



InGeoCloudS
Inspired GEOdata CLOUD Services

Towards Linked Data

Yannis Roussakis, Athina Kritsotaki, Martin Doerr, Gerald Hiebel, Kyriakos Kritikos, Vassilis Christophides, Dimitris Kotzinos

FORTH-ICS



Outline



- Introduction
- Transformation into Linked Data
- Linked Data Management
- Linked Data export into INSPIRE

Introduction (I)



- In INGEOCLOUDS project we deal with scientific data described in different formats from diverse scientific areas
 - ground waters, boreholes, chemical analyses, etc.
 - earthquakes, recordings, stations, etc.
 - Landslides, susceptibility maps, etc.
- Purpose: integrate, describe and query such heterogeneous data in a uniform way

Introduction (II)



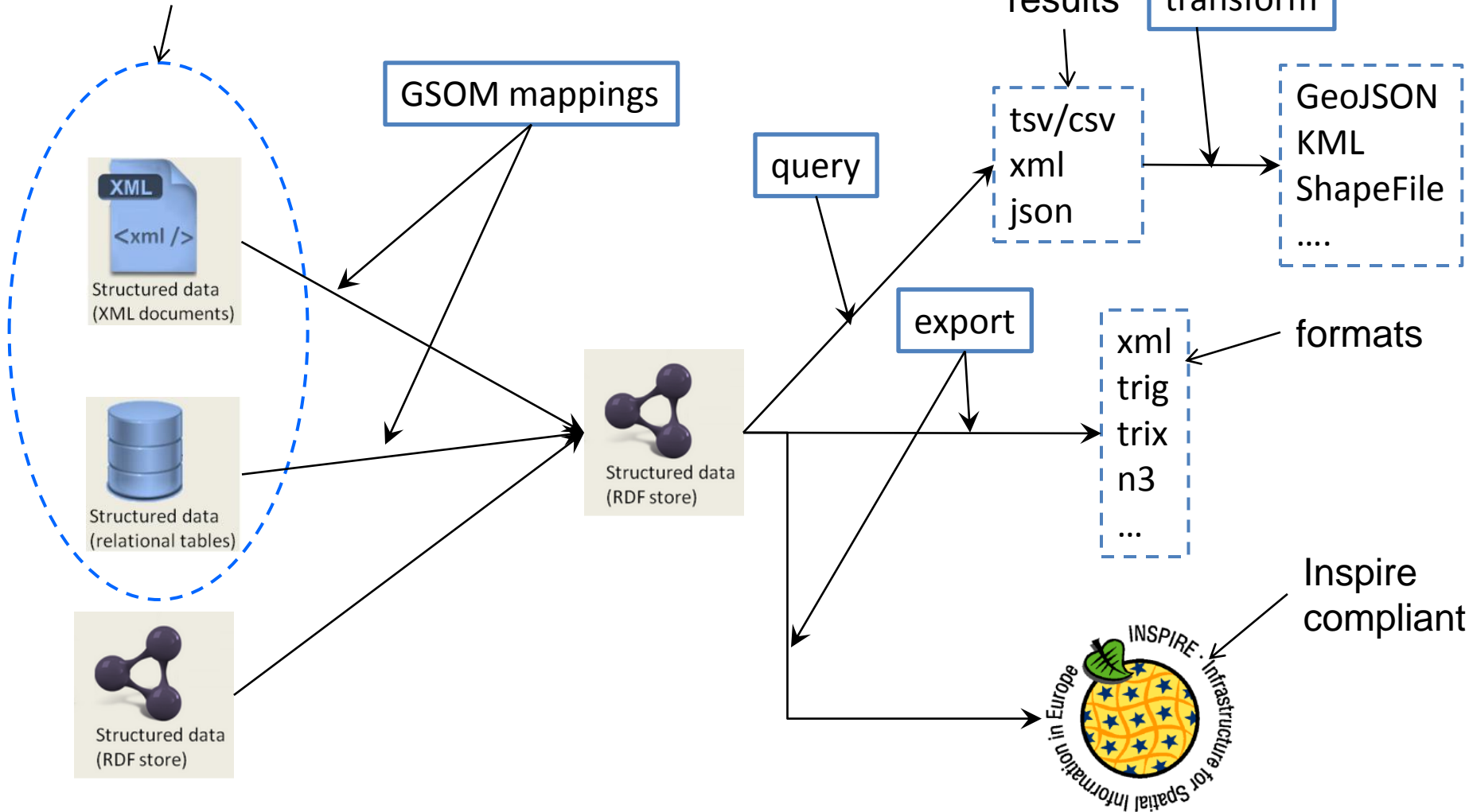
- Approach
 - Creation of a Conceptual Model to integrate and cover all the thematic fields
 - Data providers mapped the source data into RDF data compliant to the Conceptual Model
 - Rely on a scalable RDF Triple Store to enforce the mappings and enable the storage and query of the RDF data
 - Provide an API for the RDF data management



Overview



Providers' Data



R2RML-based Linked Data Transformation Scenario (I)



- We will show how the previously defined steps (for R2RML-based mappings) can be followed to create and upload the corresponding Linked Data for a specific data provider – IGME
- IGME data
 - Original data are stored in relational tables
 - Ground waters sample takings taken from the regions of Mygdonia and Thriasio in Greece
 - Chemical analyses results over sample takings
 - Information about the sampling date and origin (e.g., boreholes, springs, etc)

R2RML-based Linked Data Transformation Scenario (II)



Chem. Analysis

| |
|-------|
| CODE |
| SDATE |
| pH |
| COND |
| Ca |
| Mg |
| Na |
| ... |
| |
| V |
| Be |
| ... |

Borehole code

Sampling date

**Chemical substances
(mgr/lit or µgr/lit)**

Borehole

| |
|--------------|
| <u>CODE</u> |
| <u>PCODE</u> |
| TYPE |
| LATITUDE |
| LONGITUDE |
| CODE_WDIST |
| NAME_WDIST |
| CODE_AQUIF |
| NAME_AQUIF |
| DEPTH |
| DRILL_DIAM |
| GEOLOGY |
| |

Borehole code

Borehole name

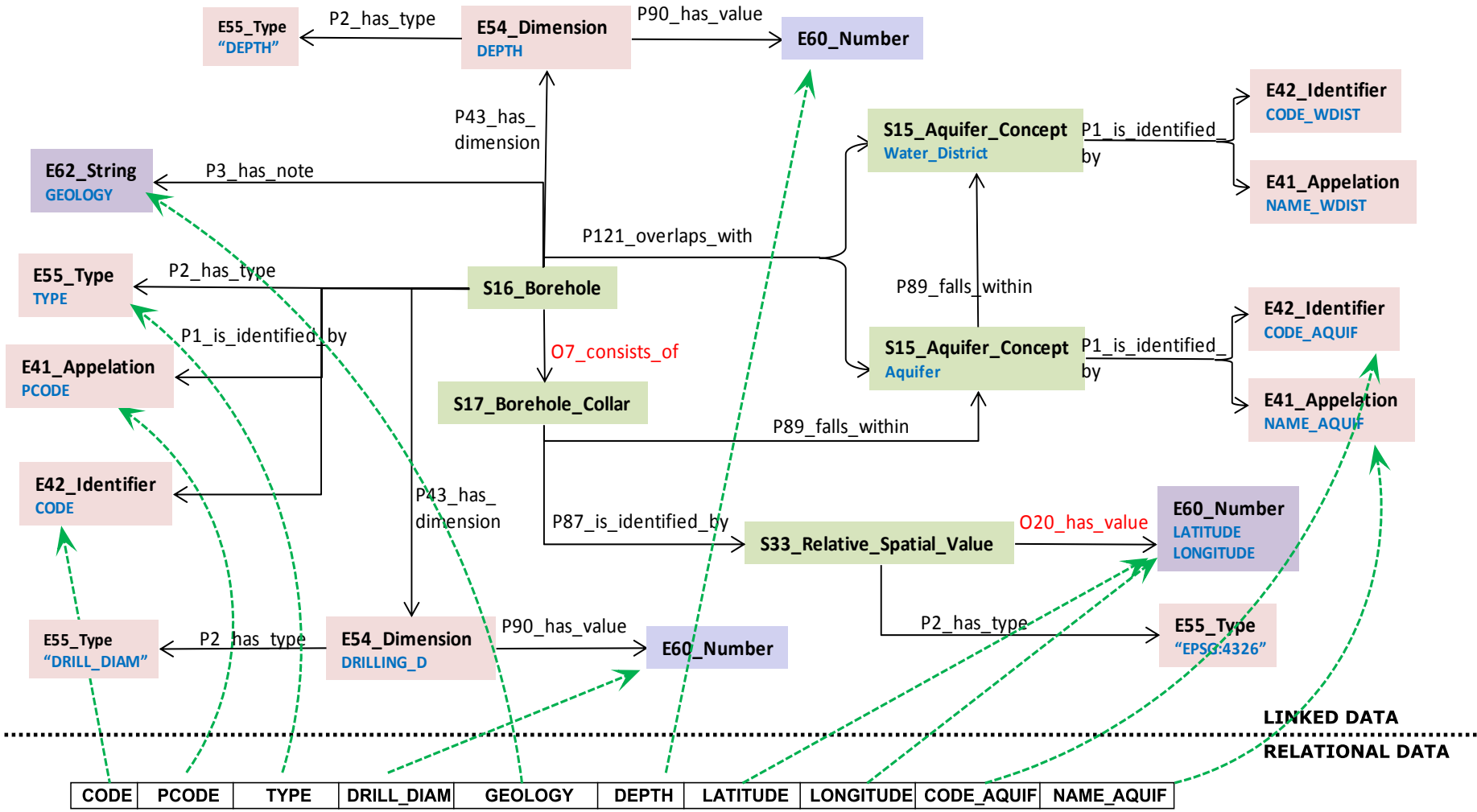
Sampling source

location

Water district

aquifer

Soil composi.



R2RML-based Linked Data Transformation Scenario (III)



- The informal Relational-to-Linked data mappings have to be formalized using R2RML language
- The final step is to call a specific API method (***addR2RMLMappings***), to enforce the R2RML mappings and create the corresponding Linked Data
- As a result:
 - the original Relational Data are transformed into Linked Data w.r.t. the R2RML mappings
 - There exists an one-way synchronization from the Relational Data to Linked Data

Querying the Linked Data



- The transformation of providers' data into Linked Data enables the evaluation of uniform queries over the underlying triple store (Virtuoso) which can be of the following types:
 - Provider-specific queries
 - Cross-theme or cross-provider queries
- Queries are expressed via the W3C's SPARQL language (<http://www.w3.org/TR/rdf-sparql-query/>)

Querying the Linked Data – Provider-Specific Scenario



- **Query Description:** Select the samples of the compounds with $\text{ph} > 7$ and $\text{Na} > 50$. Moreover select the locations and the boreholes(names) which were analyzed.
- This query is also applied only over IGME data (provider specific)
- Query is first expressed in SPARQL (see next slide) and then *ldquery* API method is called



| sample | compound | cavalu | bname | asWKT |
|---|----------|--------|---|--|
| http://www.igme.gr/GrWater/655 | pH | 7.8 | http://www.igme.gr/Borehole/Name/A%2F113 | < http://www.opengis.net/def/crs/EPSSG/0/4326 > Point(23.58046 38.05195) |
| http://www.igme.gr/GrWater/700 | pH | 8 | http://www.igme.gr/Borehole/Name/A%2F113 | < http://www.opengis.net/def/crs/EPSSG/0/4326 > Point(23.58046 38.05195) |
| http://www.igme.gr/GrWater/700 | Na | 285.1 | http://www.igme.gr/Borehole/Name/A%2F113 | < http://www.opengis.net/def/crs/EPSSG/0/4326 > Point(23.58046 38.05195) |
| http://www.igme.gr/GrWater/701 | Na | 850.6 | http://www.igme.gr/Borehole/Name/A%2F113 | < http://www.opengis.net/def/crs/EPSSG/0/4326 > Point(23.58046 38.05195) |
| http://www.igme.gr/GrWater/702 | pH | 7.2 | http://www.igme.gr/Borehole/Name/A%2F113 | < http://www.opengis.net/def/crs/EPSSG/0/4326 > Point(23.58046 38.05195) |
| http://www.igme.gr/GrWater/702 | Na | 494.3 | http://www.igme.gr/Borehole/Name/A%2F113 | < http://www.opengis.net/def/crs/EPSSG/0/4326 > Point(23.58046 38.05195) |
| http://www.igme.gr/GrWater/703 | pH | 7.4 | http://www.igme.gr/Borehole/Name/A%2F113 | < http://www.opengis.net/def/crs/EPSSG/0/4326 > Point(23.58046 38.05195) |
| http://www.igme.gr/GrWater/703 | Na | 864.4 | http://www.igme.gr/Borehole/Name/A%2F113 | < http://www.opengis.net/def/crs/EPSSG/0/4326 > Point(23.58046 38.05195) |
| http://www.igme.gr/GrWater/704 | pH | 7.6 | http://www.igme.gr/Borehole/Name/A%2F113 | < http://www.opengis.net/def/crs/EPSSG/0/4326 > Point(23.58046 38.05195) |
| http://www.igme.gr/GrWater/704 | Na | 873.6 | http://www.igme.gr/Borehole/Name/A%2F113 | < http://www.opengis.net/def/crs/EPSSG/0/4326 > Point(23.58046 38.05195) |
| http://www.igme.gr/GrWater/705 | pH | 7.5 | http://www.igme.gr/Borehole/Name/A%2F113 | < http://www.opengis.net/def/crs/EPSSG/0/4326 > Point(23.58046 38.05195) |
| http://www.igme.gr/GrWater/705 | Na | 901.2 | http://www.igme.gr/Borehole/Name/A%2F113 | < http://www.opengis.net/def/crs/EPSSG/0/4326 > Point(23.58046 38.05195) |
| http://www.igme.gr/GrWater/706 | pH | 7.8 | http://www.igme.gr/Borehole/Name/A%2F113 | < http://www.opengis.net/def/crs/EPSSG/0/4326 > Point(23.58046 38.05195) |
| http://www.igme.gr/GrWater/706 | Na | 809.2 | http://www.igme.gr/Borehole/Name/A%2F113 | < http://www.opengis.net/def/crs/EPSSG/0/4326 > Point(23.58046 38.05195) |
| http://www.igme.gr/GrWater/707 | pH | 7.5 | http://www.igme.gr/Borehole/Name/A%2F113 | < http://www.opengis.net/def/crs/EPSSG/0/4326 > Point(23.58046 38.05195) |

Querying the Linked Data – Cross-Provider Scenario



- The previous query can be applied to additional data providers with similar data (i.e. BRGM, GEUS) leading to cross datasets queries
- **Query Description:** Select the of the compounds with with $\text{ph} > 7$ and $\text{Na} > 100$. Moreover select the locations and the boreholes(names) which were analyzed.
- This query is applied over all partners' datasets which are relevant
 - IGME, GEUS, BRGM



| sample | compound | cavalue | bname | asWKT |
|---|----------|---------|---|--|
| http://www.igme.gr/GrWater/704 | Na | 873.6 | http://www.igme.gr/Borehole/Name/A%2F113 | < http://www.opengis.net/def/crs/EPSSG/0/4326 > Point(23.58046 38.05195) |
| http://www.igme.gr/GrWater/705 | pH | 7.5 | http://www.igme.gr/Borehole/Name/A%2F113 | < http://www.opengis.net/def/crs/EPSSG/0/4326 > Point(23.58046 38.05195) |
| http://www.igme.gr/GrWater/705 | Na | 901.2 | http://www.igme.gr/Borehole/Name/A%2F113 | < http://www.opengis.net/def/crs/EPSSG/0/4326 > Point(23.58046 38.05195) |
| http://www.igme.gr/GrWater/706 | pH | 7.8 | http://www.igme.gr/Borehole/Name/A%2F113 | < http://www.opengis.net/def/crs/EPSSG/0/4326 > Point(23.58046 38.05195) |
| http://www.igme.gr/GrWater/706 | Na | 809.2 | http://www.igme.gr/Borehole/Name/A%2F113 | < http://www.opengis.net/def/crs/EPSSG/0/4326 > Point(23.58046 38.05195) |
| http://www.igme.gr/GrWater/707 | pH | 7.5 | http://www.igme.gr/Borehole/Name/A%2F113 | < http://www.opengis.net/def/crs/EPSSG/0/4326 > Point(23.58046 38.05195) |
| http://www.igme.gr/GrWater/707 | Na | 943 | http://www.igme.gr/Borehole/Name/A%2F113 | < http://www.opengis.net/def/crs/EPSSG/0/4326 > Point(23.58046 38.05195) |
| http://services.ades.eaufrance.fr/GrWater/4036641 | pH | 7.1 | http://services.ades.eaufrance.fr/Borehole/Name/FORAGE%20DU%20CUF%20DE%20BCEUF | < http://www.opengis.net/def/crs/EPSSG/0/4326 > Point(5.08089810166797 45.5181359777231) |
| http://services.ades.eaufrance.fr/GrWater/4037516 | pH | 7.2 | http://services.ades.eaufrance.fr/Borehole/Name/FORAGE%20DU%20CUF%20DE%20BCEUF | < http://www.opengis.net/def/crs/EPSSG/0/4326 > Point(5.08089810166797 45.5181359777231) |
| http://services.ades.eaufrance.fr/GrWater/4038347 | pH | 7.1 | http://services.ades.eaufrance.fr/Borehole/Name/FORAGE%20DU%20CUF%20DE%20BCEUF | < http://www.opengis.net/def/crs/EPSSG/0/4326 > Point(5.08089810166797 45.5181359777231) |
| http://services.ades.eaufrance.fr/GrWater/4039197 | pH | 7.3 | http://services.ades.eaufrance.fr/Borehole/Name/FORAGE%20DU%20CUF%20DE%20BCEUF | < http://www.opengis.net/def/crs/EPSSG/0/4326 > Point(5.08089810166797 45.5181359777231) |
| http://services.ades.eaufrance.fr/GrWater/4039726 | pH | 7.1 | http://services.ades.eaufrance.fr/Borehole/Name/FORAGE%20DU%20CUF%20DE%20BCEUF | < http://www.opengis.net/def/crs/EPSSG/0/4326 > Point(5.08089810166797 45.5181359777231) |

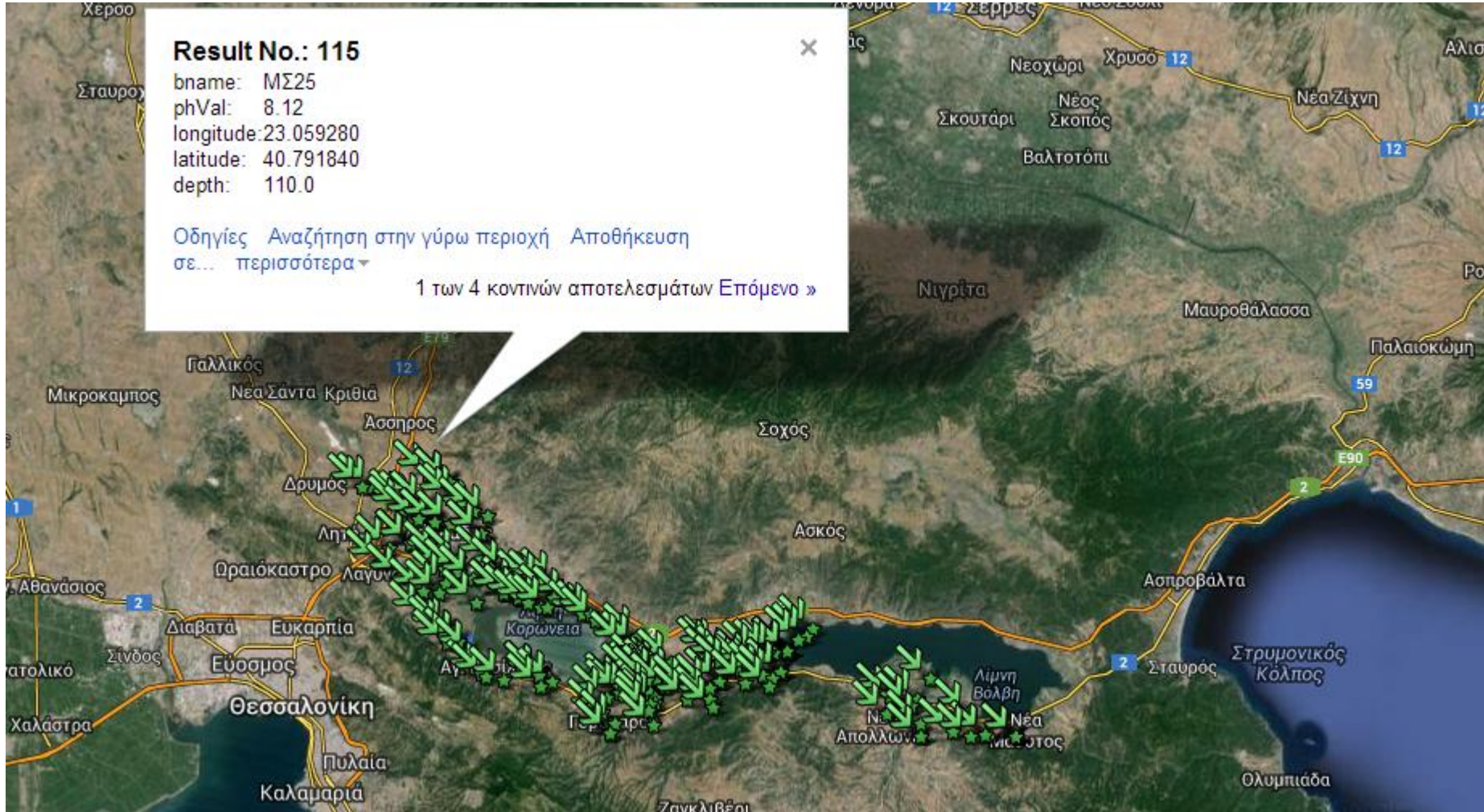
Results Visualization (I)



- As the results are provided in a particular machine-processable form, user can request a better visualization by transforming them to a format which can be loaded on various GUI tools/services, such as Google Maps
- To this end, if the SPARQL results contain spatial information, the API offers the ***geoldtransform*** method through which they can be transformed into geographical feature collection representations, such as KML, GeoJSON, GML and Shape.

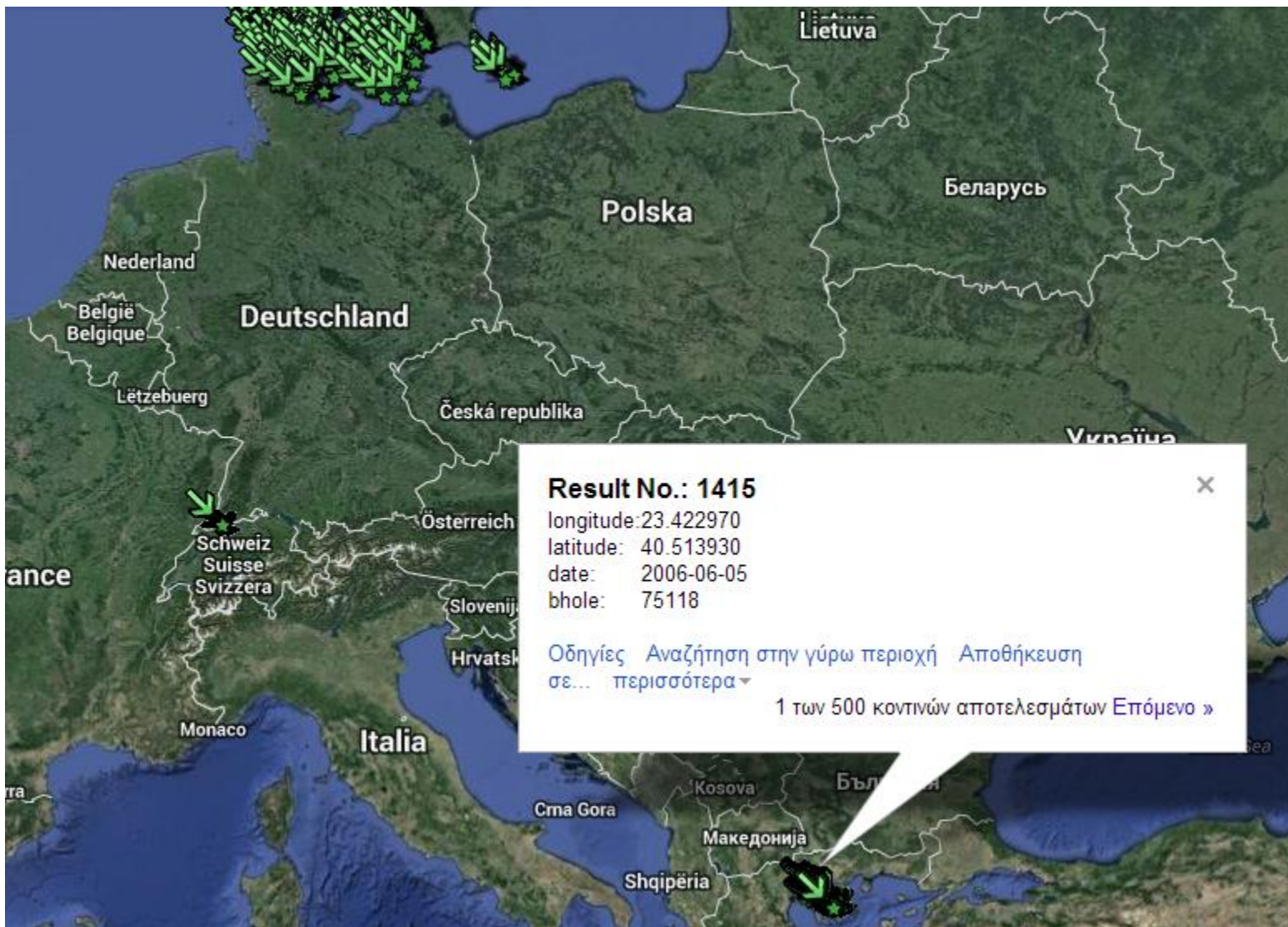


Results Visualization (II)





Results Visualization (III)



Export Linked Data into INSPIRE



- Linked Data in GSOM are not INSPIRE-compliant
- We created a GSOM-to-INSPIRE mapping mechanism which allows to transform GSOM-based RDF data, acquired from SPARQL queries, into XML-based INSPIRE-compliant data
- However INSPIRE has limitations which do not allow for a complete export of all the GSOM notions (i.e., we cannot express all the GSOM notions in INSPIRE notions)
- The generic INSPIRE conceptual schema is **not event-oriented** (especially the schema that uses the observation-measurement specification)

Approach to INSPIRE-based Exporting (I)



- Our approach was to separate our data (from all data providers) into themes.
 - chemical analyses, boreholes, landslides, earthquakes etc.
- Next we investigated the GSOM notions which could be mapped to inspire notions as only these notions will be eventually exported
- Finally, we created a set of API methods which can be used by users to export Linked Data into an INSPIRE compliant format according to two main alternative ways

Approach to INSPIRE-based Exporting (II)



- First export way:
 - The user can select one or more themes which interest him in order to export their respective data
 - Next, the user may choose to export all the data which correspond to the selected themes or he can apply SPARQL filter conditions for a selective export
 - Finally, the service ***inspire_export*** can be called with the above parameters and the INSPIRE compliant data are created

Approach to INSPIRE-based Exporting (III)



- Second export way:
 - Instead of selecting specific themes the user can provide the appropriate SPARQL query related to the data he/she desires to obtain to the *inspire_query_export* API method
 - This method, by knowing all GSOM-to-INSPIRE mappings, maps the SPARQL query variables to INSPIRE-mapped GSOM notions and makes the corresponding result transformation to INSPIRE
 - This method is appropriate for users which are familiar with GSOM and of course with SPARQL

Approach to INSPIRE-based Exporting (IV)



- Second export way:
 - Can also be used to obtain related results which span different themes and different data providers – by allowing the user to provide a SPARQL query we enable the description of any data requirement
 - Due to INSPIRE incompleteness issues raised previously, not all SPARQL query variables are guaranteed to be mapped into INSPIRE-compliant information. On the contrary, the first INSPIRE export method provides a more controlled way to produce the required INSPIRE-compliant information (all SPARQL variables are mapped to INSPIRE information)

INSPIRE Exporting Scenario (I)



- Recall that IGME data refer on Boreholes, Samples and Chemical Analyses
- The relevant themes which can be exported are those of: *boreholes, chemical analyses*
- We will show next how we can use the methods of the Linked Data Management API to export the Linked Data in INSPIRE-compliant form
- The exporting scenario considers all theme-related data even from other data providers (the latter data can be filtered by providing filtering constraints)

INSPIRE Exporting Scenario (II)



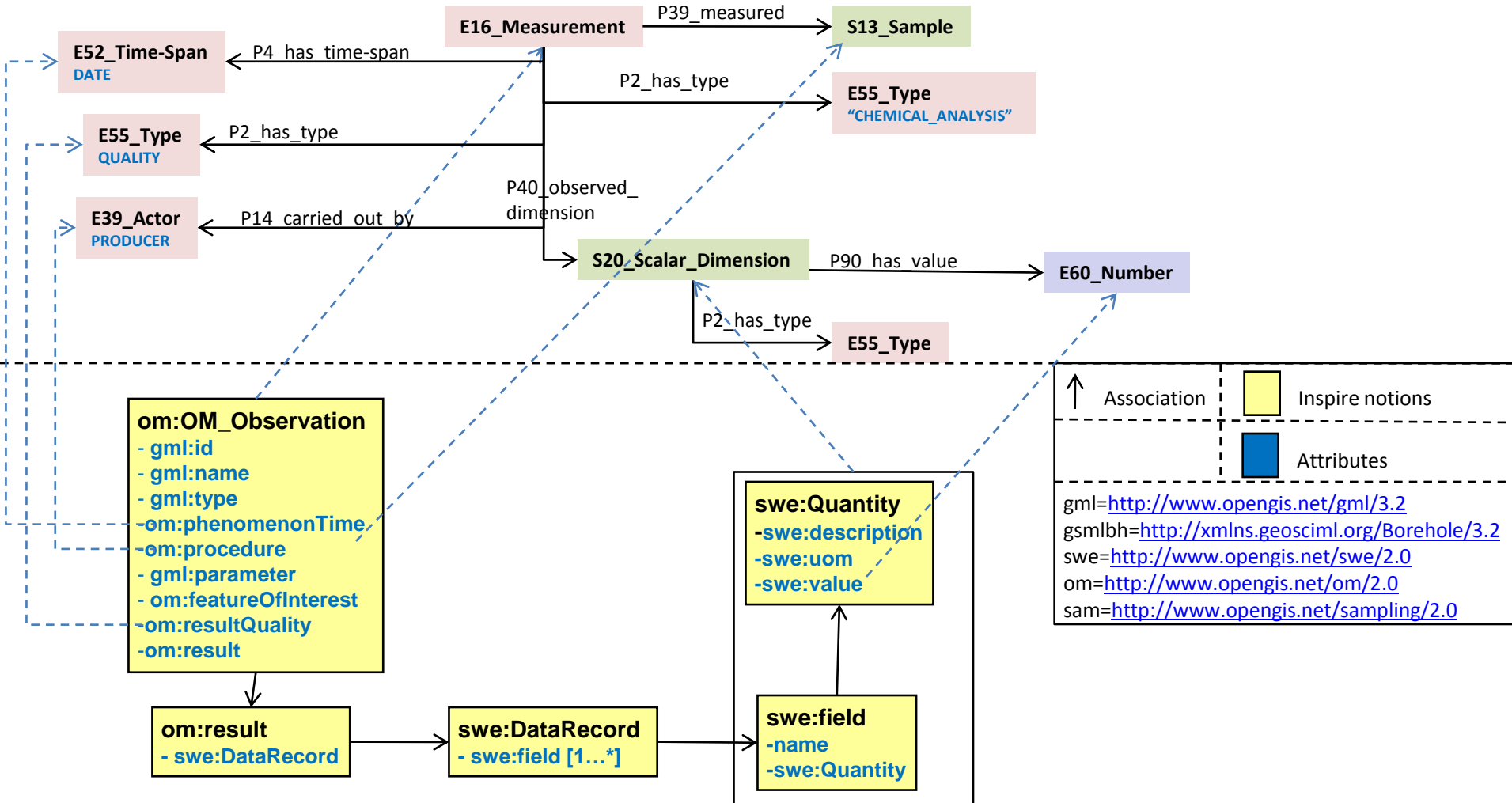
- **Scenario:** Recall the provider specific query shown in the previous section transformed into a cross datasets query
 - Select the samples of the compounds with $\text{ph} > 7$ and $\text{Na} > 100$. Moreover select the locations and the boreholes(names) which were analyzed.
- Export INSPIRE data from chemical analyses w.r.t. the above condition parameters
 - **Case 1:** Call first INSPIRE export method with parameters:
 - themes: **Chemical Analysis**
 - constraints: (compound = "pH" and value > 7) or (compound = "Na" and value > 200) or compound = "."

INSPIRE Exporting Scenario (III)



- **Case 2:** Call second INSPIRE export method with parameters:
 - query: the respective query that produces the equivalent results
- **Result:** a zipped file containing the xml specification of the chemical data is created w.r.t. the appropriate XSD specifications of INSPIRE

INSPIRE Exporting Scenario (IV)



Conclusions



- The GSOM was designed in order to cover notions and integrate data from all the diverse themes
- The Linked Data Management API developed in INGEOCLOUDS project:
 - Provides high-level functionality hiding the low-level complexities/peculiarities of the underlying RDF Store
 - Enables the complete management of Linked Data (import, update, query, transform)
 - Direct importing of RDF data
 - Indirect importing of XML/relational data
 - Supports the export of data into an INSPIRE-compliant format



InGeoCloudS
Inspired GEOdata CLOUD Services



Thanks for your attention