

Geospatial Data and the Semantic Web

The GeoKnow Project

Sebastian Hellmann

AKSW/KILT research group, Leipzig University & DBpedia Association

About the project

Collaborative Project 2012-2015

Information and Communication Technologies

Project No. 318159 Start Date 01/12/2012



The story so far

- 1 Extension of the Web with a **data commons** (> 50 Billion facts) including \approx 20% geospatial data sets
- 2 Emerging **spatial data web research community**
- 3 **Standardisation activities** (e.g. GeoSPARQL, W3C & OGC Standardisation Group started Jan 2015)
- 4 More **Spatial Triple Stores** (e.g. Strabon)
- 5 Linked Data allows combining geo services and formats (e.g. INSPIRE, OGC) with other structured information





Information Excellence

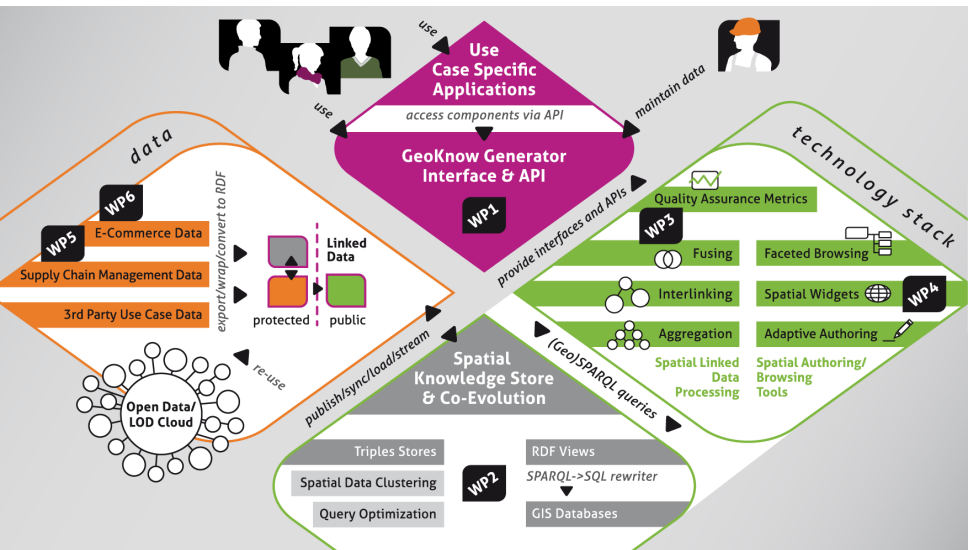


INEY IMIS



InfAI®
Institut für Angewandte Informatik

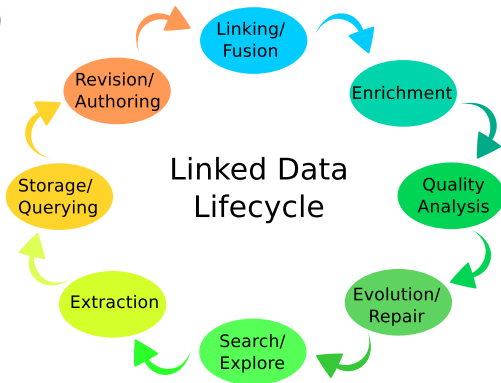





WP7: Dissemination & Exploitation

WP8: Management

- Originated in LOD2 Project
- **Linked Data Stack** (<http://stack.linkeddata.org>) provides a consolidated repository of such tools
- Lightweight integration between tools via **common vocabularies and standards**, e.g. (Geo)SPARQL
- Each tool is a **Debian** package
 - 38 GeoKnow packages
 - 87 packages in Linked Data Stack (including other projects)
 - RPM packages
 - Nightly builds
 - Planned stack use in Big Data Europe, HOBBIT, GEISER etc. research projects



- Common interface and integration of geospatial Linked Data Stack tools
 - Guided demo tour on the website
 - Co-evolution support
 - Supply chain data extraction
 - Commercial adoption at State Secretariat for Economic Affairs (SECO) Switzerland within LINDAS project



⚙️ Workbench
✎ Settings
👤 Account

Logged in as **badmin**
🚪 Logout

Data Sources

Named Graphs

Components

Roles Management

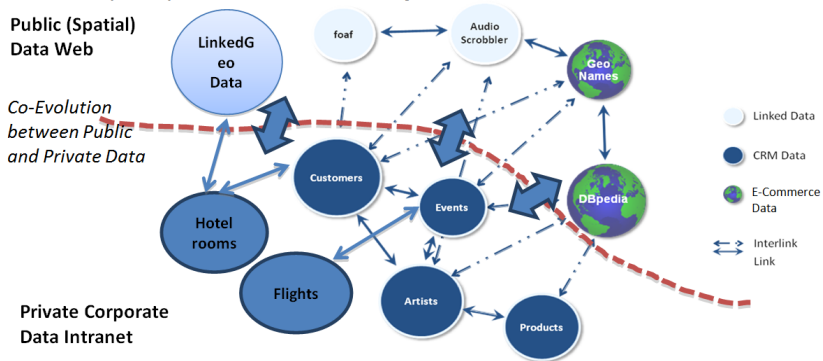
Generator Settings

| | |
|-----------------|---|
| URI Base: | http://www.xybermotive.com/resource/ |
| Settings Graph: | http://www.xybermotive.com/resource/settingsGraph |
| Endpoint: | http://generator.geoknow.eu:8890/sparql-auth |
| Workbench URL: | http://generator.geoknow.eu:8080/brox-generator/ |

Public-private spatial data co-evolution

Goal: Identify types of enterprise RDF data synchronization workflows, define and implement tools that support them. This will include ETL processes, query federation, transformation/patching of data and change set propagation.

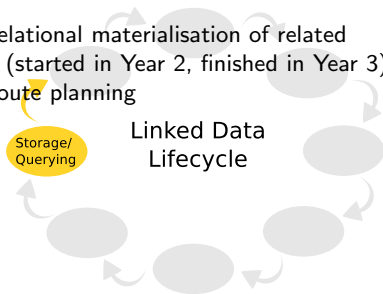
Achieved by: Named Graph Sets with changeset tracking, versioning and conflict resolution



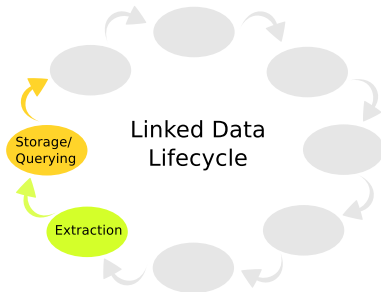
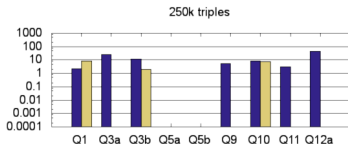
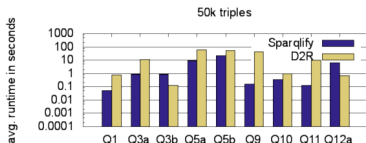
- Intensive work on benchmarking geospatial systems
- Benchmark project started: <https://github.com/GeoKnow/GeoBenchLab>
- Validation of component scalability across different lifecycle phases (and detection of potential regression)
- Current benchmarks:
 - Slippy map benchmark
 - Datacube query benchmark
 - Interlinking benchmark
 - Fusion benchmark
 - Enrichment benchmark
 - Facet Browsing Library benchmark
 - RDB2RDF benchmark



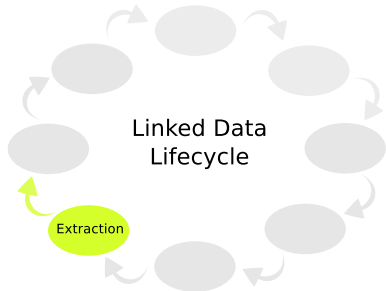
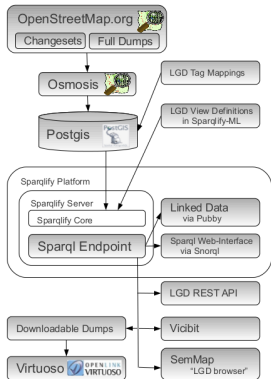
- Triplestore development in GeoKnow
 - Initial analysis for Triple Stores performed (Virtuoso, uSeekM, Parliament, AllegroGraph, OWLIM-SE, Strabon + Oracle Spatial 11g, PostGIS as reference) using fragments of OSM and Ordnance Survey data
 - Improved GeoSPARQL compliance (after tests with various stores: uSeekM, Parliament, AllegroGraph, OWLIM-SE, Strabon)
 - Continuous benchmarking of Virtuoso throughout the project (detailed in D1.3 and WP2)
 - Characteristic set support in Virtuoso - relational materialisation of related structures to improve query performance (started in Year 2, finished in Year 3)
 - Graph analytics inside Virtuoso e.g. for route planning



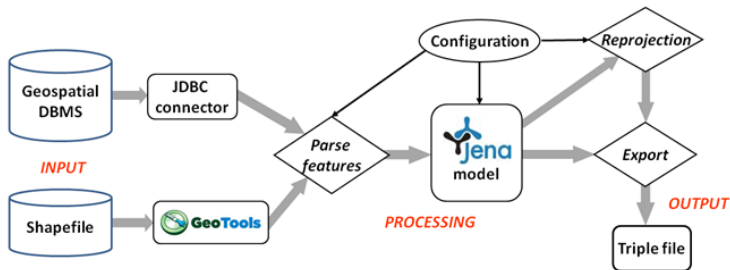
- <http://aksw.org/Projects/Sparqlify>
- Rewrites SPARQL SELECT queries to a **single** SQL query allowing the underlying database to perform optimizations
- Support for using **geospatial RDBMS (PostGIS) functions**
- Web interface with syntax highlighting and live data generation for easy mapping creation
- Docker image now available



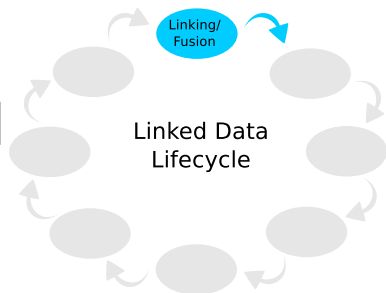
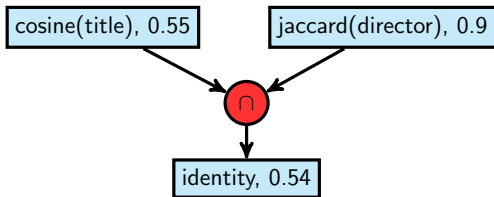
- LinkedGeoData conversion simplified to a set of SQL files and Sparqlify Mapping Definitions
- New full release (January 2016) with more than 1.2 billion triples based on the OpenStreetMap planet file from 2015-11-02 is now online!
- Updated/new interlinks with Wikidata, DBpedia, Wikimapia, Geonames, Natural Earth in progress



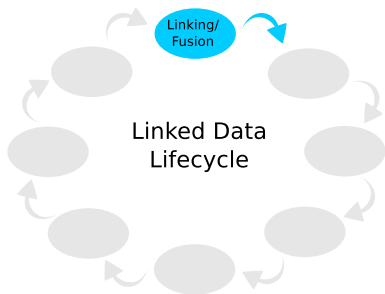
- Converts shapefiles, spatial DBMS output to RDF
- Input: ESRI shapefiles, Oracle Spatial, PostGIS, MySQL, IBM DB2, INSPIRE-aligned GML (Geography Markup Language) data for seven Data Themes (in Annex I)
- Output: GeoSPARQL WKT, WGS84 + several RDF serialisations supported
- Support for EU INSPIRE directive added, in particular INSPIRE-aligned GML (Geography Markup Language) data for seven Data Themes (Annex I): Addresses, Administrative Units, Cadastral Parcels, GeographicalNames, Hydrography, Protected Sites, and Transport Networks (Roads)



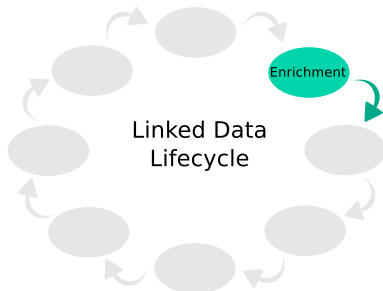
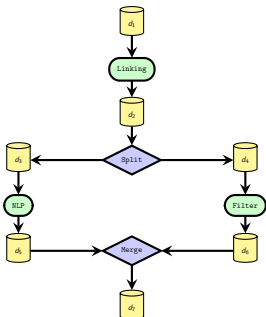
- Orthodromic distance only supported by LIMES and SILK
- Linking algorithm ORCHID for geospatial data in LIMES
- Reduces number of comparisons of resources in datasets which should be integrated
- Linking papers at ESWC'13 (linking in the cloud, best paper award) and ISWC'13 (ORCHID algorithm) and ISWC'14 ORCHID
 - Between 1 and 2 orders of magnitude scalability improvement on geospatial data
 - Large-scale evaluation of ORCHID including industrial settings
 - Improved caching and load balancing



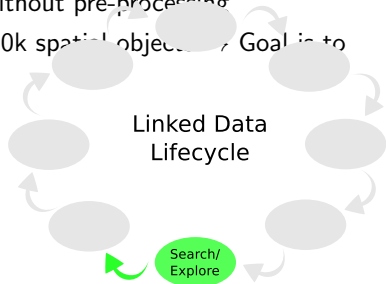
- Geospatial transformations and fusion
- Merges geospatial data using several configurable fusing actions
 - Map interactive geospatial transformations
 - Higher efficiency
 - 2 orders of magnitude performance improvements
 - Extended set of fusion actions



- Goal: enrich RDF datasets with (spatial) information
- Modular pipeline approach for dereferencing, linking and NLP extraction
- Progress:
 - Machine learning techniques for learning DEER specifications using desired input-output pairs
 - Learning of more complex ML specs work in progress



- Generic faceted browser for HTTP SPARQL endpoints
- Vocabulary abstraction layer for support of arbitrary geo vocabularies (e.g. GeoSPARQL, WGS)
- Automatically finds map data for entities
- Nested facets and client side pagination without pre-processing
- Client side querying with support for $\approx 100k$ spatial objects. Goal is to work with DBpedia/LGD



- RDF authoring support
- Implicit support for tracking changes
- Support for explicitly registering Changesets using the Co-Evolution API

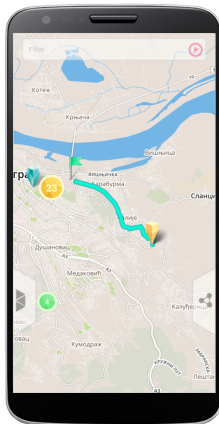
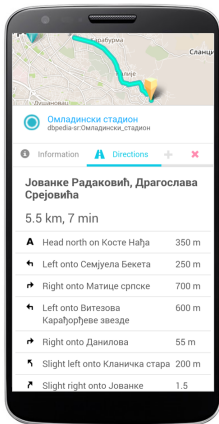
The screenshot shows the Facete II application interface. On the left, there are navigation and search controls. The main area displays a map of Europe with numerous red location markers. Below the map, a data table lists funding entries.

Facete II Find Place... Submit Main Edit Tools Config Notifications

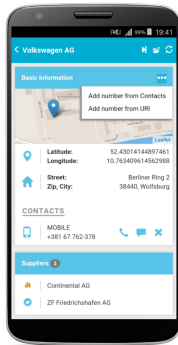
Data

| Items | fundings/partner/address/country | fundings/amount | year |
|--------------|----------------------------------|-----------------|------|
| ECHORD | Germany | 7033790 | 2009 |
| HAVE-it | Germany | 4422353 | 2007 |
| PIANO+ | Germany | 3918750 | 2010 |
| FI-WARE | Germany | 3509783 | 2011 |
| OLED100.eu | Germany | 3307528 | 2008 |
| SMErobotics | Germany | 2709866 | 2011 |
| SEAL | Germany | 2444184 | 2010 |
| SLA@SOI | Germany | 2370399 | 2008 |
| PLANTCockpit | Germany | 2310720 | 2010 |
| BrainScaleS | Germany | 2123580 | 2010 |

- Mobile spatial exploration - extended with better routing, authoring, recommendations



- Mobile dashboard developed:



- Integration of prototype with Xybermotive commercial Web-EDI system
- On-the-fly calculation of metrics, for example weather influences and news aggregation (see demo)
- Distributed metrics calculation via Apache Spark

- Motive based search prototype
- Internal evaluation and crowdsourcing approach on BlueKiwi web application

The screenshot shows a mobile application interface for a hotel listing. At the top, there is an orange header with the text "TOP Panorama Inn & Boardinghaus" and "★★★ Hamburg, Hamburg". Below the header, there are three main sections: a photo of a hotel room, a rating of "3,8 - Sehr gut Bewertung", and a price tag "ab 257€" with a "Flug + Hotel" icon. Below these, there are sections for "Beschreibung", "Gästemeinungen", "Einrichtungen", and "Themen". The "Gästemeinungen" section is expanded, showing positive and negative reviews. The "Angebote zum Hotel" section at the bottom allows users to search for offers, with fields for dates (25.11.15 - 09.12.15) and duration (1 Woche), and a "Personen: 2" indicator.

TOP Panorama Inn & Boardinghaus
★★★ Hamburg, Hamburg

3,8 - Sehr gut Bewertung
ab 257€
Flug + Hotel

Beschreibung
Gästemeinungen
Einrichtungen
Themen

Positives
Hotel gutes Hotel
sauberes Hotel
Zimmer sauberes Zimmer
saubere Zimmer
mehr ...

Negatives
Hotel kostenpflichtige Garage
begrenzte Auswahl
Zimmer kleiner Fernseher
sehr kleiner Fernseher

Angebote zum Hotel Pauschalreisen ✕ + nur Hotel

25.11.15 - 09.12.15 1 Woche
Früheste Hinreise- und spätestes Rückreisedatum Reisezeitraum

Personen: 2 Angebote finden

main repository: <https://github.com/GeoKnow>

Software components:

- DataCubeValidation - quality assessment for statistical data
- DEER – annotates existing RDF datasets with geospatial information
- ESTA-LD - spatiotemporal analysis tool
- FAGI – a tool for fusion and aggregation of geospatial information
- GEM - mobile geospatial exploration and filtering
- Jassa – a JavaScript library for SPARQL-based applications
- LDViewer - Linked Data presentation and browsing
- Lodtenant - RDF/SPARQL workflow toolkit
- Mappify – allows to create simple geospatial web applications
- TripleGeo – converts shapefiles and other geospatial structures to triples
- Sparqlify – SPARQL-2-SQL rewriter
- LinkedGeoData – RDF version of OpenStreetMap
- Facete (2) – a generic, facet-based RDF browser
- LIMES – framework for RDF data integration
- Virtuoso – triple store and middleware



Scientific Project Leader

Jens Lehmann

<http://jens-lehmann.org>

lehmann@informatik.uni-leipzig.de

Institute for Applied Informatics (InfAI)



Project Manager

Nadine Jänicke

<http://aksw.org/NadineJaenicke.html>

jaenicke@uni-leipzig.de

Institute for Applied Informatics (InfAI)



<http://geoknow.eu>

Thanks for your attention!