeoNetwork's Administrator Survival Tool

applications, that produce metadata, such as ESRI ArcCatalog, can also access (for example, through the Oracle service bus, which handles the authentication, etc.) the back-end services of GeoNetwork.

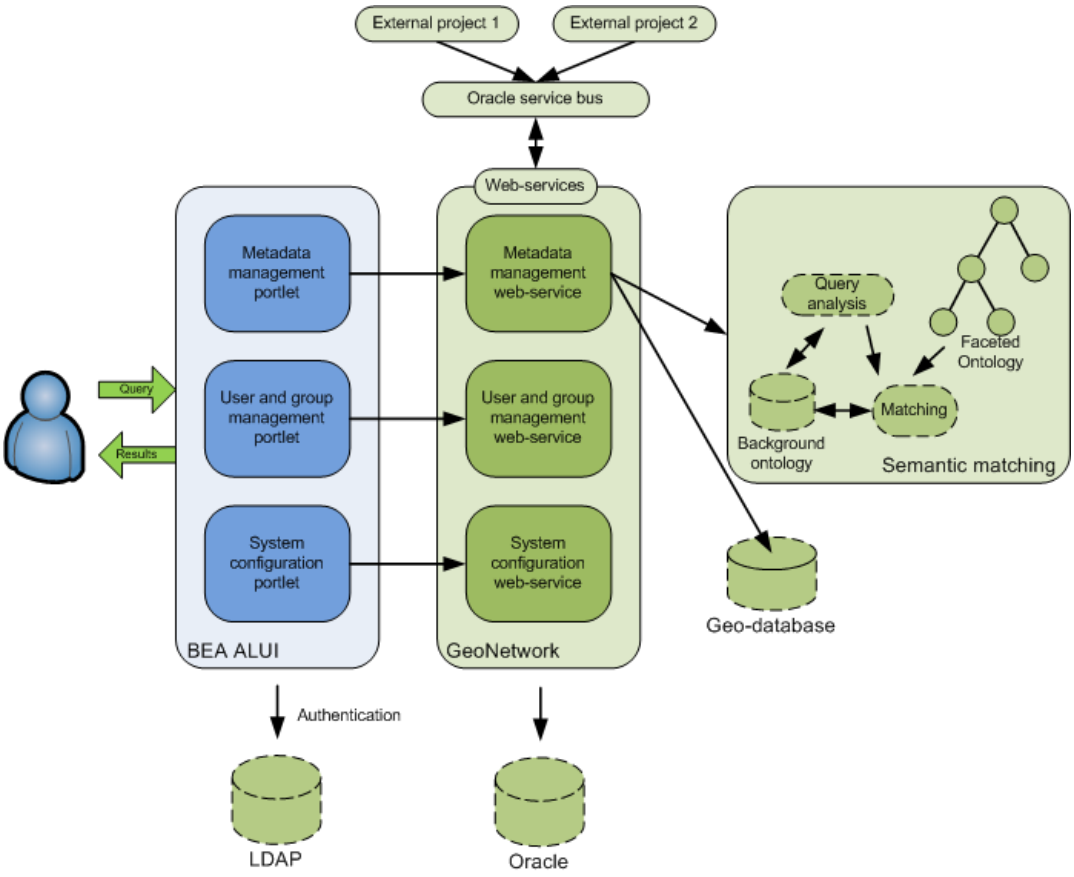

Figure 1. System architecture: the integration view.

Semantic extension for the geo-catalogue. We extend the GeoNetwork catalogue search function by adding semantic query processing methods. In particular, we have analyzed the available work in the field, such as [2,5,11,13,16,17] provided a summary of these in [14] with the identification of work on search [6] and matching technologies [7,8,9] that provide the semantic extension of the geo-catalogue. Specifically, *semantic matching* is a technique for the identification of semantically related information. Given two graph-like structures (e.g., XML schemas) a semantic matching operation identifies the pairs of nodes in the two structures that semantically correspond to each other. For instance, it can identify that two nodes labelled field and meadow are semantically equivalent because the two terms are synonyms in English. This allows similar information to be identified that would be more difficult to find using traditional information retrieval approaches.

S-Match [8] is an example of semantic matching operation, recently released as open source software[9]. The query analysis component (see Figure 1) uses natural language text to search (which can be seen as a classification composed by a single node) and translates it in a formal language according to the knowledge codified in the background ontology. The formal representation of the query is then given as input to the matching component that matches it against a faceted ontology [10]. The faceted ontology is an ontology composed of several subtrees, each one codifying a different aspect of the domain. In our case, it codifies the knowledge about geography and includes (among others) the administrative divisions (e.g., municipalities, villages), the water bodies (e.g., lakes, rivers) and the land formations (e.g., mountains, peaks, hills) of Trentino.

Conclusions. We briefly reported our experience with the geo-catalogue integration into the SDI of Trentino, which we think should be of help also to similar initiatives. Since this is an integration exercise (rather than building a new system from scratch), it required a substantial amount of time for the coordination and integration with the other systems already in place. This allowed us to take advantage of these systems and to arrive at synergetic activities. Some of the major findings include:

- The key technical issues were: handling authentication and authorization procedures, re-implementation of the GeoNetwork front-end under BEA ALUI (including all the usability issues); and re-integrating the already available services for geo-data upload and download into geo-catalogue within the geo-portal.
- GeoNetwork provides the so-called universally unique identifiers (UUID), that complies with the INSPIRE metadata implementing rules. In turn, in Italy there is the IntesaGIS-CNIPA initiative that produced the corresponding technical specifications - *Regolamento Repertorio Nazionale Dati Territoriali*[10] *(RNDT)* - that extend the INSPIRE definition; so that UUIDs (as implemented in GeoNetwork) cannot be used, and that such IDs should contain the IPA prefix[11] of the organization being responsible for the metadata.
- S-Match is being initially designed as a standalone application, so its integration with a production system (GeoNetwork) was implemented through a wrapper that provides web services to be invoked by GeoNetwork. This approach mitigates risks of failure in experimental code while still following strict uptime requirements of the production system. Another advantage of this approach is the possibility to reuse this service in other applications with similar needs.

In summary, we have found it useful to start realizing our geo-catalogue service from the available open-source implementation conforming to the INSPIRE Directive. However, it had to be adapted further based on the available technological infrastructure (such as the geo-portal of Trentino) and specificities of the implementation rules further constrained at the national level, e.g., by RNDT. Future work includes an extensive empirical evaluation of the semantic extension of the geo-catalogue.

[9] http://sourceforge.net/projects/s-match/
[10] http://www.cnipa.gov.it/site/it-IT/Attivit%c3%a0/Sistemi_Informativi_Territoriali/Specifiche_tecniche/Regolamento_Repertorio/
[11] http://www.indicepa.gov.it/

Acknowledgments. This work has been supported by the TasLab network project funded by the European Social Fund under the act n. 1637 (30.06.2008) of the Autonomous Province of Trento. We are thankful to Giuliana Ucelli e Daniela Ferrari for many fruitful discussions on the use of the geo-catalogue within the SDI of Trentino. We are grateful to Fausto Giunchiglia for the project leadership on the side of University of Trento, in charge of development of the semantic layer for the geo-catalogue and for the technical discussions that leaded to the actual realization of this idea. We thank Veronica Rizzi for some technical support. Finally, we are grateful to the colleagues from the production of Informatica Trentina for many fruitful technical discussions on the implementation of the geo-catalogue within the geo-portal of the Autonomous Province of Trento.

References

[1] J. Crompvoets, M. Wachowicz, F. de Bree, and A. Bregt. Impact assessment of the INSPIRE geo-portal. *In Proc. of the 10th EC GI&GIS workshop*, 2004.
[2] I. Cruz and W. Sunna. Structural alignment methods with applications to geospatial ontologies. *Transactions in Geographic Information Science*, 12(6):683–711, 2008.
[3] European Commission, "COMMISSION REGULATION (EC) No 976/2009 implementing Directive 2007/2/EC as regards the Network Services," 2009.
[4] European Parliament, "Directive 2007/2/EC establishing an Infrastructure for Spatial Information in the European Community (INSPIRE)", 2009.
[5] J. Euzenat and P. Shvaiko. Ontology matching. *Springer*, 2007.
[6] F. Giunchiglia, U. Kharkevich, and I. Zaihrayeu. Concept search. *In Proc. of ESWC*, 2009.
[7] F. Giunchiglia, F. McNeill, M. Yatskevich, J. Pane, P. Besana, and P. Shvaiko. Approximate structure-preserving semantic matching. *In Proc. of ODBASE*, 2008.
[8] F. Giunchiglia, P. Shvaiko, M. Yatskevich. S-Match: an algorithm and an implementation of semantic matching. *In Proc. of ESWS*, 2004.
[9] F. Giunchiglia, V. Maltese, A. Autayeu. Computing minimal mappings. *In Proc. of the 4th workshop on Ontology Matching at ISWC*, 2009.
[10] F. Giunchiglia, B. Dutta, V. Maltese. Faceted lightweight ontologies. In Conceptual Modeling: Foundations and Applications, *Springer*, 2009.
[11] K. Janowicz, M. Wilkes, and M. Lutz. Similarity-based information retrieval and its role within spatial data infrastructures. *In Proc. of GIScience*, 2008.
[12] M. Lutz, N. Ostlander, X. Kechagioglou, and H. Cao. Challenges for Metadata Creation and Discovery in a multilingual SDI - Facing INSPIRE. *In Proc. of ISRSE*, 2009.
[13] P. Maué. An extensible semantic catalogue for geospatial web services. *Journal of Spatial Data Infrastructures Research*, 3:168–191, 2008.
[14] P. Shvaiko, L. Vaccari, G. Trecarichi. Semantic Geo-Catalog: A Scenario and Requirements. *In Proc. of the 4th workshop on Ontology Matching at ISWC*, 2009.
[15] P. Smits and A. Friis-Christensen. Resource discovery in a European Spatial Data Infrastructure. *Transactions on Knowledge and Data Engineering*, 19(1):85–95, 2007.
[16] K. Stock, M. Small, Y. Ou, and F. Reitsma. OGC catalogue services - OWL application profile of CSW. Technical report, Open Geospatial Consortium, 2009.
[17] L. Vaccari, P. Shvaiko, and M. Marchese. A geo-service semantic integration in spatial data infrastructures. *Journal of Spatial Data Infrastructures Research*, 4:24–51, 2009.