InGeoCloudS
Inspired GEOdata CLOUD Services

A cloud based platform for Linking and Managing Geodata

Dimitris Kotzinos
University of Cergy Pontoise and FORTH-ICS
Kyriakos Kritikos and Yannis Roussakis
FORTH-ICS

This content by the InGeoCloudS consortium members is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License. Based on a work at http://www.ingeoclouds.eu/.

(cc) BY-NC-SA

cip competitiveness and innovation framework programme 2007–2013

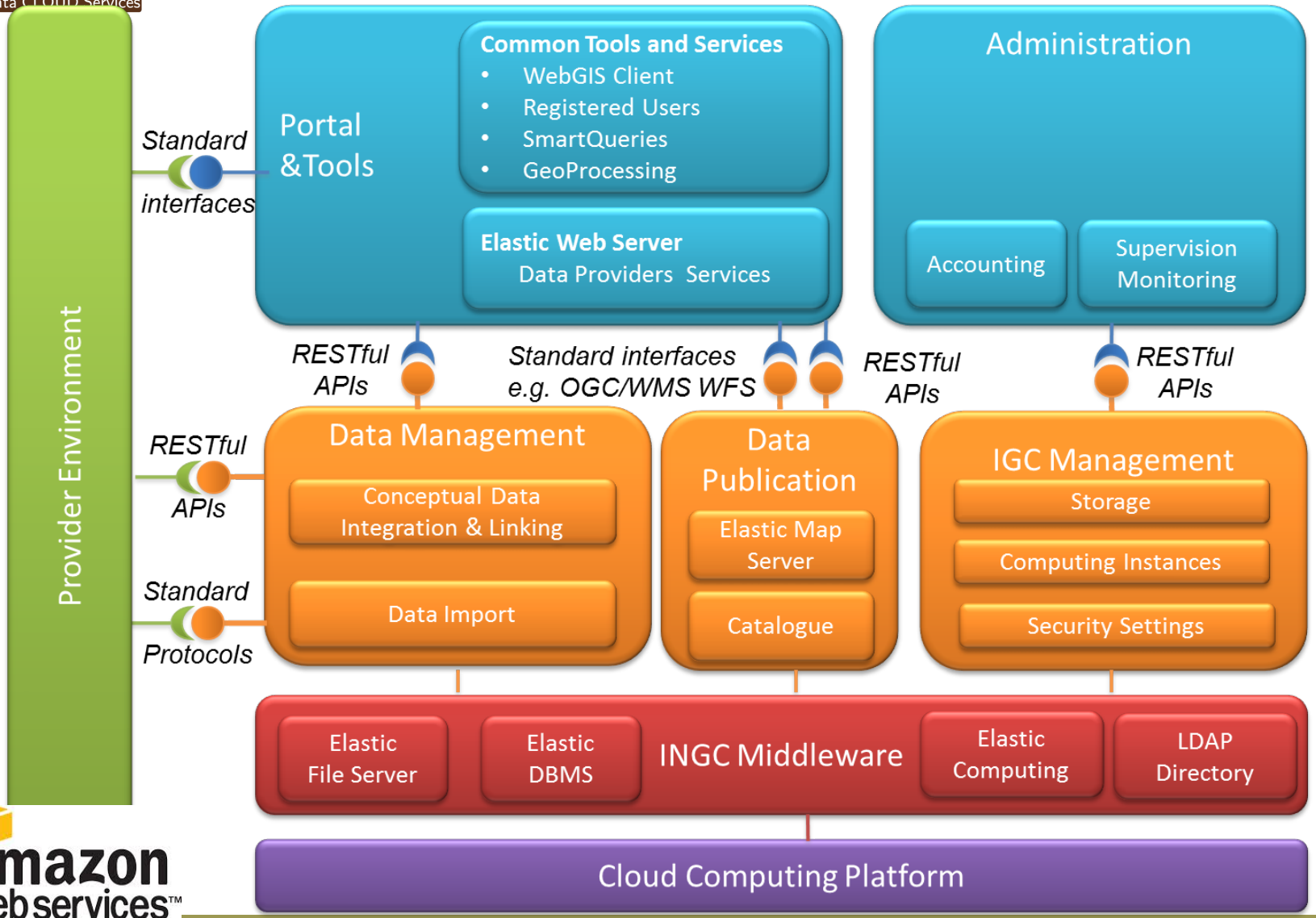
InGeoCloudS
Inspired GEOdata CLOUD Services

Objectives in a nutshell

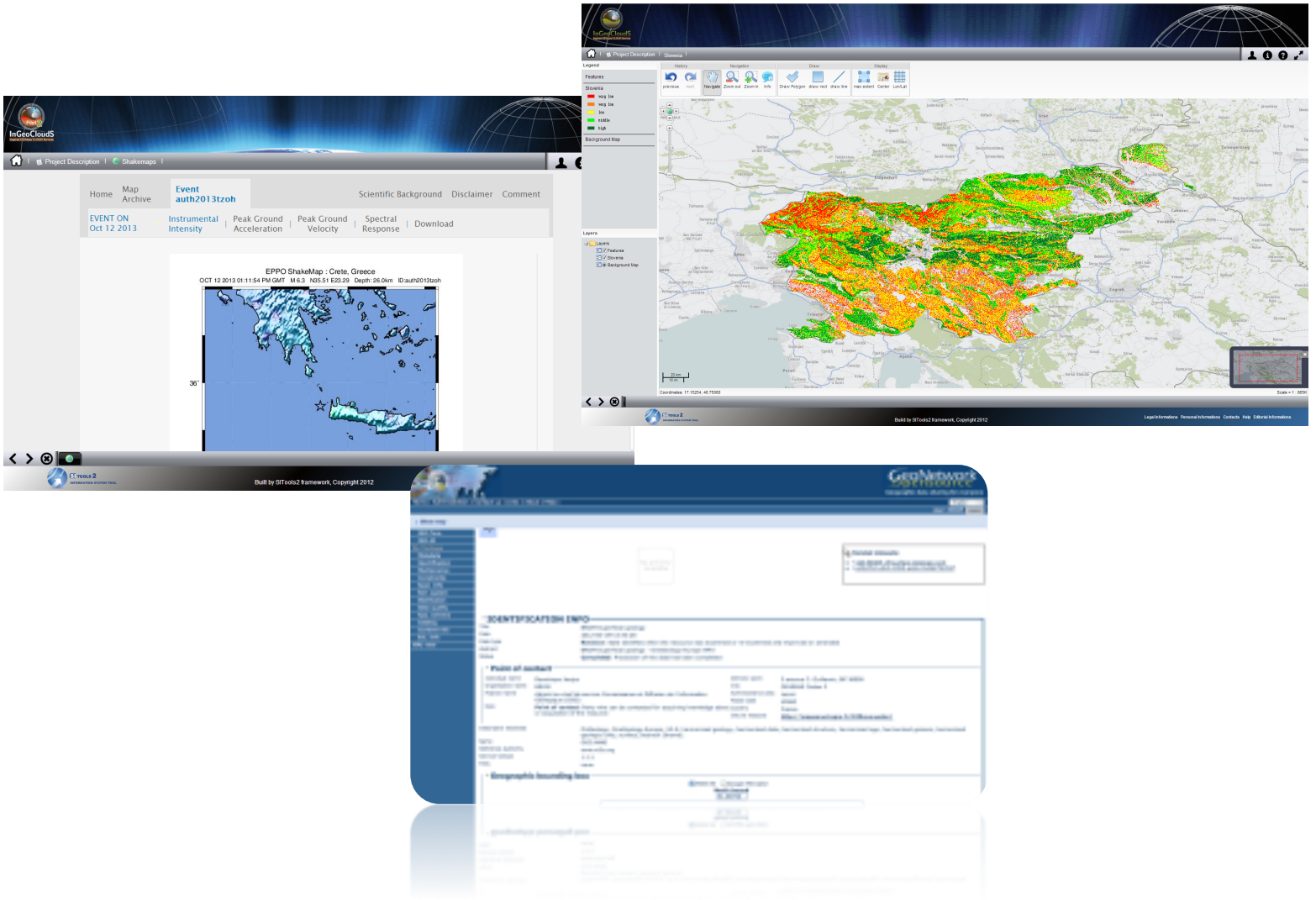
- Design and build a cloud infrastructure for public agencies in the spatial-environmental field
- Provide an innovate (flexible,…) infrastructure for geo-data services
- “Move” public services to a cloud-based infrastructure.
- Integrate geo-data by exploiting a Linked Data “model”

Infrastructure Description

The Big Picture



Integrated Geo Applications (www.ingeoclouds.eu)



Linked (Open) Data as a Service

Publishing Linked Data

- URI construction
- Conceptual Model
- Storage as RDF files or SPARQL endpoints

Querying Linked Data

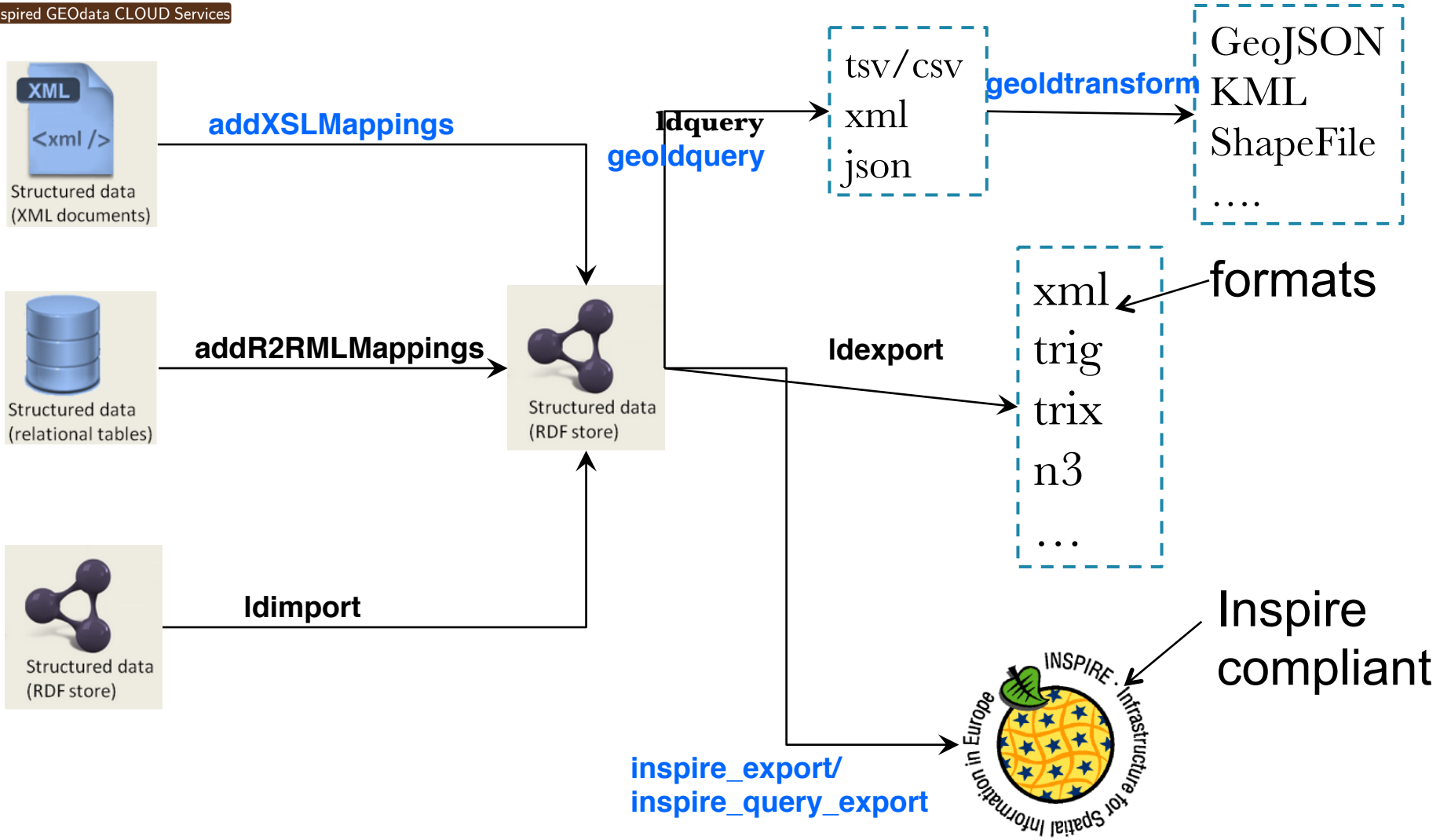
- SPARQL
- GeoSPARQL

Updating Linked Data

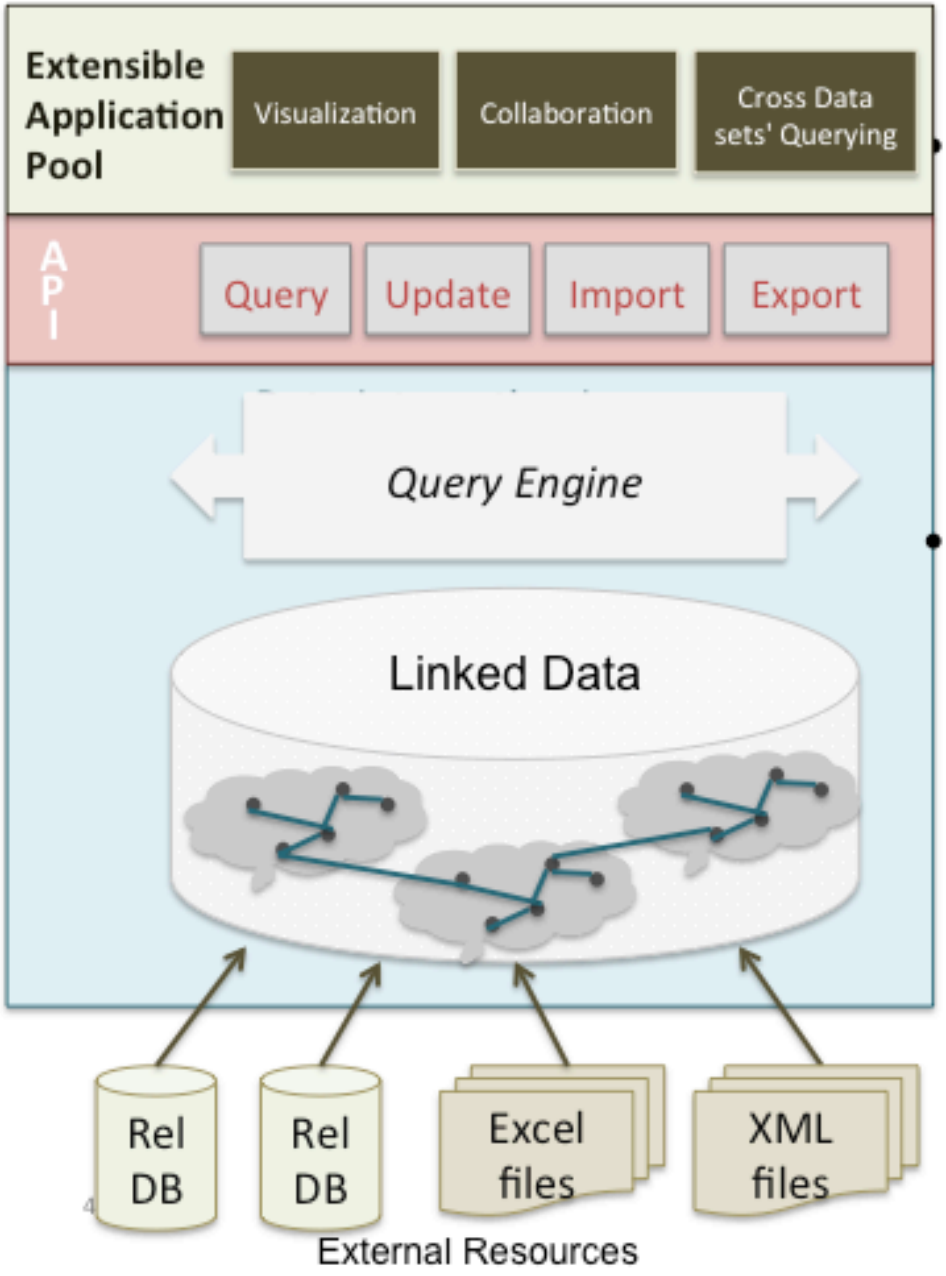
- SPARUL
- Synchronization with original sources

Linked Data Services

InGeoCloudS
Inspired GEOdata CLOUD Services



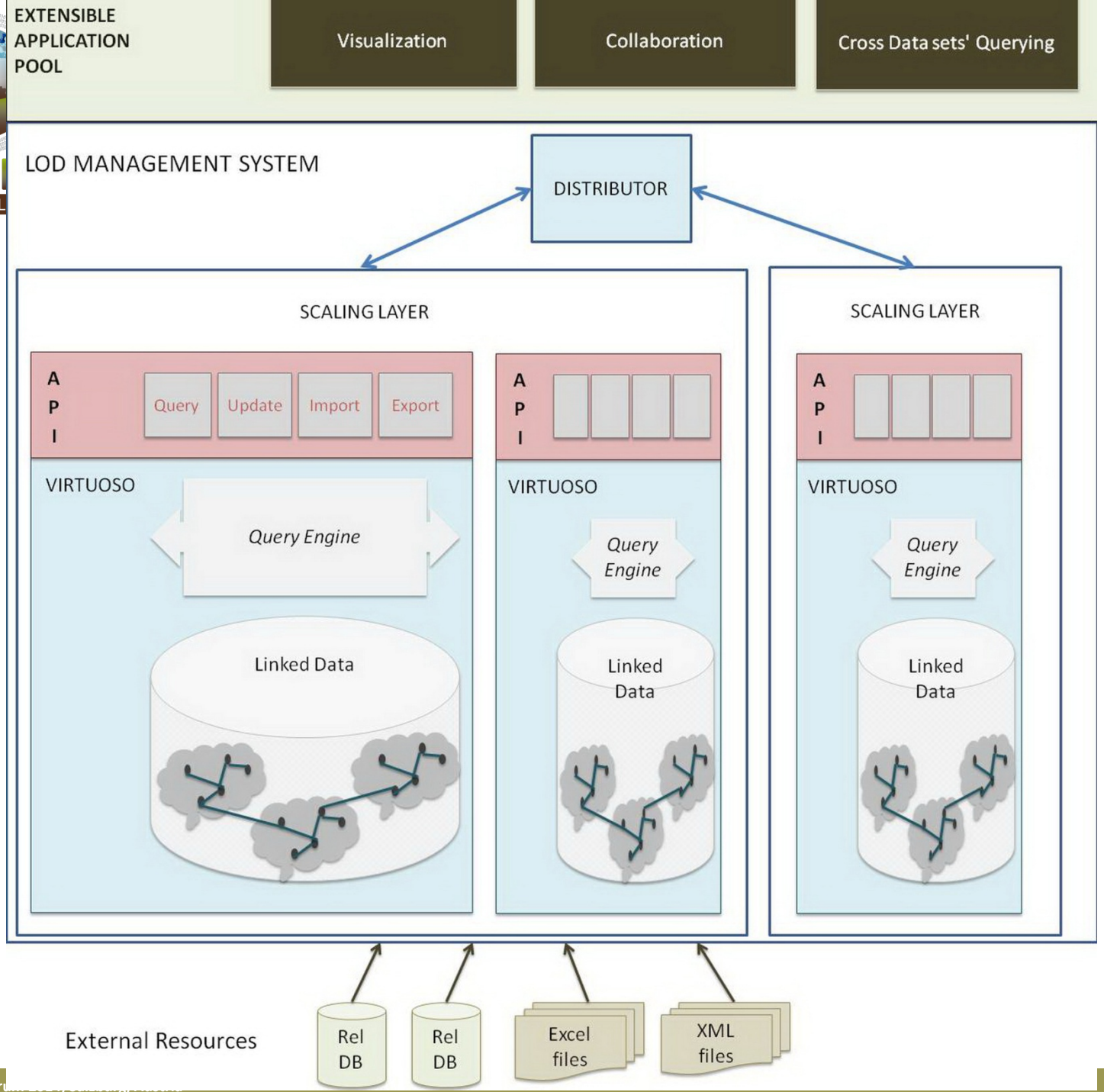
Linked (Open) Data as a Service



- Abstraction layer for data access

  abstract the applications from the specific setup of the data management service (such as local vs. remote, federation, and distribution)

- Beyond Data Access
  - Enabling automation of discovery, composition, and use of datasets
  - Data Markets
  - Online Visualization Services
  - Data Publishing Solutions
  - Data Aggregators
  - BI / Analytics as a Service


EXTENSIBLE APPLICATION POOL
Visualization
Collaboration
Cross Data sets' Querying
LOD MANAGEMENT SYSTEM
DISTRIBUTOR
SCALING LAYER
SCALING LAYER
API
Query
Update
Import
Export
VIRTUOSO
Query Engine
Linked Data
VIRTUOSO
Query Engine
Linked Data
VIRTUOSO
Query Engine
Linked Data
External Resources
Rel DB
Rel DB
Excel files
XML files

LOD Integration & Publishing (I)

URI Creation

(semi) automatic creation of URIs if needed

Two alternatives for publishing LOD:

1. Create and import RDF-based descriptions of data-sets via particular LMS method
   - Data update process must be controlled by performing SPARUL updates via particular LMS method
   - Data provider responsibility to keep synchronized relational & RDF data
     - A perfect synchronization may be also not required as it may incur costs -> second alternative becomes more preferable

LOD Integration & Publishing (II)

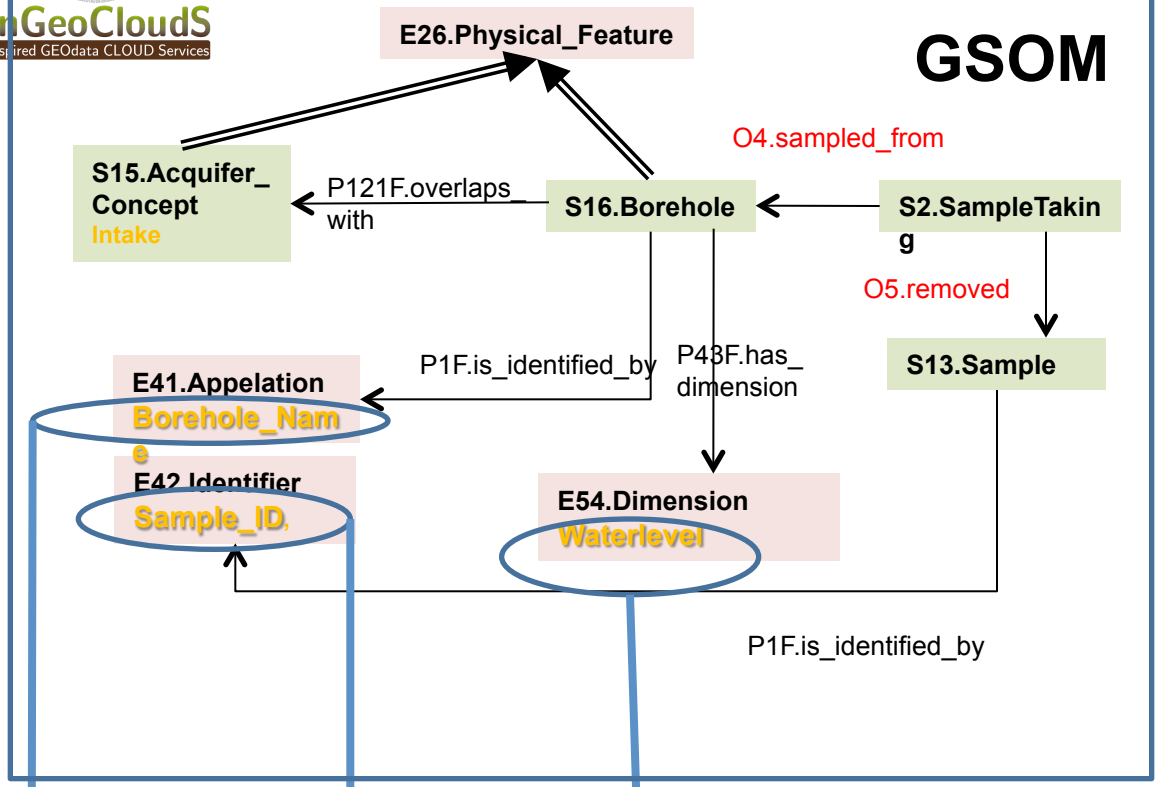
2. Data provider publishes relational data of his/her data sets + provides a mapping file in R2RML to enable the synchronization of relational to RDF data (by executing LMS method)
   - System takes care of this synchronization
   - Relational storage in the way used many years + additional RDF storage for the data with automatic one-way synchronization between the two
   - Provider should have a good knowledge of RDF

R2RML:

- W3C recommendation since 2012
- Can specify customized mappings between RDB & RDF data
- R2RML specification is just a RDF graph in Turtle
- No specific implementation is imposed

R2RML

GSOM



URI Identification:
http://orgURL/SampleID/XYZ

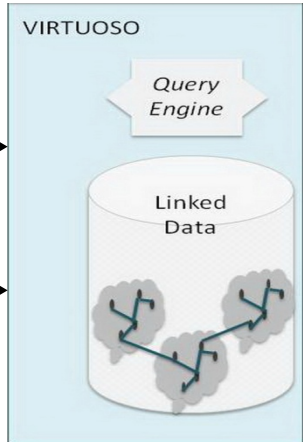
Borehole Name	Sample ID	WaterLevel
B1	XYZ	Level1

Borehole Relational Model

RDB

Publication

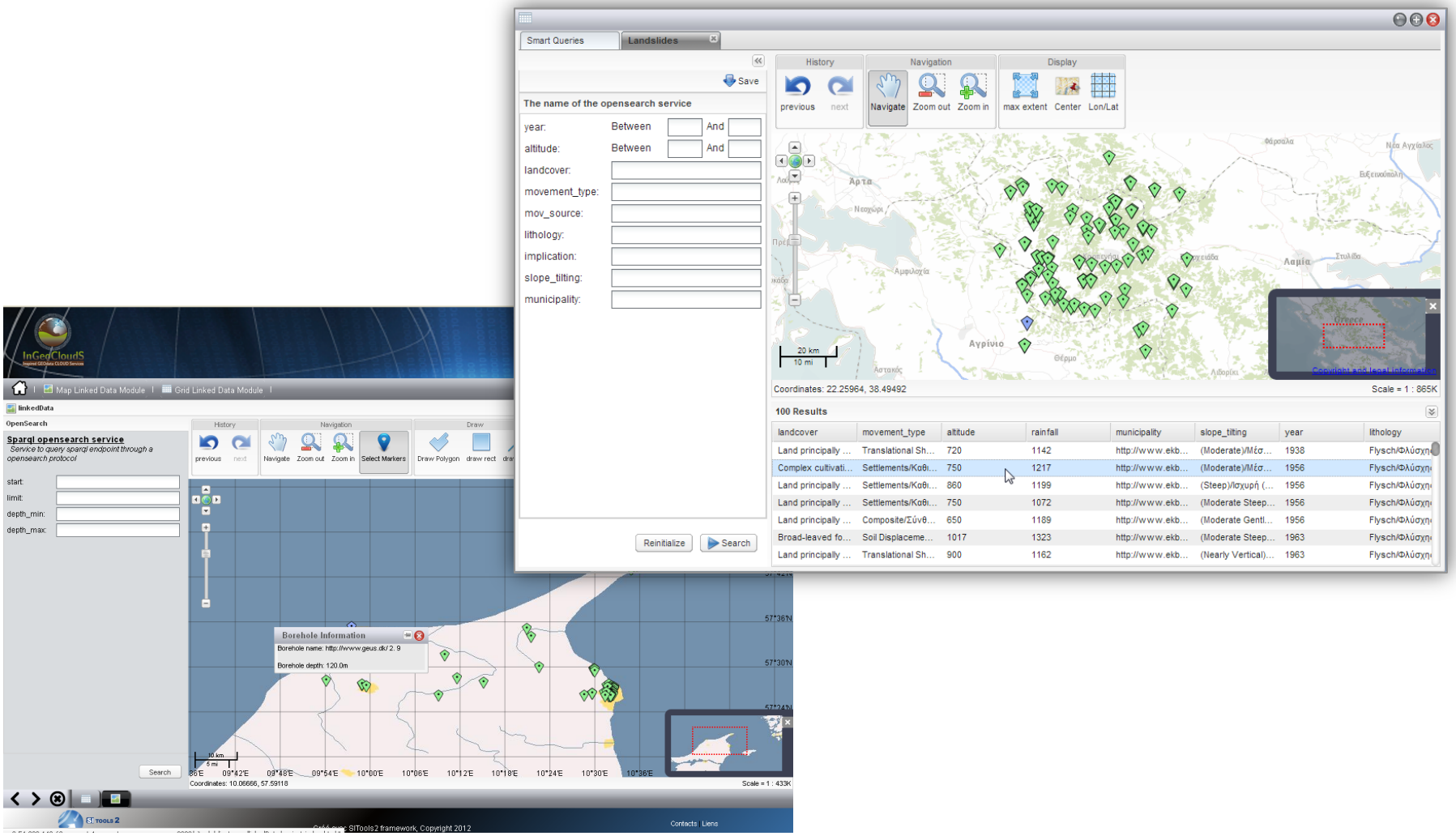
Synchronization



Export Linked Data into INSPIRE

- We created a our model(GSOM)-to-INSPIRE mapping mechanism which allows to transform GSOM-based RDF data, acquired from SPARQL queries, into XML-based INSPIRE-compliant data
- Exported Data are made INSPIRE-compliant
- 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)
- Ability to export:
  - Directly the results of the queries
  - Whole INSPIRE themes

Querying Linked Data Store



GeoProcessing

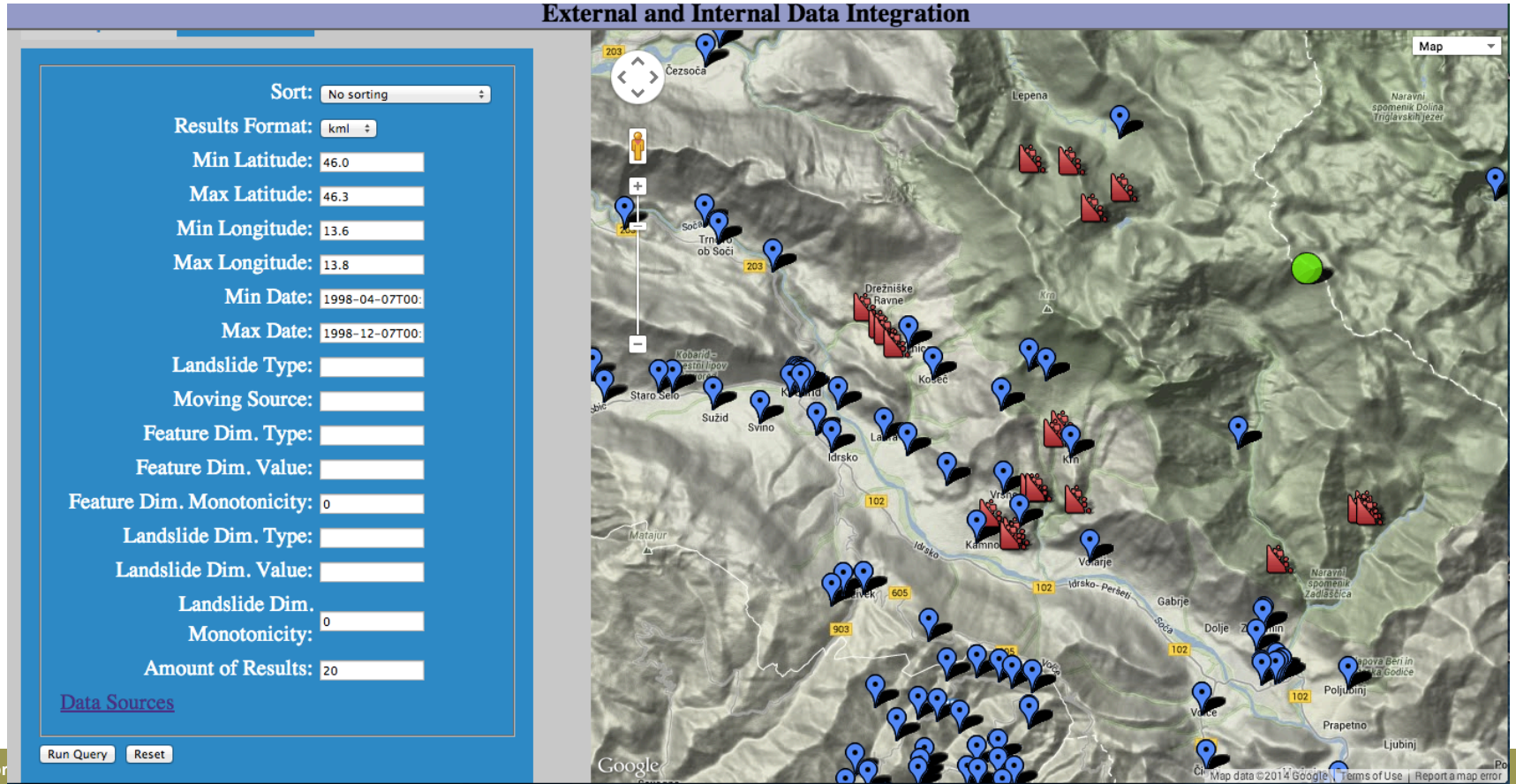
Geoprocessing implemented as a WPS service:
refers to ordinary kriging interpolation

WPS is provisioned in the ElasticWebServer component and uses parameters:

- Given by the user
- Fetched from InGeoCloudS Triplestore using LD-API

Linking Data Services

- Combine external and internal data sources
- Linking data to each other
- Retrieve information on demand



Challenges related to Linked Data and the Cloud - "Political" Challenges

- User adoption
  - Users are happy with what they have and they patiently wait until the problem surfaces
  - Users with weak or no infrastructure are most receptive to turn to Cloud and/or Linked Data paradigms than others
  - User see the cloud as a platform with more resources (memory, storage, processing) but they still want their applications to run there unchanged
  - **Security, privacy and trust**
- Public vs. Private Clouds
- Publicly owned vs. Company-owned Clouds

Challenges related to Linked Data and the Cloud - Scientific Challenges

- (Big) Linked Data storage
  - No “native” RDF triplestore that exploits the Cloud capabilities
  - Data security / privacy / control also a technical issue, especially when we deal with big data that you cannot monitor precisely
- Querying (Big) Linked Data
  - Fast querying for big data (many times the queries are simple but the volume is big)
  - Support for geospatial queries
  - (Fast) Indexing for (RDF) geospatial data
- “Exchanging” (Big) Linked Data
  - Data providers want to keep their own infrastructures and synchronize data between them and the cloud! Issues on synchronization and exchange of big data …

InGeoCloudS
Inspired GEOdata CLOUD Services

Conclusions

Proposed a scalable, geo-spatial LOD as-a-Service management system deployed on Amazon cloud

- Distributes query load + scales-up/down when CPU utilization surpasses specific thresholds
- Exposes REST-based service with LOD management methods
- Provides two different ways for publishing open geo-spatial data sets
- Provides various export possibilities

InGeoCloudS
Inspired GEOdata CLOUD Services

Thank you and Questions

www.ingeoclouds.eu

Dimitrios.Kotzinos@u-cergy.fr