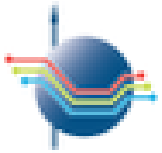
INSPIRE & Linked Data: Bridging the Gap

Part II: Tools for linked INSPIRE data



Athena
Research Center

Michalis Alexakis
Spiros Athanasiou
Nikos Georgomanolis
Giorgos Giannopoulos
Kostas Patroumpas
Thodoris Stratiotis
Nikos Vitsas

Objectives in this Session

- Geospatial linked data, metadata & services
  - Impact on INSPIRE Spatial Data Infrastructures (SDIs)
- Expose INSPIRE via (Geo)SPARQL endpoints?
  - making transformation to RDF triples possible with little work
  - *Interactive*: showcase open-source tools & (Geo)SPARQL interface
    - tripleGEO
    - tripleGEO/CSW
    - http://geodata.gov.gr/sparql
  - known limitations, good practices
- Benefits from reusing existing INSPIRE SDIs
  - currently for data discovery & retrieval
  - potentially for interlinking, fusion & reasoning
- Towards INSPIRE-aligned, semantic SDIs
  - harness the wealth of LOD cloud for novel applications

Who We Are

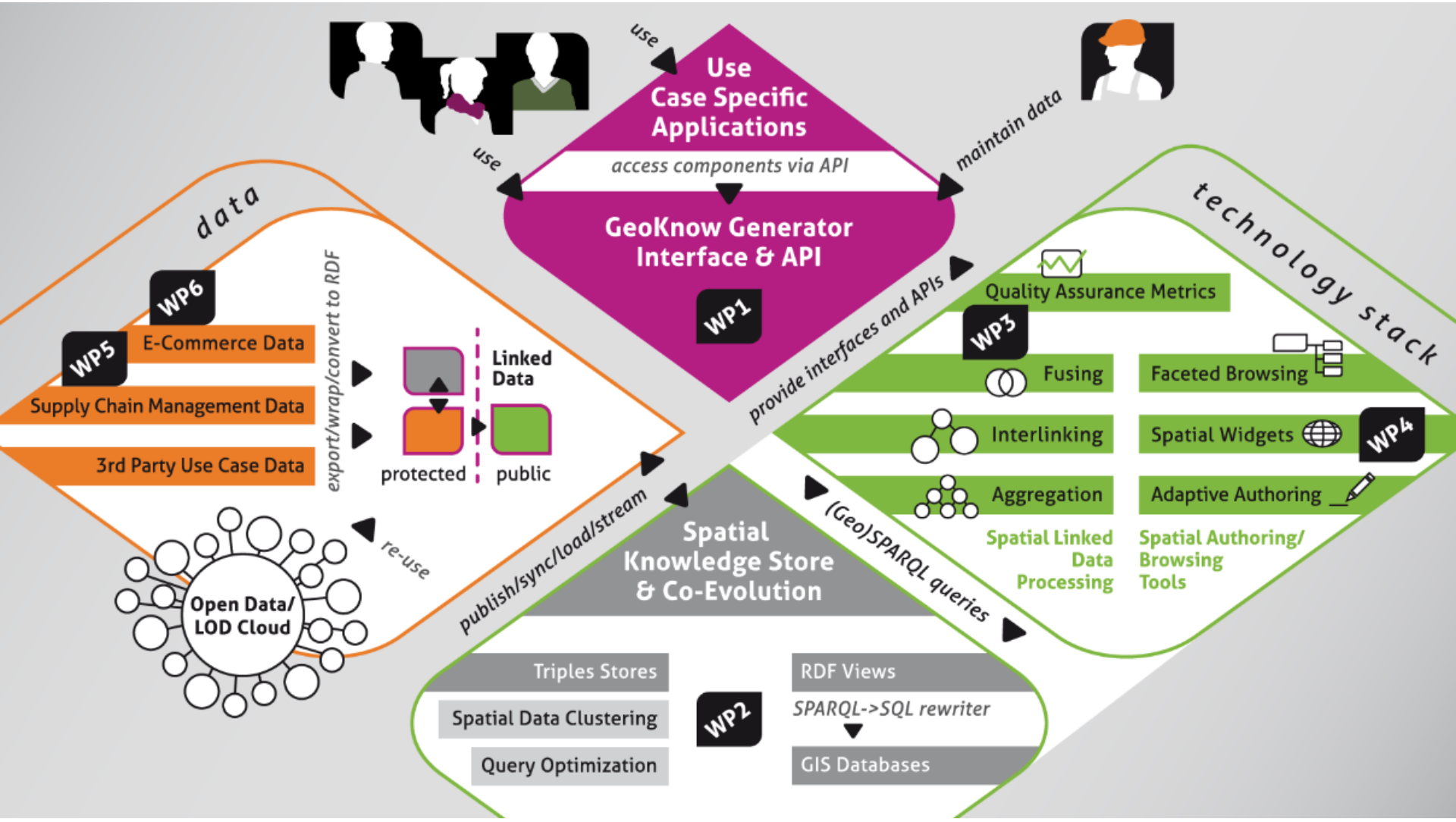
- "Athena" Research & Innovation Center
  - a research & technology body for Information, Communication & Knowledge Technologies
  - founded in 2001 by the Greek Ministry of Development
- Institute for the Management of Information Systems
  - founded in 2007
  - ~30 active RTD projects

Who We Are

- Research in **-data* management & information systems
  - big, web, geospatial, linked, open, scientific, cultural, meta-, … :-*data*
  - information retrieval
  - data mining / analytics
  - Applications

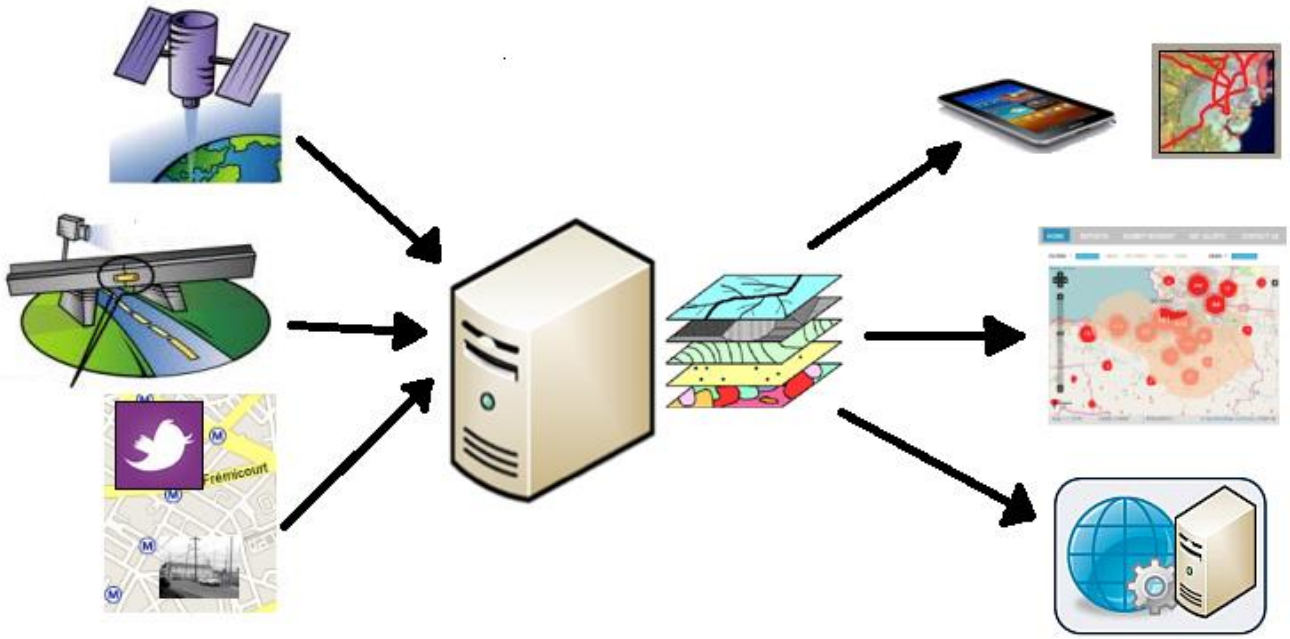
  - Expertise in geospatial data management
    - maintaining Greek geodata portal: http://geodata.gov.gr
    - taking part into INSPIRE WGs

Scope of the GeoKnow Project



Geospatial Data

- Geospatial data is ubiquitous in the Web
  - In a *variety* of formats, schemas, platforms, services
    - maps • satellite imagery • addresses • geotagged photos …
  - ➢ Crowdsourcing: OpenStreetMap GeoNames wikimapia
  - ➢ Standardization initiatives: *OGC, ISO, EU INSPIRE*



Geospatial Data

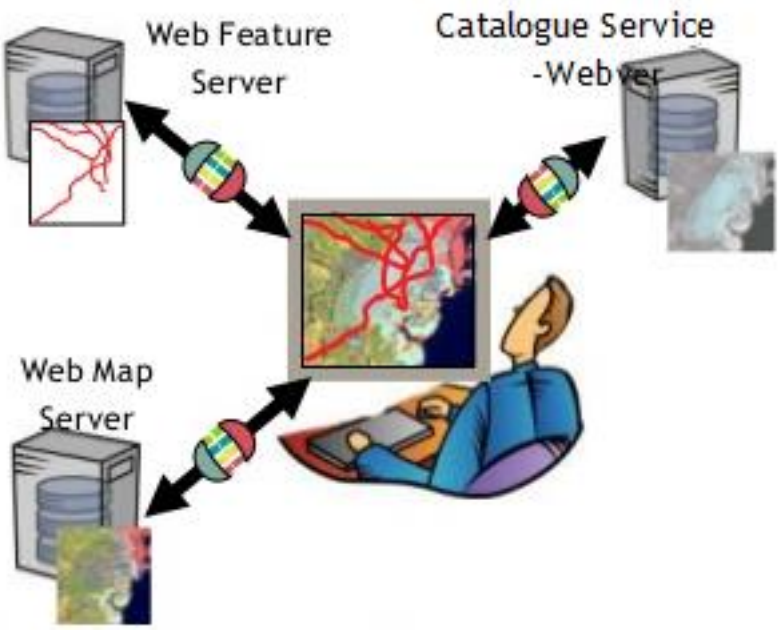
- Geospatial data is ubiquitous in the Web
  - In a *variety* of formats, schemas, platforms, services
    - maps • satellite imagery • addresses • geotagged photos …
  - ➢ Crowdsourcing:  OpenStreetMap  GeoNames wikimapia
  - ➢ Standardization initiatives: *OGC, ISO, EU INSPIRE*
- Semantic technologies
  - for exposing, sharing, and linking data resources in RDF
- Enhancing RDF with geospatial support
  - Not much such data published. *Why?*
    - Complexity of geometry representations & georeferencing
    - Expressing topological relationships in queries → OGC GeoSPARQL
    - Lack of conversion tools to enable data integration
  - ➢ *Bridge the gap between GIS and Semantic Web communities*

Geospatial Metadata

- Assess quality of geodata with reliable *metadata*
  - A brief summary on the spatial dataset & its creation procedures
    - Content, purpose, owner, license, status, location/coverage, …
- ISO 19115:2003 & ISO 19115-1:2014
  - standardized metadata for discovery & retrieval of geodata
  - ➢ *Assist users in choosing spatial features that best suit their needs*
    - Coordinate reference system (CRS)
    - Resolution → the map scale used in digitization
    - Date of last update
    - Textual keywords on the content of digital maps
- Spatial Metadata as Linked Data Sources
  - Take advantage of Data Catalogue (DCAT) & VoID vocabularies
  - ➢ RDF mappings for INSPIRE metadata elements [JRC@EC]
    - based on DCAT-AP, DCT, SKOS, vCard, ... → *our starting point!*

Geospatial Catalogue Services

- OGC Catalogue Services for the Web (*CSW*)
  - *application profiles* of services based on spatial metadata *in XML*
    - resource characteristics can be published, queried, and processed
  - Accept user requests through **`GET/POST`** HTTP protocols



Geospatial Catalogue Services

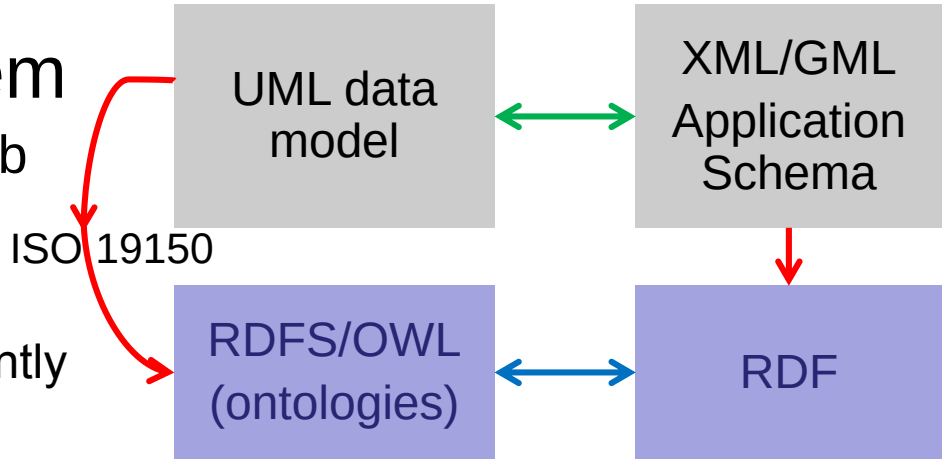
- OGC Catalogue Services for the Web (*CSW*)
  - *application profiles* of services based on spatial metadata *in XML*
    - resource characteristics can be published, queried, and processed
  - Accept user requests through **`GET/POST`** HTTP protocols
  - Return *metadata records* with standard schema & encoding
    - title of dataset
    - geographical extent & CRS
    - licensing policies …
    - format (e.g., GML)
    - links to associated metadata
- CSW still disjoint from the Semantic Web
    - Machine readable metadata? Interlinked with third-party information?
  - ➢ *Recent initiatives may act as catalysts:*
    - OGC GeoSPARQL standard: expose & query spatial metadata *in RDF*
    - EU INSPIRE infrastructure: accessible via SPARQL endpoints?

Use Case on INSPIRE

- INSPIRE Directive 2007/2/EC
  - by 2020: harmonize spatial metadata, data & services in the EU
    - common access policies & standards for spatial datasets
    - *catalogues* of available resources using metadata
    - *network services* for data discovery, view, download, transformation, etc.
- Exposing INSPIRE as Linked Data
  - can be a game-changer for industry, research & governance
    - without affecting established operation of SDIs
- Opportunities for Linked Data from INSPIRE
  - "official" public data → proven high-quality & large coverage
  - volume exceeds any available crowdsourced data (e.g., OSM)
  - reuse/interlink data across diverse domains & applications

A "*Missing Link*"?

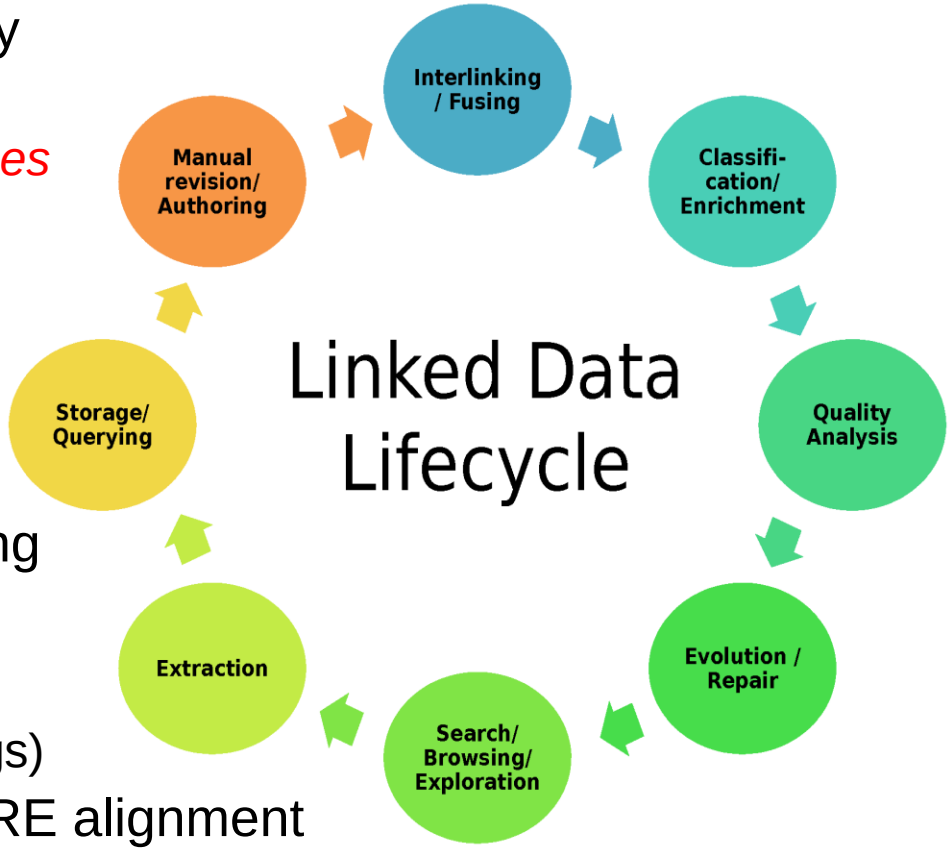
- The INSPIRE ecosystem
  - is disjoint from Semantic Web
  - based on standards
  - towards *diachronic* SDIs
    - …but very few data is currently
    - published as linked data!



- No complete INSPIRE ontology in RDF/OWL *yet*
  - The "*closed world*" assumption of UML models in INSPIRE …
  - … incompatible with the "*open world*" view of RDF
- *Our objectives*
  - Provide (semi-)automatic *tools* to expose INSPIRE data & metadata
  - Repurpose/reuse geospatial information from INSPIRE SDIs
  - Interface to (Geo)SPARQL endpoints: http://geodata.gov.gr/sparql

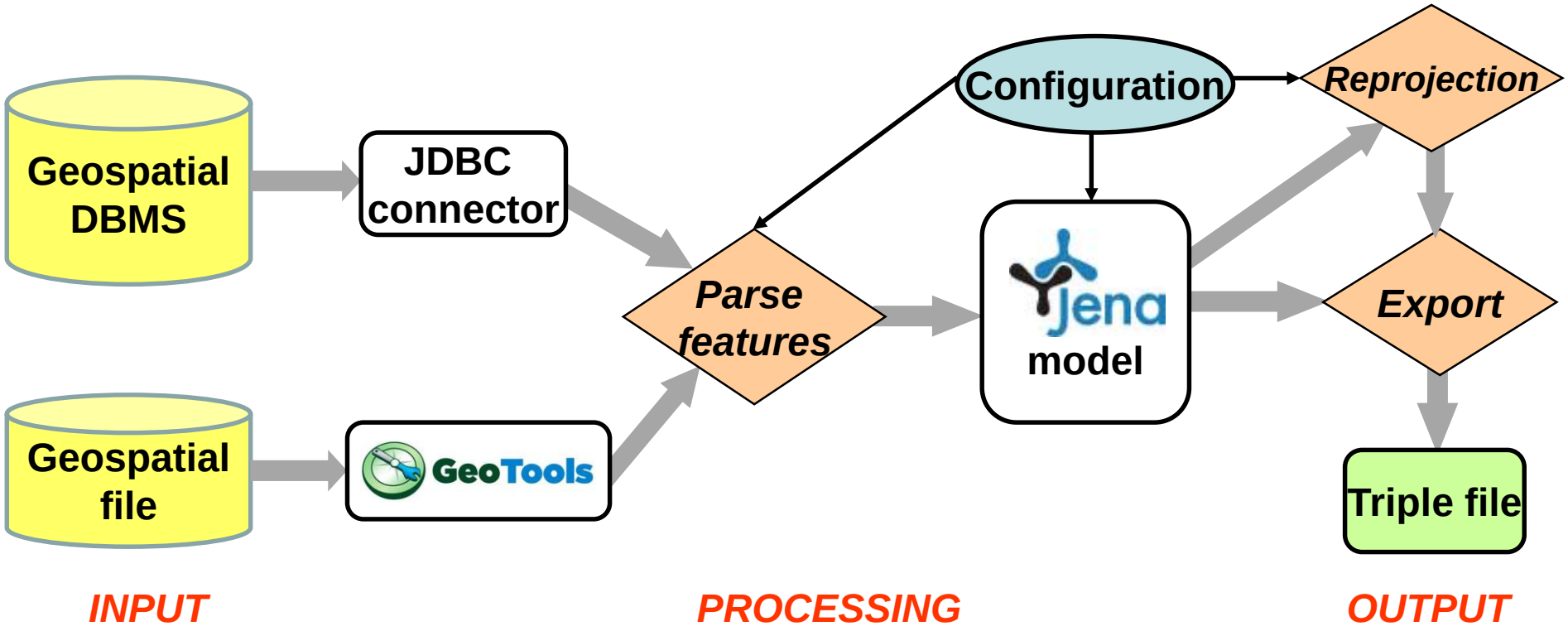
Tools for Linked INSPIRE

- **tripleGEO/CSW** → *Search/Browsing/Exploration*
  - Explore the quantity & quality of INSPIRE-compliant datasets available from *Catalogue Services* on metadata across Europe
  - Expose existing CSWs through a SPARQL endpoint
- **tripleGEO** → *Extraction*
  - reusable scripts for translating INSPIRE *data* & *metadata* from GML/XML into RDF (XSL stylesheets, RDF mappings)
  - recommendations for INSPIRE alignment
  - best practices for such transformations

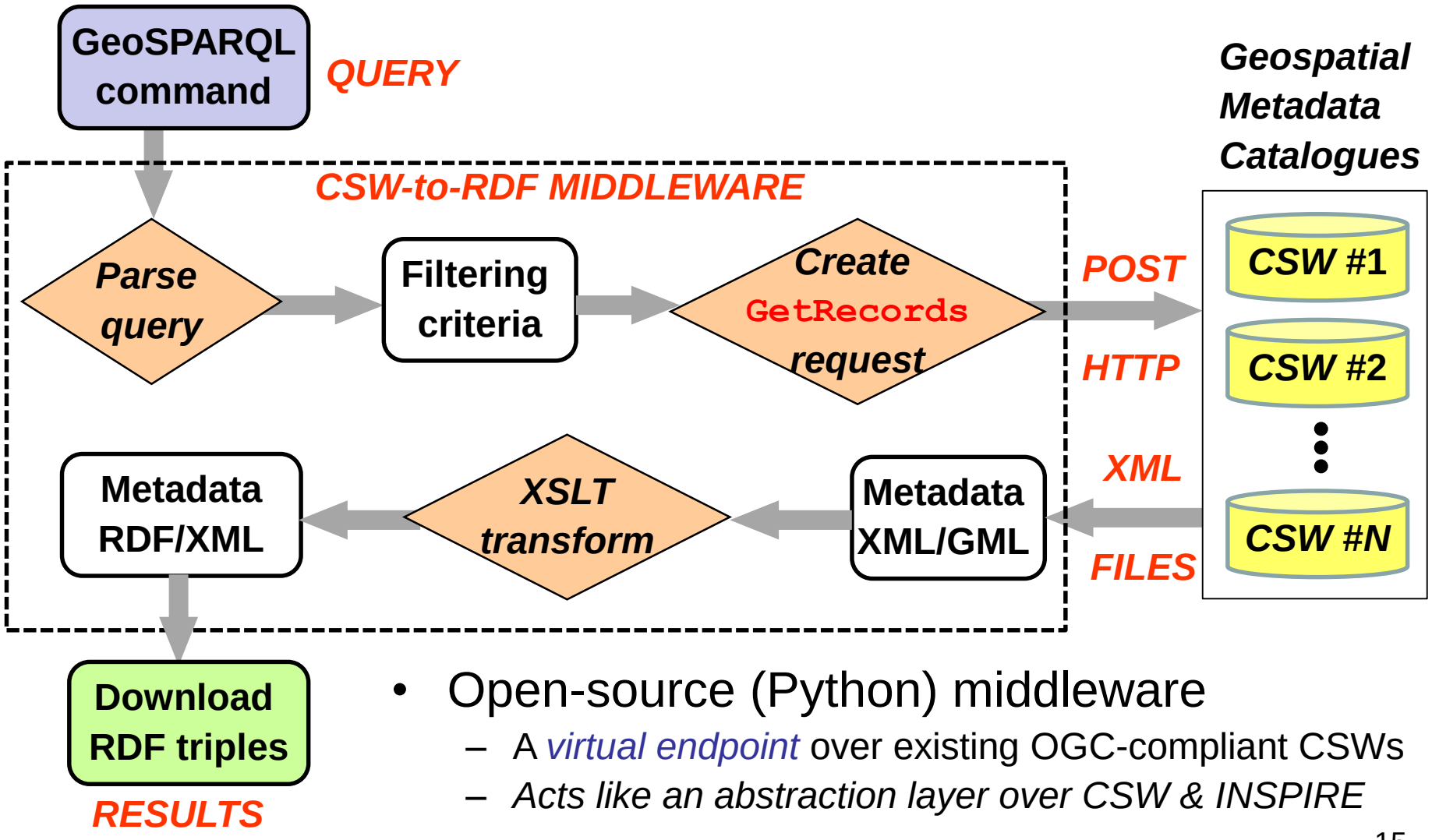


tripleGEO – Overview

- Open-source (Java) *spatial ETL tool* for RDF repositories
  - v1.0: from files (shp, GML, KML) or DBMS (Oracle, PostGIS, …) to RDF
  - *v1.1* integrates XSL stylesheets for transforming INSPIRE to RDF



tripleGEO/CSW – Overview



- Open-source (Python) middleware
  - A *virtual endpoint* over existing OGC-compliant CSWs
  - *Acts like an abstraction layer over CSW & INSPIRE*

Complying with INSPIRE

- Metadata
  - Published according to *ISO standards* 19115, 19119
  - Enrich with *terms* from code lists & multi-lingual thesauri
  - Specify spatial coverage & temporal evolution
- Data
  - Data models specified in UML at conceptual level
    - Geospatial encoding using GML
  - INSPIRE *XML schemas* for harmonized data themes
    - Finalized for Annex I & Annex II
    - Draft implementations for Annex III
- Services according to *OGC standards*
  - Web Map Services (*WMS*) → mapping, visualisation
  - Web Feature Services (*WFS*, *WFS-T*) → query, editing
  - Catalogue Services for the Web (*CSW*) → discovery

i) Translate INSPIRE *metadata* to RDF

- Define RDF mappings based on
  - RDF mappings to vocabularies: DCAT, DCT, SKOS, vCard, etc.
  - *preliminary version* by courtesy of the JRC@*European Commission*:

INSPIRE metadata element	RDF mapping of attribute
Resource title	`dct:title`
Resource language	`dct:language`
Keyword	`dcat:keyword`
Geographic Bounding Box	`dct:spatial`
Responsible organization - Owner	`dct:rightsHolder`

  - Main elements:
    - spatial coverage (as *MBR*)
    - temporal range
    - descriptions (e.g., *title, abstract, subject, keywords*)
    - content assessments (like *quality, provenance,* or *conformity*)
    - legal status (*owner, license, point of contact*, etc.)

➢ Transform to RDF via a generic XSL stylesheet (integrated in tripleGEO)

ii) Translate INSPIRE *data* to RDF

- INSPIRE data from GML to RDF
  - No guidelines for exposing such information as Linked Data
- Our approach is *data-centric*, not *model-based*
  - a custom XSL stylesheet per INSPIRE Data Theme (Annex I)
    - encodes elements from INSPIRE GML application schemata into RDF
    - imitates INSPIRE schema for a data theme
    - produces a compliant RDF representation
    - XSLT *not aware* of INSPIRE vocabulary, re-used with any parser
  - Challenging issues:
    - Resolvable *http URIs*
    - Meaningful *INSPIRE identifiers*
    - Geometry serializations: *GeoSPARQL*, others?
    - Peculiarities in thematic attributes: *multi-linguality, voidable properties*
  - Method integrated into tripleGEO tool

INSPIRE Data Themes

ANNEX I

1 Coordinate reference systems
2 Geographical grid systems
3 Geographical names
4 Administrative units
5 Addresses
6 Cadastral parcels
7 Transport networks
8 Hydrography
9 Protected sites

ANNEX II

1 Elevation
2 Land cover
3 Orthoimagery
4 Geology

ANNEX III

1 Statistical units
2 Buildings
3 Soil
4 Land use
5 Human health and safety
6 Utility and governmental services
7 Environmental monitoring Facilities
8 Production and industrial facilities
9 Agricultural and aquaculture facilities
10 Population distribution and demography
11 Area management / restriction / regulation zones & reporting units
12 Natural risk zones
13 Atmospheric conditions
14 Meteorological geographical features
15 Oceanographic geographical features
16 Sea regions
17 Bio-geographical regions
18 Habitats and biotopes
19 Species distribution
20 Energy Resources
21 Mineral Resources

iii) GeoSPARQL Requests over *CSW*

- Predefined application profile for spatial metadata
  - a set of XSLT templates, also *handling GML geometries*
  - **tripleGEO/CSW**: transform XML/GML elements into RDF on-the-fly
- **`SELECT / CONSTRUCT`** queries in (Geo)SPARQL
  - **`WHERE`** clause: graph pattern to match against the metadata
  - \+ optional **`FILTER`** criteria:
    - *Textual:* matching regular expressions (e.g., keywords)
    - *Temporal:* date comparisons (e.g., last update)
    - *Spatial:* region of interest
  - Parser recognizes typical GeoSPARQL topological predicates
    - **`sfWithin()`**
    - **`sfContains()`**
    - **`sfIntersects()`**
    - **`sfOverlaps()`**, …

Example: Query → Request → Results

```
PREFIX dcat: <http://www.w3.org/ns/dcat#>
PREFIX dc: <http://purl.org/dc/terms/>
PREFIX geo: <http://www.opengis.net/ont/geosparql#>
PREFIX geof: <http://www.opengis.net/def/geosparql/function/>
CONSTRUCT { ?m dcat:keyword ?k .
            ?s dc:subject ?sub .
            ?f geo:hasGeometry ?fWKT
          }
WHERE { ?m dcat:keyword ?k .
        ?s dc:subject "Environment" .
        ?f geo:hasGeometry ?fWKT .
        FILTER (REGEX(str(?k),"^water*")  &&
                geof:sfWithin(?fWKT, "BOX2D(-8.24 54.02,-5.18 55.32)"^^geo:wktLiteral))
      };
```

GeoSPARQL query

Example: Query → Request → Results

```
<?xml version='1.0' encoding='utf-8'?>
<GetRecords
     xmlns="http://www.opengis.net/cat/csw/2.0.2"
     xmlns:csw="http://www.opengis.net/cat/csw/2.0.2"
      :
      :
     xsi:schemaLocation="http://www.opengis.net/cat/csw/2.0.2
     http://schemas.opengis.net/csw/2.0.2/CSW-discovery.xsd">
  <Query typeNames="gmd:MD_Metadata">
    <ElementSetName typeNames="gmd:MD_Metadata">full
    </ElementSetName>
    <Constraint version="1.1.0">
      <ogc:Filter>
        <ogc:And>
          <ogc:PropertyIsEqualTo>
            <ogc:PropertyName>dc:subject</ogc:PropertyName>
            <ogc:Literal>Environment</ogc:Literal>
          </ogc:PropertyIsEqualTo>
          <ogc:PropertyIsLike wildCard="^" singleChar="_">
            <ogc:PropertyName>dcat:keyword</ogc:PropertyName>
            <ogc:Literal>^water*</ogc:Literal>
          </ogc:PropertyIsLike>
          <ogc:Within>
            <ogc:PropertyName>ows:BoundingBox</ogc:PropertyName>
            <gml:Envelope>
              <gml:lowerCorner>-8.24 54.02</gml:lowerCorner>
              <gml:upperCorner>-5.18 55.32</gml:upperCorner>
            </gml:Envelope>
          </ogc:Within>
        </ogc:And>
      </ogc:Filter>
    </Constraint>
  </Query>
</GetRecords>
```

CSW request

Example: Query → Request → Results


```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:xs="
http://www.w3.org/TR/2008/REC-xml-20081126#" xmlns:xsd="http://www.w3.org/2001/XMLSchema" xmlns:dct="
http://purl.org/dc/terms/" xmlns:dcat="http://www.w3.org/ns/dcat#" xmlns:xsi="
http://www.w3.org/2001/XMLSchema-instance" xmlns:gmd="http://www.isotc211.org/2005/gmd" xmlns:foaf="
http://xmlns.com/foaf/0.1/" xmlns:locn="http://www.w3.org/ns/locn#" xmlns:prov="http://www.w3.org/ns/prov#"
xmlns:rdfs="http://www.w3.org/1999/01/rdf-schema#" xmlns:schema="http://schema.org" xmlns:skos="
http://www.w3.org/2004/02/skos/core#" xmlns:vcard="http://www.w3.org/2006/vcard/ns#" xmlns:gml="
http://www.opengis.net/gml/" xmlns:gco="http://www.isotc211.org/2005/gco">
  <rdf:Description rdf:about="urn:uuid:60fabe5d-73ad-4523-a23a-5d2eb3981f1f">
    <foaf:primaryTopicOf rdf:resource="urn:uuid:{89C8EB27-B20F-4908-A961-A01375FBE079}"/>
    <dct:language rdf:datatype="http://purl.org/dc/terms/ISO639-2">eng</dct:language>
    <dct:title xml:lang="eng">Rivers Agency (NI) Coastal High Water Outline (Metadata)</dct:title>
    <dct:description xml:lang="eng">This WMS service represents a combined service of 11 datasets that make up
    Annex I &amp;amp; II dataset Natural England is supplying as part of INSPIRE directive.
The datasets are Limestone Pavement Orders (England), RAMSAR (England), Biosphere Reserves (England), Areas of
Outstanding Natural Beauty (England), Local Nature Reserves (England), Marine Conservation Zones (England),
National Nature Reserves (England), National Parks (England),  Sites of Special Scientific Interest (England),
Special Areas of Conservation (England), and Special Protection Areas (England).
By using this data you are accepting the Terms of Use for Natural England&amp;#8217;s Information and Data as
published at: http://www.naturalengland.org.uk/copyright. If you wish to use the data for commercial purposes
you should contact Natural England's Enquiry Service, tel: 0845 600 3078</dct:description>
    <rdf:type rdf:resource="http://www.w3.org/ns/dcat#Dataset"/>
    <dcat:landingPage rdf:resource=""/>
    <dct:identifier rdf:datatype="http://www.w3.org/2001/XMLSchema#string"/>
    <dct:subject>environment</dct:subject>
    <dcat:keyword xml:lang="eng">Hydrography</dcat:keyword>
    <dcat:keyword xml:lang="eng">hydrography</dcat:keyword>
    <dcat:keyword xml:lang="eng">coastal water</dcat:keyword>
    <dcat:keyword xml:lang="eng">Hydrography</dcat:keyword>
    <dcat:keyword xml:lang="eng">INSPIRE</dcat:keyword>
    <dcat:keyword xml:lang="eng">Rivers Agency</dcat:keyword>
```

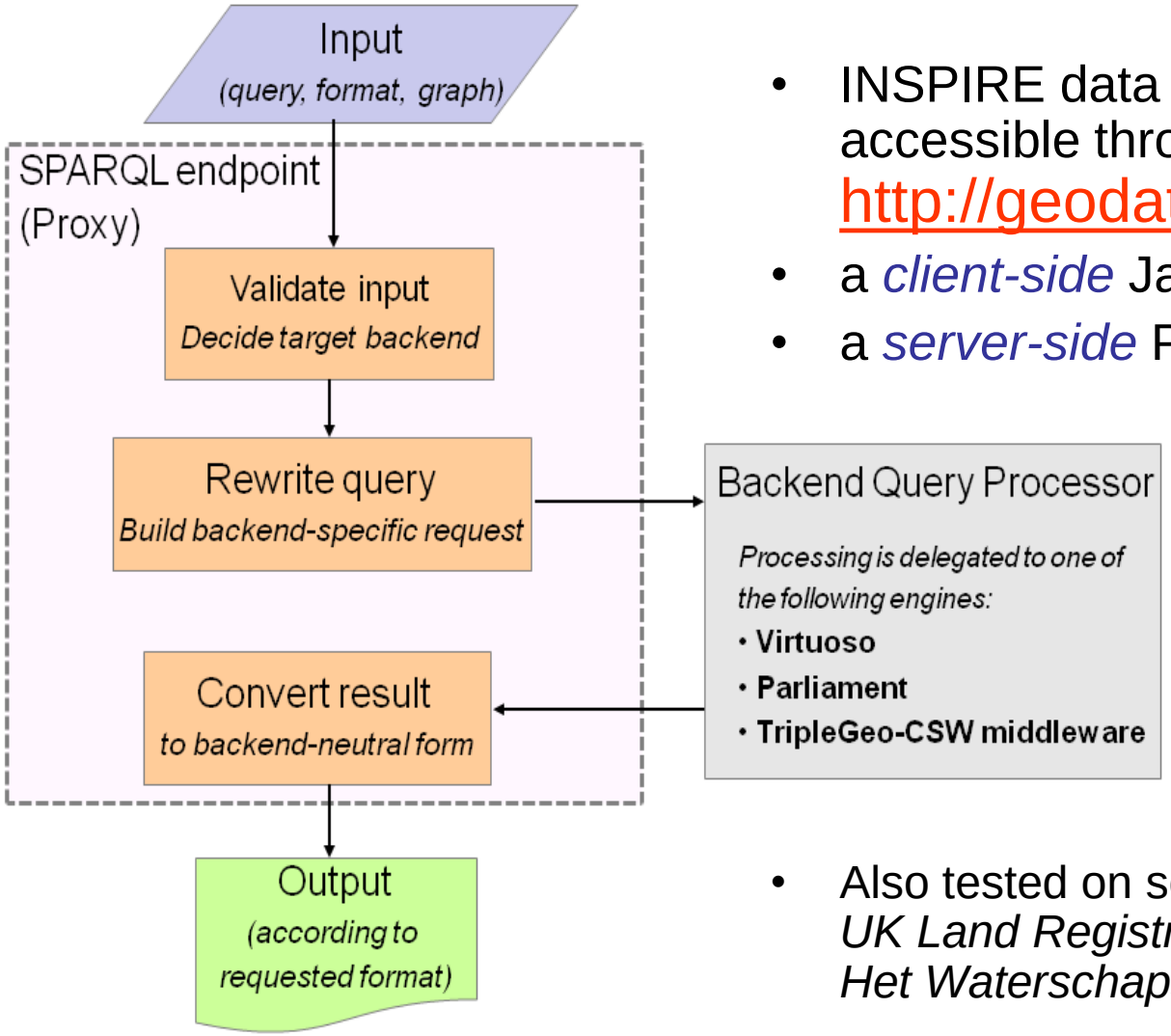

RDF triples

Use Case on a Real-World SDI

- *Proof-of-concept* on INSPIRE data/metadata
  - Repository of Greek public geodatasets: http://geodata.gov.gr
  - One indicative dataset for seven INSPIRE data themes (*Annex I*)
    - *administrative units*
    - *protected sites, ...*
  - Original datasets were *not* INSPIRE-aligned
    - we created schema mappings into INSPIRE GML using HALE tool
  - Transforming features using customized *XSL stylesheets*
    - thanks to parameterised URIs & geometry serialisations
- RDF triples loaded in two repositories:
  - *Virtuoso*: for scalability, query performance
  - *Parliament*: for GeoSPARQL compliance
  - two separate graphs (data, metadata) in each triple store
  + a virtual *GeoSPARQL* endpoint from a list of CSWs across Europe

INSPIRE-aligned CSWs in Use

(Geo)SPARQL Endpoints for INSPIRE



- INSPIRE data & metadata accessible through a web interface http://geodata.gov.gr/sparql
- a *client-side* JavaScript
- a *server-side* PHP

- Also tested on selected datasets from: *UK Land Registry, Localities (IT), Het Waterschapshuis (NL)*

Summary

- A *data-centric* approach for Linked INSPIRE
  - Encoding elements from INSPIRE GML appl. profiles into RDF
  - Using XSLT to produce an INSPIRE-aligned RDF representations
    - Customisable for both data & metadata
    - Geometry serialisations comply with GeoSPARQL
- *Proof-of-concept*
  - *Open-source* software tools:
    - tripleGEO
    - tripleGEO/CSW
  - Web interface: http://geodata.gov.gr/sparql
    - data + metadata: reusing features from an SDI from Greece
    - CSW: across Europe

Lessons Learnt

- (Geo)SPARQL queries tend to be quite verbose
  - because graphs reflect the detailed INSPIRE schemata
- Increased response times for evaluation
  - too many triple bindings → *optimisation?*
  - depend on the capabilities of RDF stores
  - ➢ orthogonal to our methodology
- Method inherently extensible for Annex II & III data
  - simply authoring extra XSLT transformations!
- Explore interlinking of INSPIRE RDF with LOD
  - e.g., OpenStreetMap

Discussion

- Ongoing effort by JRC@EC
  - GeoDCAT-AP for *metadata*
    - https://joinup.ec.europa.eu/node/139283
  - \+ Other initiatives
    - ARE3NA RDF + PIDs, W3C, SmartOpenData, ...
- Exposing INSPIRE data on the Semantic Web?
  - Special use cases ?
  - Subtle issues ?
  - Recommendations ?
  - Potential applications ?

INSPIRE & Linked Data: Bridging the Gap

Part II: Tools for linked INSPIRE data

Thank you!

Find out more:

http://geodata.gov.gr/sparql/

http://geoknow.eu

https://github.com/GeoKnow/TripleGeo

https://github.com/GeoKnow/TripleGeo-CSW