GeoKnow:
Making the Web an Exploratory Place
for Geospatial Knowledge

by Spiros Athanasiou, Daniel Hladky, Giorgos Giannopoulos, Alejandra Garcia Rojas and Jens Lehmann

The GeoKnow project aims to make geospatial data accessible on the Web of Data, transforming the Web into a place where geospatial data can be published, queried, reasoned, and interlinked, according to Linked Data principles.

In recent years, Semantic Web methodologies and technologies have strengthened their position in the areas of data and knowledge management. Standards for organizing and querying semantic information, such as RDF(S) and SPARQL are adopted by large academic communities, while corporate vendors adopt semantic technologies to organize, expose, exchange and retrieve their data as Linked Data [1]. RDF stores have become robust enough to support volumes of billions of records (RDF triples), and also offer data management and querying functionalities very similar to those of traditional relational database systems. Currently, there are three major sources of open geospatial data in the Web: Spatial Data Infrastructures (SDI), open data catalogues, and crowdsourced initiatives. Crowdsourced geospatial data are emerging as a potentially valuable source of geospatial knowledge. Among various efforts we highlight OpenStreetMap, GeoNames, and Wikipedia as the most significant. Recently, GeoSPARQL [2] has emerged as a promising standard from W3C for geospatial RDF, with the aim of standardizing geospatial RDF data modelling and querying. Integrating Semantic Web with geospatial data management requires the scientific community to address two challenges: (i) the definition of proper standards and vocabularies that describe geospatial information according to RDF(S) and SPARQL protocols, that also conform to the principles of established geospatial standards, (e.g. OGC), (ii) the development of technologies for efficient storage, robust indexing, and native processing of semantically organized geospatial data.

Geoknow is an EU funded, three-year project that started in December 2012. While several research projects, such as LOD2[4], are supporting the Linked Data LifeCycle, Geoknow addresses the key issues of integrating geographically related information on the Web, scalable reasoning over billions of geographic features within the Linked Data Web, as well as efficient crowd-sourcing and collaborative authoring of geographic information. In particular, GeoKnow will apply the RDF model and the GeoSPARQL standard as the basis for representing and querying geospatial data and will contribute to the following areas:

- Efficient geospatial RDF querying: Existing RDF stores lack performance and geospatial analysis capabilities compared to geospatially-enabled relational DBMS. We will focus on introducing query optimization techniques for accelerating geospatial querying by at least an order of magnitude.
- Fusion and aggregation of geospatial RDF data: Given a number of different RDF geospatial data for a given region containing similar knowledge
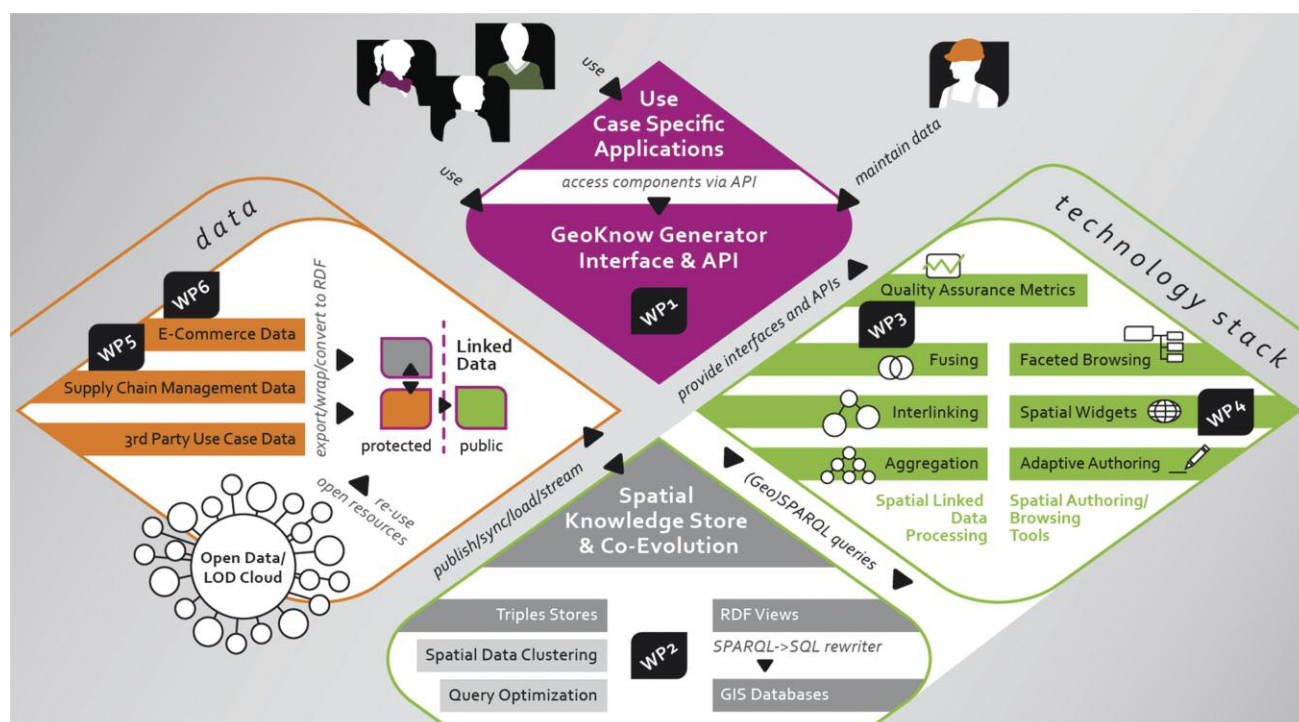

Figure 1: Overview of GeoKnow technologies

(e.g. OSM, PSI and closed data) we will devise automatic fusion and aggregation techniques in order to consolidate them and provide a dataset of increased value and quantitative quality metrics of this new data resource.

- Visualization and authoring: We will develop reusable mapping components, enabling the integration of geospatial RDF data as an additional data resource in web map publishing. Further, we will support expert and community-based authoring of RDF geospatial data within interactive maps, fully embracing crowdsourcing.
- Public-private geospatial data: To support value added services on top of open geospatial data, we will develop enterprise RDF data synchronization workflows that can integrate open geospatial RDF with closed, proprietary data.
- GeoKnow Generator: This will consist of a full suite of tools supporting the complete lifecycle of geospatial Linked Data. The GeoKnow Generator will enable publishers to triplify geospatial data, interlink them with geospatial and non-geospatial data sources, fuse and aggregate linked geospatial data to provide new data of increased quality, and finally, visualize and author link geospatial data in the Web.

The GeoKnow Generator

A prototype of the GeoKnow Generator is already available at http://generator.geoknow.eu. It allows the user to triplify geospatial data, such as ESRI shapefiles and spatial tables hosted in major DBMSs using the GeoSPARQL, WGS84 or Virtuoso RDF vocabulary for point features geospatial representations (TripleGeo). Non-geospatial data in RDF (local and online RDF files or SPARQL endpoints) or data from relational databases (via Sparqlify) can also be entered into the Generator's triple store. With these two sources of data it is possible to link (via LIMES), to enrich (via GeoLift), to query (via Virtuoso) , to visualize (via Facete) and to generate light-weight applications as JavaScript snippets (via Mappify) for specific geospatial applications. Most steps in the Linked Data lifecycle [1] have been integrated in the Generator as a graph-based workflow, which allows the user to easily manage new generated data. The components

comprising it are available in the Linked Data Stack (http://stack.linked-data.org)

Achievements and Future Work

Geoknow is concluding its first year and has already achieved important advancements. The first step was to perform a thorough evaluation of the current standards and technologies for managing geospatial RDF data and identify major shortcomings and challenges [3]. The next step has already produced significant output in the form of ready-for-use tools comprising the GeoKnow Generator. These components are being further enhanced and enriched. For example, Virtuoso RDF store is being extended in order to fully support OGC geometries and the GeoSPARQL standard and FAGI is being developed to support fusion of thematic and geospatial metadata of resources, either manually or automatically. Also, within 2014 the consortium will start testing the use cases and evaluating the performance and scalability of the GeoKnow Generator. Finally, future activities include, among others, the enhancement of the already developed frameworks, as well as the development of sophisticated tools for (a) aggregation of crowdsourced geospatial information and (b) exploration and visualization of spatio-temporal RDF data.

Acknowledgement

Links:
GeoKnow project: http://geoknow.eu
LOD2 project: http://lod2.eu
Linked Data Stack: http://stack.linkeddata.org
GeoKnow Github: https://github.com/GeoKnow
Open Geospatial Consortium: http://www.opengeospatial.org/

References:
[1] S. Auer, J. Lehmann: "Making the web a data washing machine - creating knowledge out of interlinked data", Semantic Web Journal, volume 1, number 12, p. 97-104, IOS Press, 2010, http://www.semantic-web-journal.net/sites/default/files/swj24_0.pdf
[2] M. Perry, J. Herring (eds): "OGC GeoSPARQL standard - A geographic query language for RDF data", Open Geospatial Consortium Inc, v.1.0, 27/04/2012, https://portal.opengeospatial.org/files/?artifact_id=47664
[3] K. Patroumpas et al.: "Market and Research Overview", GeoKnow EU/FP7 project deliverable 2.1.1, 2013, http://svn.aksw.org/projects/GeoKnow/Public/D2.1.1_Market_and_Research_Overview.pdf
[4] S. Auer et al.: "Managing the lifecycleof Linked Data with the LOD2 Stack", in proc. of ISWC'11, Springer, 2012
[5] A. G. Rojas, et al.: "GeoKnow: Leveraging Geospatial Data in the Web of Data", in Open Data Workshop, W3C, London, 2013.

Please contact:
Spiros Athanasiou
Institute for the Management of Information Systems
Athena Research Center, Greece
E-mail: spathan@imis.athena-innovation.gr

Daniel Hladky
Ontos AG, Switzerland
E-mail: daniel.hladky@ontos.com

Jens Lehmann
InfAI, University of Leipzig
Email: lehmann@informatik.uni-leipzig.de

Giorgos Giannopoulos
Institute for the Management of Information Systems
Athena Research Center, Greece
E-mail: giann@imis.athena-innovation.gr