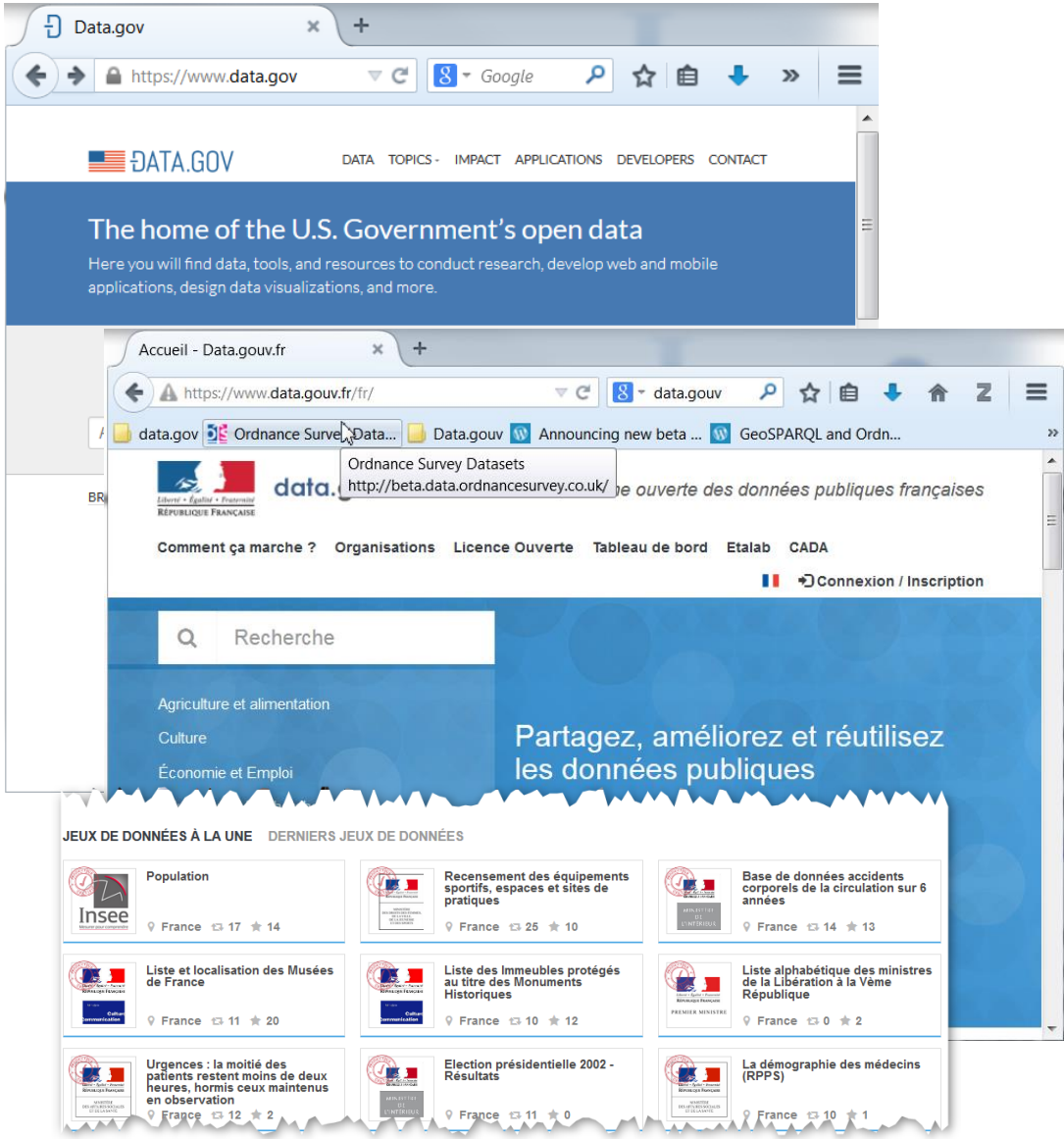
Données spatio-temporelles ouvertes et liées : surveillance des ressources en eau.

Danielle ZIEBELIN, Philippe GENOUD – Université Grenoble Alpes (France) Kim HOBUS – Yukon Water Board (Canada)

Participation : Sylvain Bouveret & Camille Bernard (LIG-Steamer)

Open Data

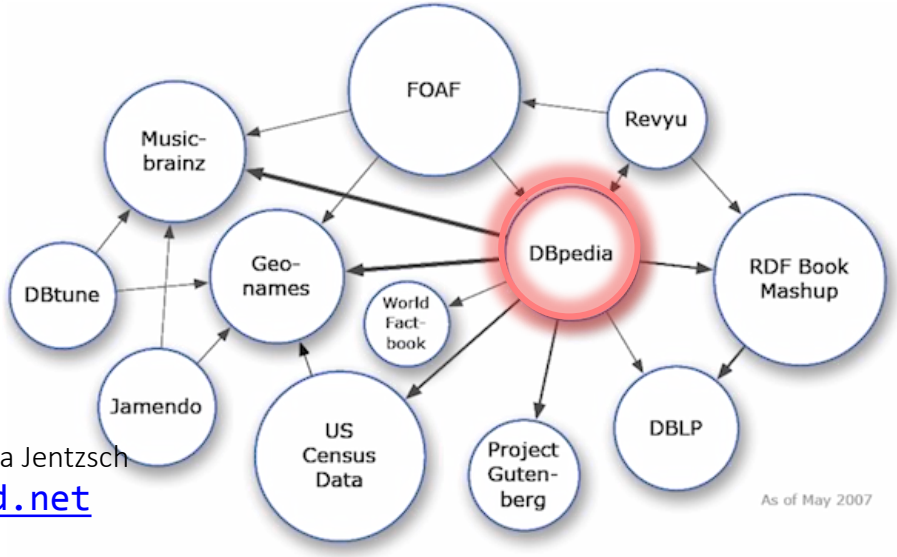
- Open Data Movement
  - *"A piece of content or data is open if anyone is free to use, reuse, and redistribute it — subject only, at most, to the requirement to attribute and/or share-alike."* http://opendefinition.org/
  - an old idea but a recent term gaining popularity
    - with the rise of the Internet and World Wide Web
    - with the launch of open-data government initiatives such as Data.gov (USA), data.gouv.fr (FR)…



Linked Open Data

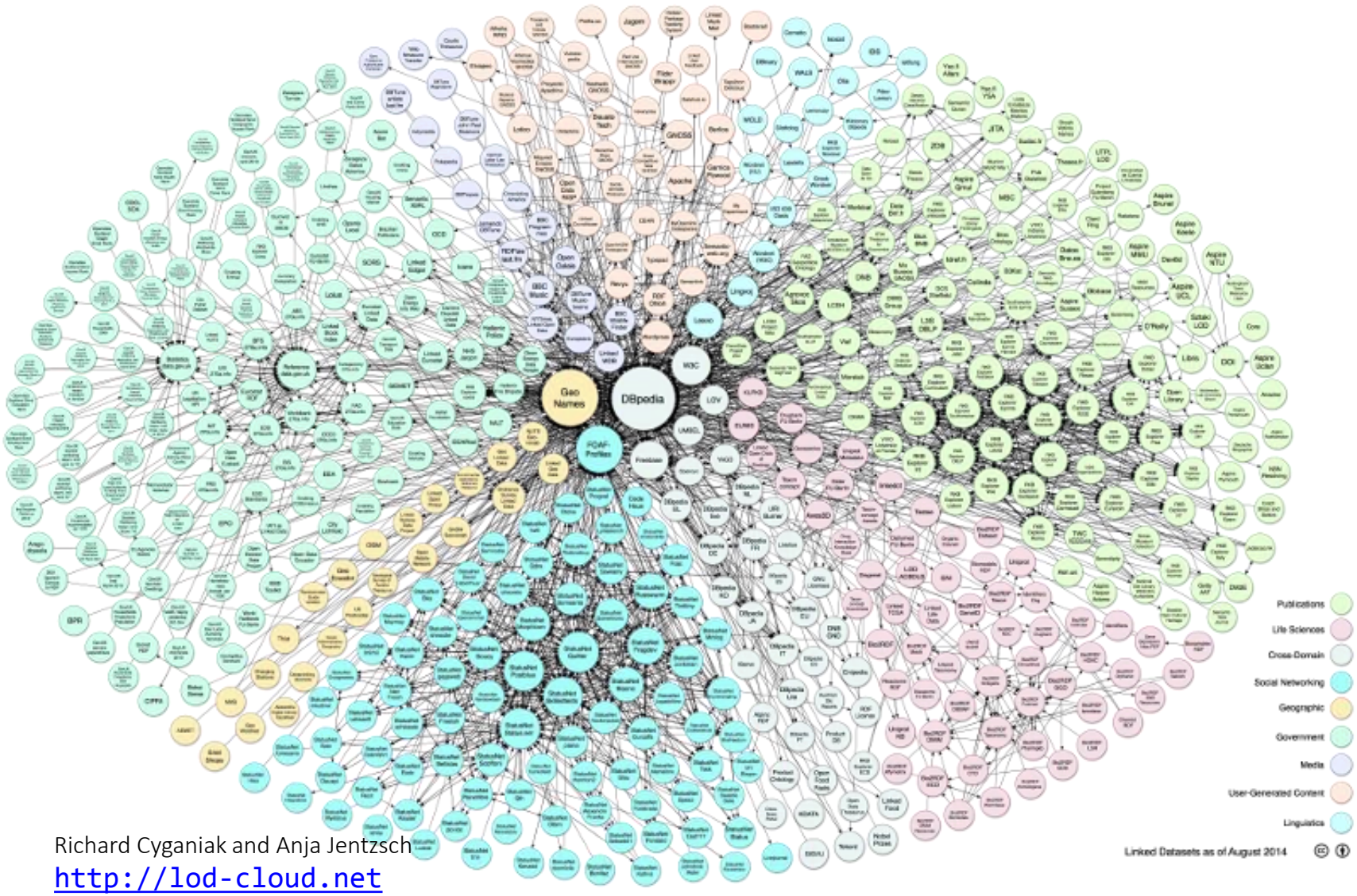
The Open Data Movement aims at making data freely available to everyone. There are already various interesting open data sets available on the Web. Examples include Wikipedia, Wikibooks, Geonames, MusicBrainz, WordNet, the DBLP bibliography and many more which are published under Creative Commons or Talis licenses.

The goal of the W3C SWEO Linking Open Data community project is to extend the Web with a data commons by publishing various open data sets as RDF on the Web and by setting RDF links between data items from different data sources.



Richard Cyganiak and Anja Jentzsch
http://lod-cloud.net

Linked Open Data



Richard Cyganiak and Anja Jentzsch
http://lod-cloud.net

From Open Data to Linked Open Data

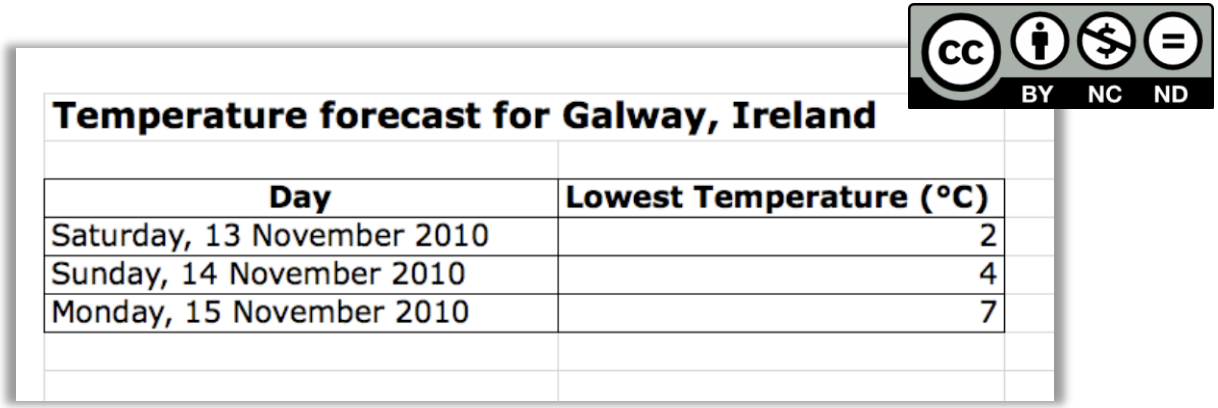
- 2006: defines basic principles for publishing Linked Data
- 2010: added a 5 star rating system for Linked Open Data (LOD)

  "in order to encourage people -- especially government data owners -- along the road to good linked data…"

Tim Berners-Lee
http://www.w3.org/DesignIssues/LinkedData.html

★ Open Data

make your stuff available on the Web (whatever format)
under an open license



Temperature forecast for Galway, Ireland

Day	Lowest Temperature (°C)
Saturday, 13 November 2010	2
Sunday, 14 November 2010	4
Monday, 15 November 2010	7

Michael Hausenblas http://5stardata.info/

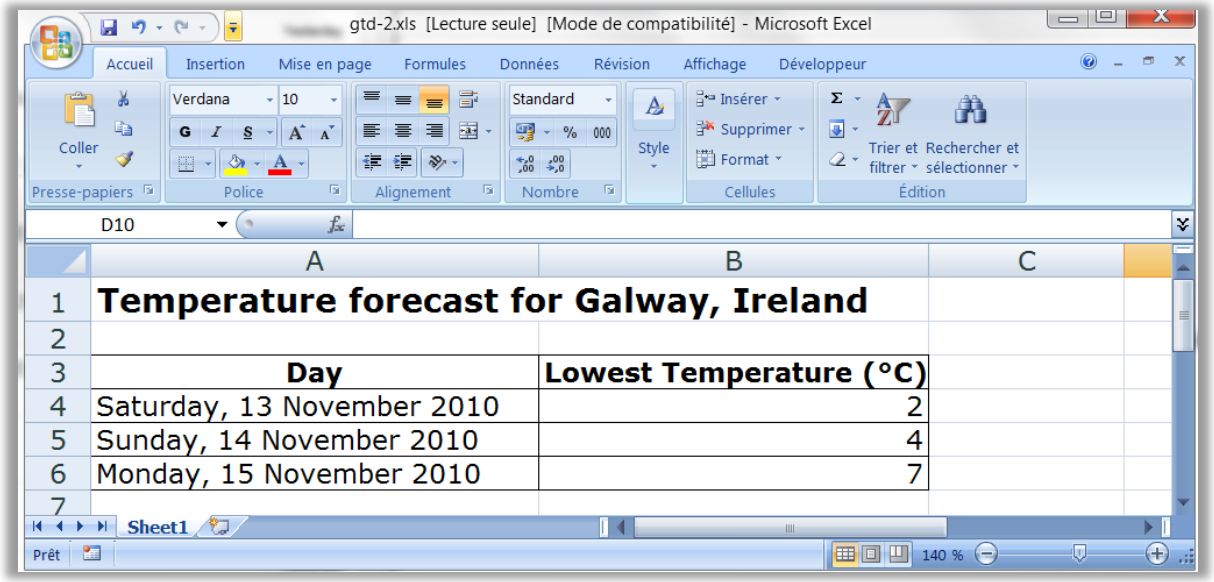
Costs & benefits of ★ Web data

- As a consumer ...
  - ✓ You can look at it.
  - ✓ You can print it.
  - ✓ You can store it locally (on your hard drive or on an USB stick).
  - ✓ You can enter the data into any other system.
  - ✓ You can change the data as you wish.
  - ✓ You can share the data with anyone you like.
- As a publisher ...
  - ✓ It's simple to publish.
  - ✓ You do not have to explain repeatedly to others that they can use your data.

Michael Hausenblas http://5stardata.info/

★★ Open Data

make it available as structured data (e.g., Excel instead of image scan of a table)



Michael Hausenblas http://5stardata.info/

Costs & benefits of ★★ Web data

- As a consumer, you can do all what you can do with ★ Web data and additionally:
  - ✓ You can directly process it with proprietary software to aggregate it, perform calculations, visualise it, etc.
  - ✓ You can export it into another (structured) format.
- As a publisher ...
  - ✓ It's still simple to publish.

★★★ Open Data

use non-proprietary formats (e.g., CSV instead of Excel)

gtd-3.csv

```
"Temperature forecast for Galway, Ireland",

"Day","Lowest Temperature (C)"
"Saturday, 13 November 2010",2
"Sunday, 14 November 2010",4
"Monday, 15 November 2010",7
```

length : 164 lines : 6 Ln : 1 Col : 1 Sel : 0 Macintosh ANSI as UTF-8 INS

Costs & benefits of ★★★ Web data

- As a consumer, you can do all what you can do with ★ ★Web data and additionally:
  - ✓ You can manipulate the data in any way you like, without being confined by the capabilities of any particular software.
- As a publisher ...
  - ⚠ You might need converters or plug-ins to export the data from the proprietary format.
  - ✓ It's still rather simple to publish.

From Open Data to Linked Open Data : COIN web application

WATERLINE Yukon Water Board

geomatics yukon

Environment Canada Canada

Natural Resources Canada

DBpedia

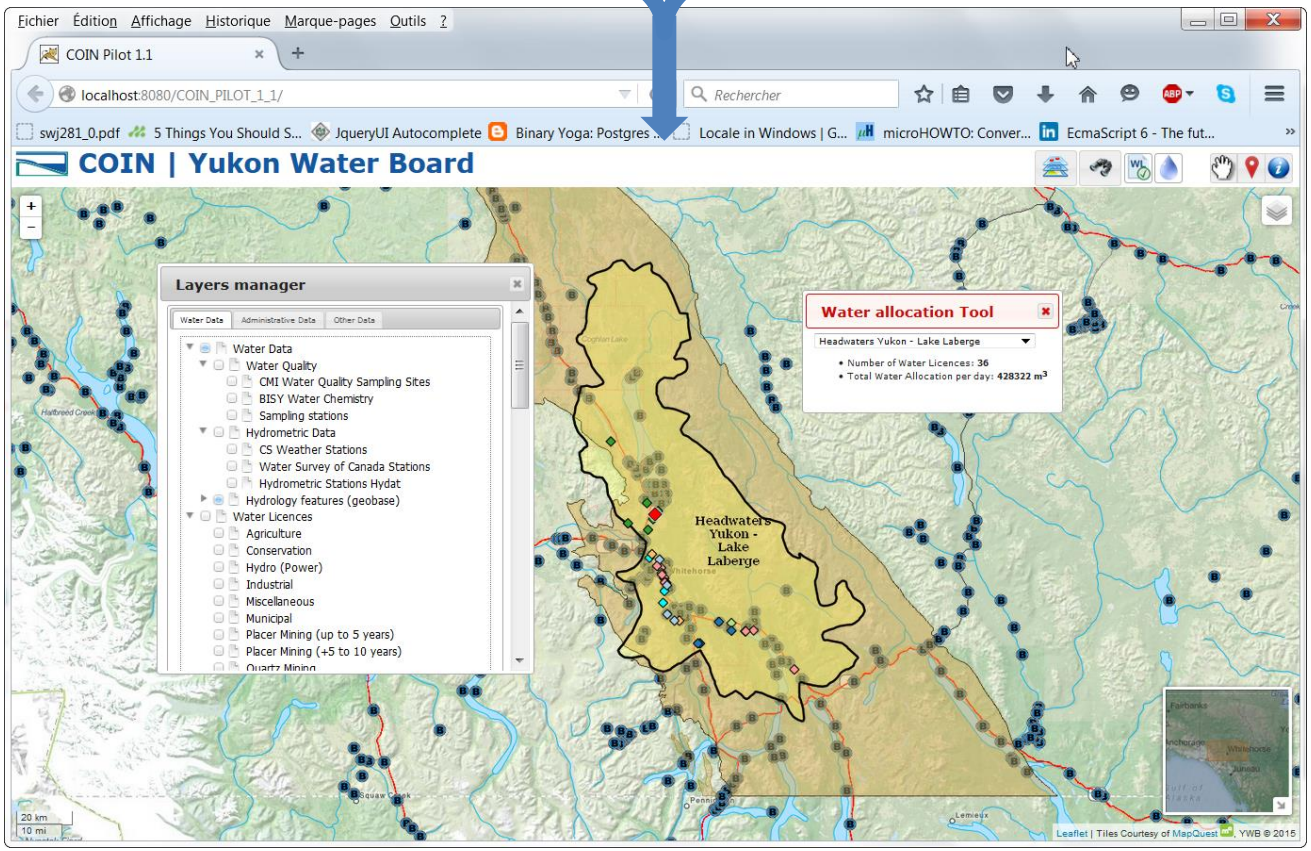
WIKIPEDIA The Free Encyclopedia

HYDAT Data

Bisy Data

Geogratis

Canadian Council of Ministers of the Environment CCME



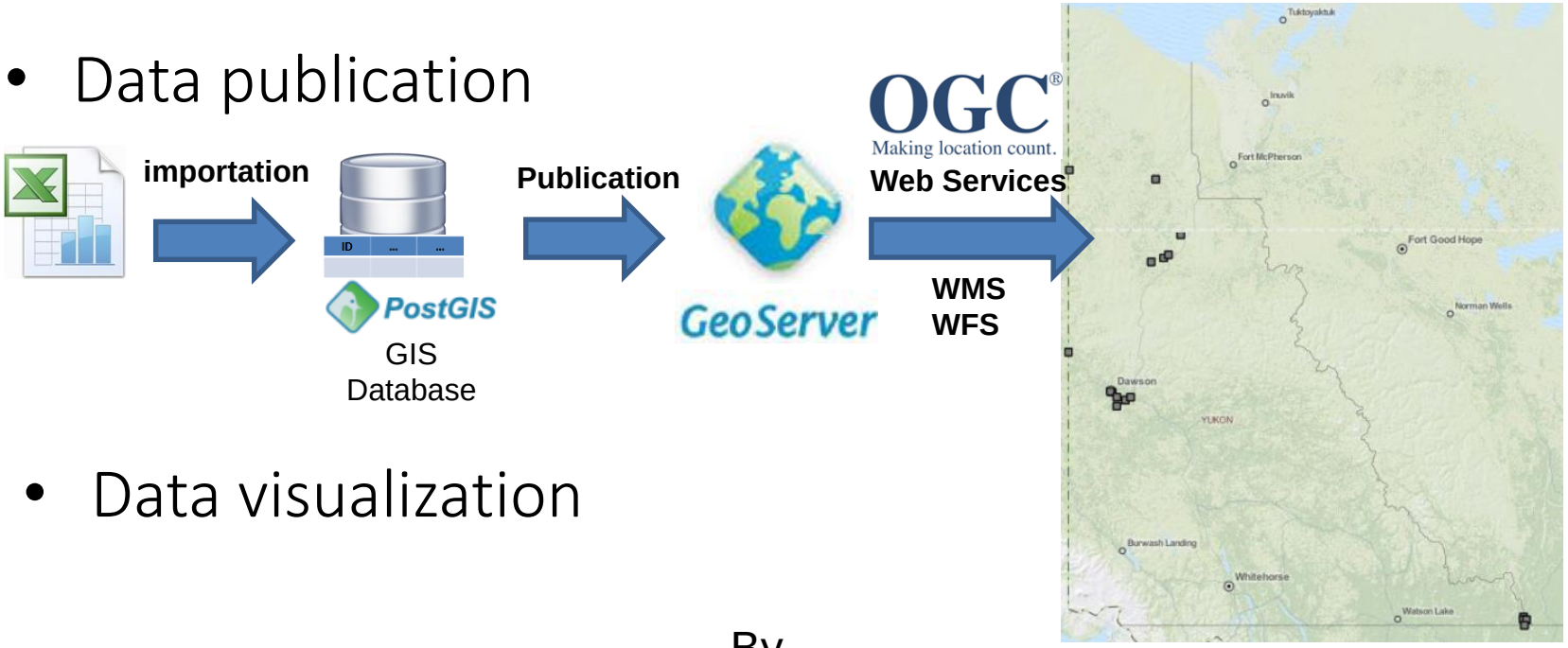
Water Quality Sampling data

- A set of sampling stations
  - For each station a set of measures at different dates

Station Name	Smpl Date	Parameter Name	Param.Description	Param.Code	Param.Class	Result	Units
Site-A	01/01/2013	1,2,6-Trimethylphenanthrene	1,2,6-Trimethylphenanthrene	126-Trimethylphenanthrene	org	<0.0499	ng/L
Site-A	01/01/2013	1,2-Dimethylnaphthalene	1,2-Dimethylnaphthalene	12-Dimethylnaphthalene	org	<0.236	ng/L
Site-A	01/01/2013	1,4,6,7-Tetramethylnaphthalene	1,4,6,7-Tetramethylnaphthalene	1467-Tetramethylnaphthalene	org	<0.158	ng/L
Site-A	01/01/2013	1,7-Dimethylfluorene	1,7-Dimethylfluorene	17-Dimethylfluorene	org	<0.193	ng/L
Site-A	01/01/2013	1,7-Dimethylphenanthrene	1,7-Dimethylphenanthrene	17-Dimethylphenanthrene	org	<0.202	ng/L
Site-A	01/01/2013	1,8-Dimethylphenanthrene	1,8-Dimethylphenanthrene	18-Dimethylphenanthrene	org	<0.203	ng/L
Site-A	01/01/2013	1-Methylchrysene	1-Methylchrysene	1-Methylchrysene	org	<0.0486	ng/L
Site-A	01/01/2013	1-Methylnaphthalene	1-Methylnaphthalene	1-Methylnaphthalene	org	0,646	ng/L
Site-A	01/01/2013	1-Methylphenanthrene	1-Methylphenanthrene	1-Methylphenanthrene	org	<0.194	ng/L
Site-A	01/01/2013	2,3,5-Trimethylnaphthalene	2,3,5-Trimethylnaphthalene	235-Trimethylnaphthalene	org	NDR 0.194	ng/L
Site-A	01/01/2013	2,3,6-Trimethylnaphthalene	2,3,6-Trimethylnaphthalene	236-Trimethylnaphthalene	org	NDR 0.193	ng/L
Site-A	01/01/2013	2,4-Dimethyldibenzothiophene	2,4-Dimethyldibenzothiophene	24-Dimethyldibenzothiophene	org	<0.0507	ng/L
Site-A	01/01/2013	2,6-Dimethylnaphthalene	2,6-Dimethylnaphthalene	26-Dimethylnaphthalene	org	<0.192	ng/L
Site-A	01/01/2013	2,6-Dimethylphenanthrene	2,6-Dimethylphenanthrene	26-Dimethylphenanthrene	org	<0.203	ng/L
Site-A	01/01/2013	2-Methylanthracene	2-Methylanthracene	2-Methylanthracene	org	<0.203	ng/L
Site-A	01/01/2013	2-Methylfluorene	2-Methylfluorene	2-Methylfluorene	org	<0.347	ng/L
Site-A	01/01/2013	2-Methylnaphthalene	2-Methylnaphthalene	2-Methylnaphthalene	org	1,07	ng/L
Site-A	01/01/2013	2-Methylphenanthrene	2-Methylphenanthrene	2-Methylphenanthrene	org	<0.193	ng/L
Site-A	01/01/2013	2/3-Methyldibenzothiophenes	2/3-Methyldibenzothiophenes	23-Methyldibenzothiophenes	org	<0.163	ng/L
Site-A	01/01/2013	3,6-Dimethylphenanthrene	3,6-Dimethylphenanthrene	36-Dimethylphenanthrene	org	<0.205	ng/L
Site-A	01/01/2013	3-Methylfluoranthene/Benzo[a]f	3-Methylfluoranthene/Benzo[a]fluore	3-Methylfluoranthene-Benzo-a-fluore	org	NDR 0.238	ng/L
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

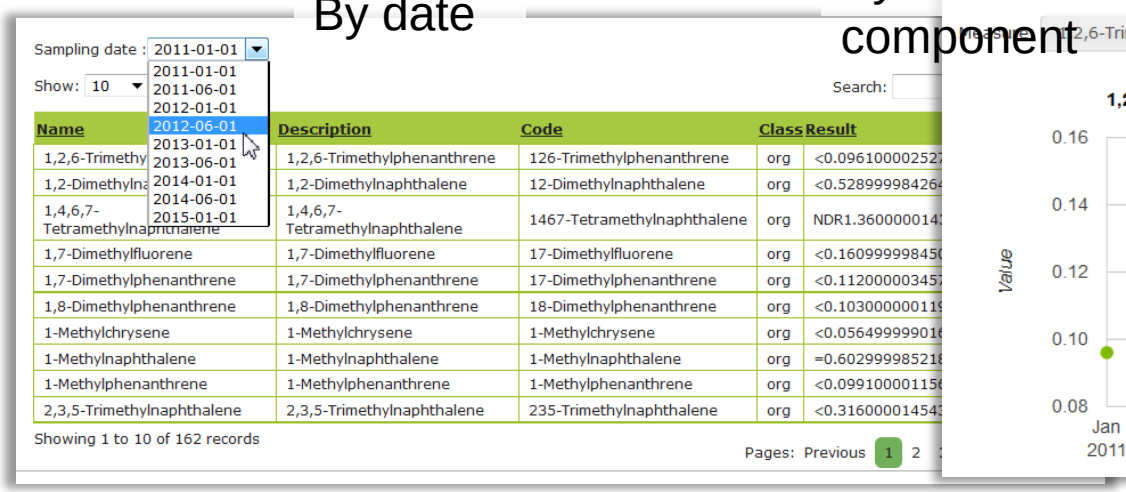
Water Quality Sampling data

- Data publication

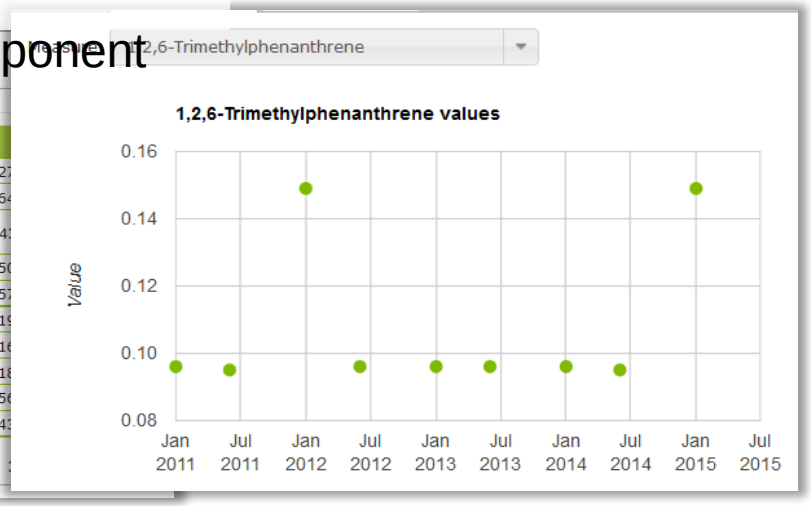


- Data visualization

By date

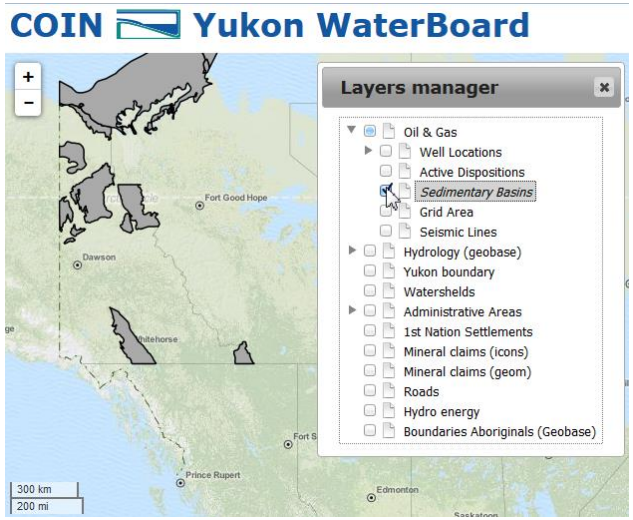


By component



Web Map Service (WMS)

HTTP client (web browser)



HTTP server

HTTP request to the WMS service

url:
`http://lig-coin.imag.fr/geoserver/coin_ws/wms`
Query String:
`SERVICE=WMS`
`REQUEST=GetMap`
`VERSION=1.1.1`
`LAYERS=yukon_oilgas_sedimentary_basins`
`STYLES=`
`FORMAT=image/png`
`TRANSPARENT=true`
`SRS=EPSG:3857`
`HEIGHT=256`
`WIDTH=256`
`BBOX=...`

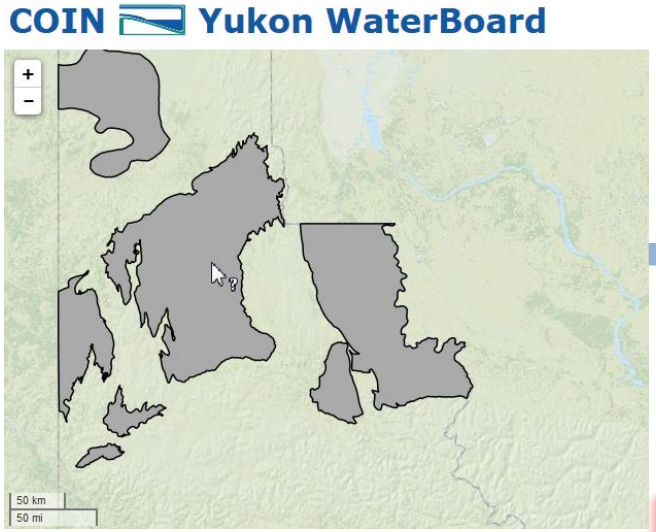
other requests with bounding box fixed for each tile

HTTP response

png image with transparency for each tile

GeoServer

Web Feature Service



HTTP server

HTTP request to the WFS service

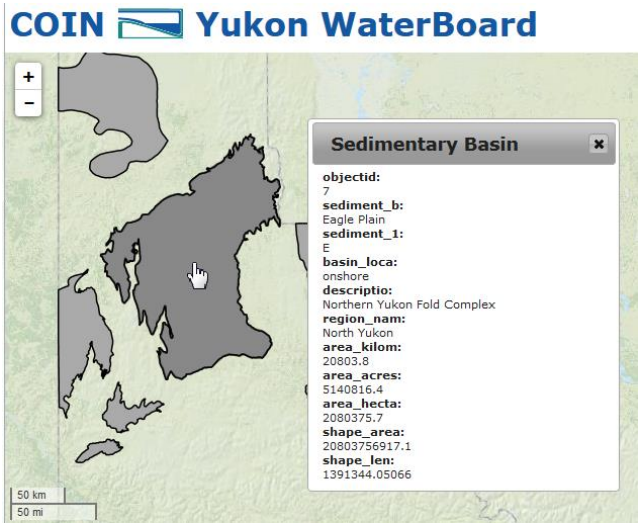
url:
`http://lig-coin.imag.fr/geoserver/coin_ws/wfs`

Query String:

```
version=2.0.0
request=GetFeature
typeName=yukon_oilgas_sedimentary_basins
outputFormat=application/json
cql_filter=CONTAINS(geom, POINT(66.58321725728175 -138.14208984375))
```

GeoServer

HTTP response

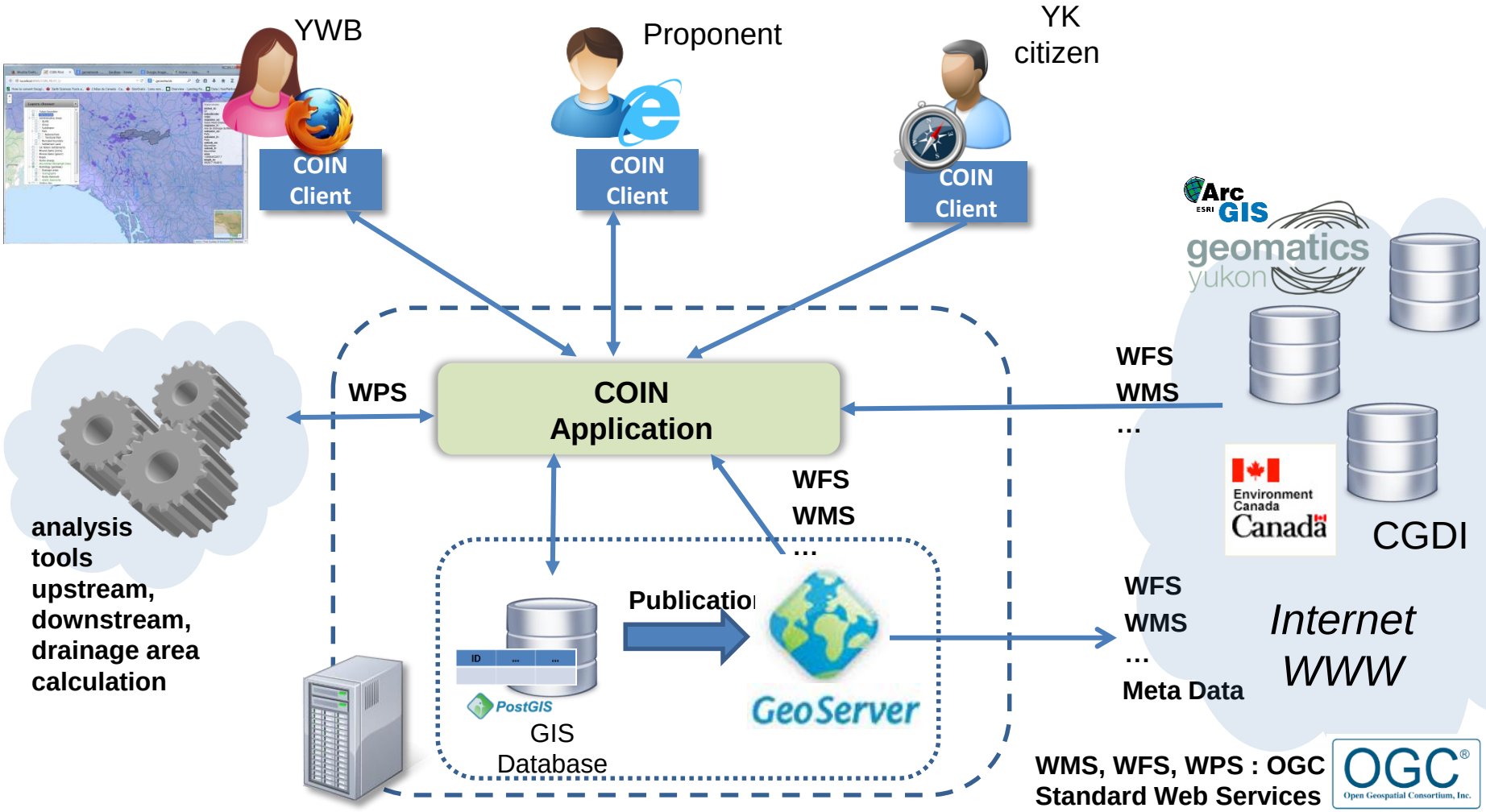



```
{
  "type":"FeatureCollection",
   "totalFeatures":1,
   "features":[{"type":"Feature",
   "id":"yukon_oilgas_sedimentary_basins.7",
    "geometry": {
        "type":"MultiPolygon",
        "coordinates":[[[
          [66.58613633961849,-139.99753640064932],
          [66.60835459019349,-139.97916364962316],
        ]]]
    }
}
```


geoJSON

COIN: Coordinated Online Information Network

natural resource (water) licensing and allocation procedures



From Open Data to Linked Open Data

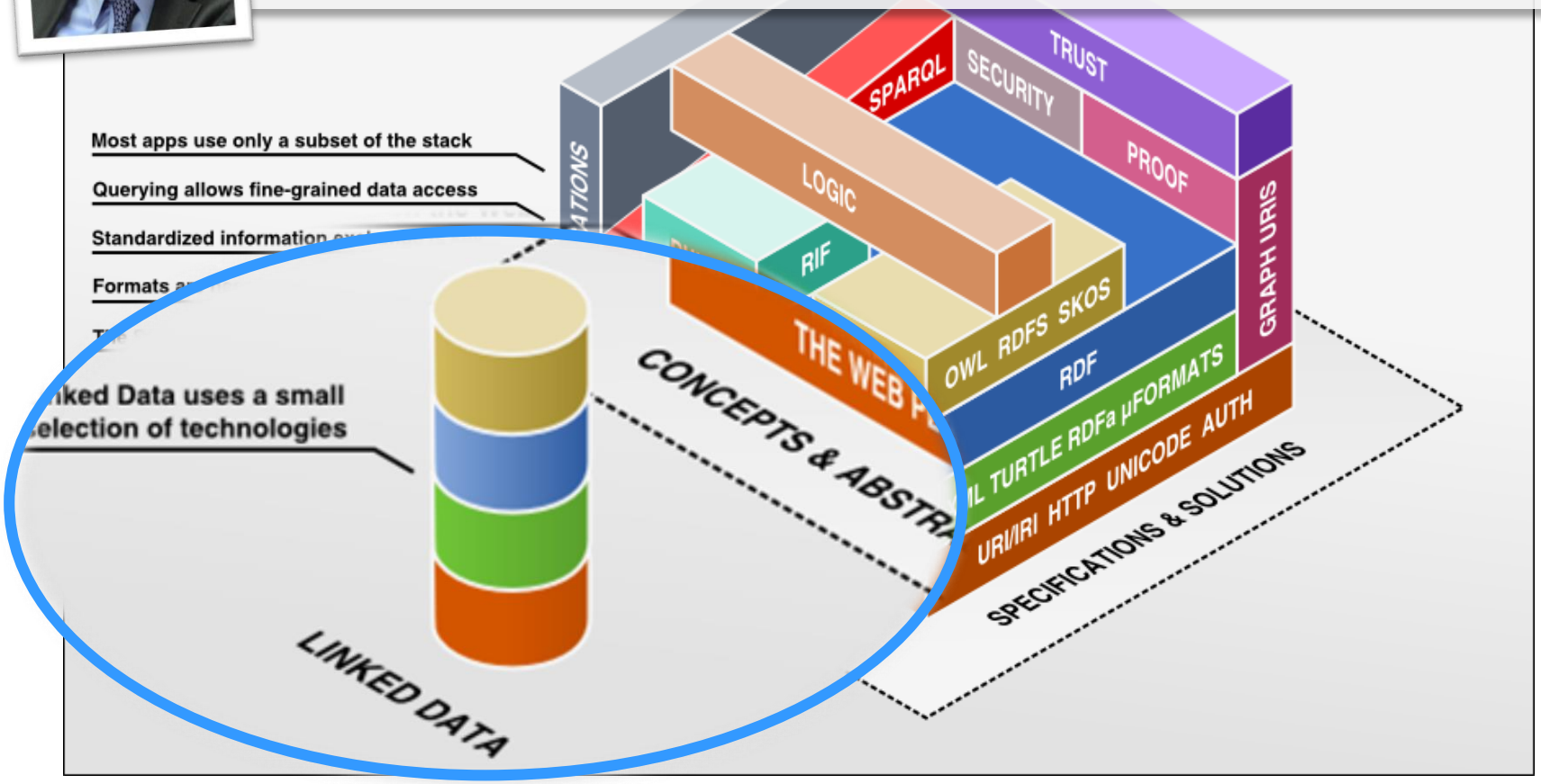
- 2006: defines basic principles for publishing Linked Data
- 2010: added a 5 star rating system for Linked Open Data (LOD)

  "in order to encourage people -- especially government data owners -- along the road to good linked data…"

Tim Berners-Lee
http://www.w3.org/DesignIssues/LinkedData.html

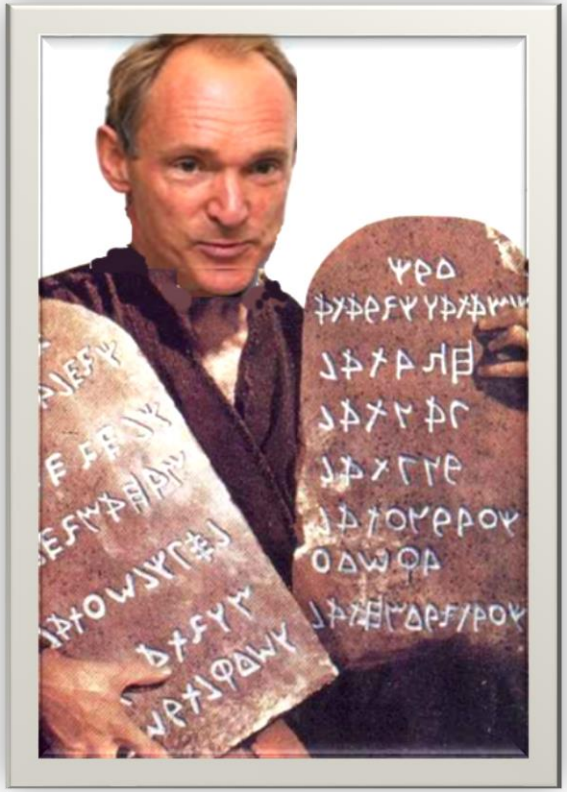
Linked Data

"The Semantic Web isn't just about putting data on the web. **It is about making links, so that a person or machine can explore the web of data**. With **linked data**, when you have some of it, you can find other, related, data."
Tim Berners-Lee - 2006 http://www.w3.org/DesignIssues/LinkedData.html



http://www.bnode.org/blog/tag/layer%20cake (Benjamin Nowack)

Linked Data Principles



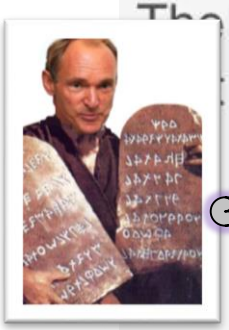
Tim Berners-Lee
http://www.w3.org/DesignIssues/LinkedData.html

a set of best practices for publishing and interlinking structured data on the Web

Basic idea: to apply the general architecture of the World Wide Web to the task of sharing structured data on global scale.

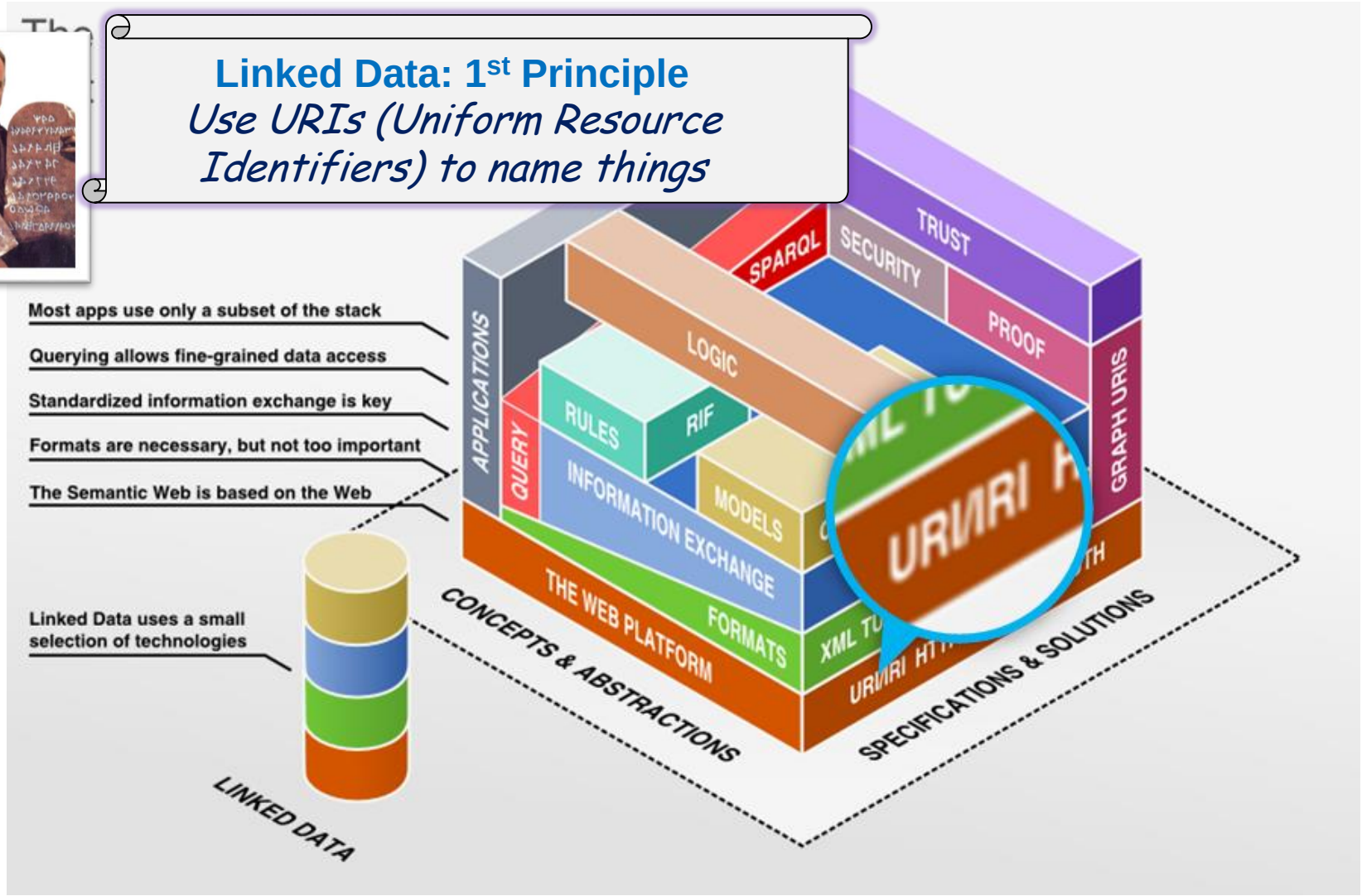
1. Use **URIs** as names for things.
2. Use **HTTP URIs**, so that people can look up those names.
3. When someone looks up a URI, **provide useful information, using the standards** (RDF, SPARQL).
4. Include **links** to other URIs, so that they **can discover more things**.

Uniform Resource Identifiers (URIs)



Linked Data: 1st Principle

Use URIs (Uniform Resource Identifiers) to name things



http://www.bnode.org/blog/tag/layer%20cake (Benjamin Nowack)

URI: syntax

- URI : generic syntax

scheme ":" ["//" authority "/"] [path] ["?" query] ["#" fragment]

- **scheme:** http, ftp, mailto, …
- **authority:** [userinfo@]host[:port]
  - **userinfo:** authentication section e.g: username:password
  - **host:** domain name, IP address
  - **port:** port number, ex: 80 for HTTP standard port
- **path:** a sequence of segments separated by slashes, e.g. : a path in the hierarchical file system of the HTTP server.
- **query:** a query string of non-hierarchical data. (e.g: a sequence of attribute–value pairs separated by a delimiter (&) for HTTP requests)
- **fragment:** a fragment identifier providing direction to a secondary resource (e.g.: anchor id in a HTML document)

Internationalized Resource Identifier (RFC 3987): extension to support Universal Character Set (Unicode/ISO 10646))

http://fa.dbpedia.org/resource/ژرژ_برسنس

URI: examples

- URI : generic syntax

scheme ":" ["//" authority "/"] [path] ["?" query] ["#" fragment]

example :

http://myserver.mydomain.org:8080/people/ziebelin/test.html#chapitre1

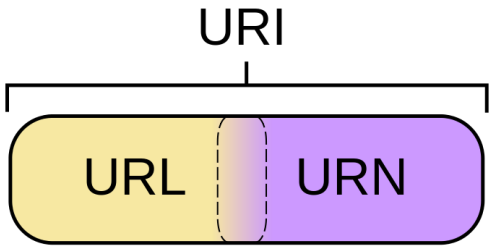
- http: — Used for HTTP protocol
- myserver.mydomain.org — Identification of the web server
- 8080 — IP port number
- /people/ziebelin/test.html — path to the HTML document `test.html`
- #chapitre1 — position inside the document

other URIs examples:

ftp://server.example.com/foo
mailto:person@example.fr
urn:isbn:978-0553283686

URI - URL - URN

- an URI/IRI **doesn't necessarily** identifies a resource that is resolvable on the web



- Address (Locator)
  - **U**niform **R**esource **L**ocator (RFC 1738)
  - Tells where and how a resource can be found in the internet
  - *Can change during the life cycle of a resource*

- Identity (Name)
  - **U**niform **R**esource **N**ame (RFC 2141)
  - identifies a resource by name in a particular namespace. A URN can be used to talk about a resource without implying its location or how to access it.
  - *Remains unchanged during life cycle of the resource*

URIs, URLs, and URNs: Clarifications and Recommendations 1.0
Report from the joint W3C/IETF URI Planning Interest Group- W3C Note 21 September 2001
http://www.w3.org/TR/uri-clarification/

Resources in the Web of data

- Data describe elements (things) for a domain of interest through their properties and relationships.
- Theses elements can be:

RESOURCES

URIs

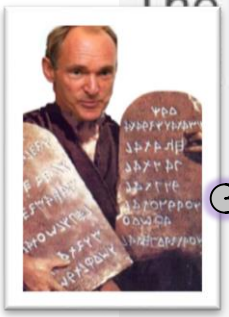
URLs

URNs

  - documents
    - HTML pages, images, data accessed through a web service…
  - but also:
    - real world entities:
      - place: **Dawnson City**
      - a chemical substance: **Dimethyl sulfate**
      - …
    - abstracts concepts:
      - **Set of Carcinogens substances,**
      - **Year Mean Temperature** relationship
      - …

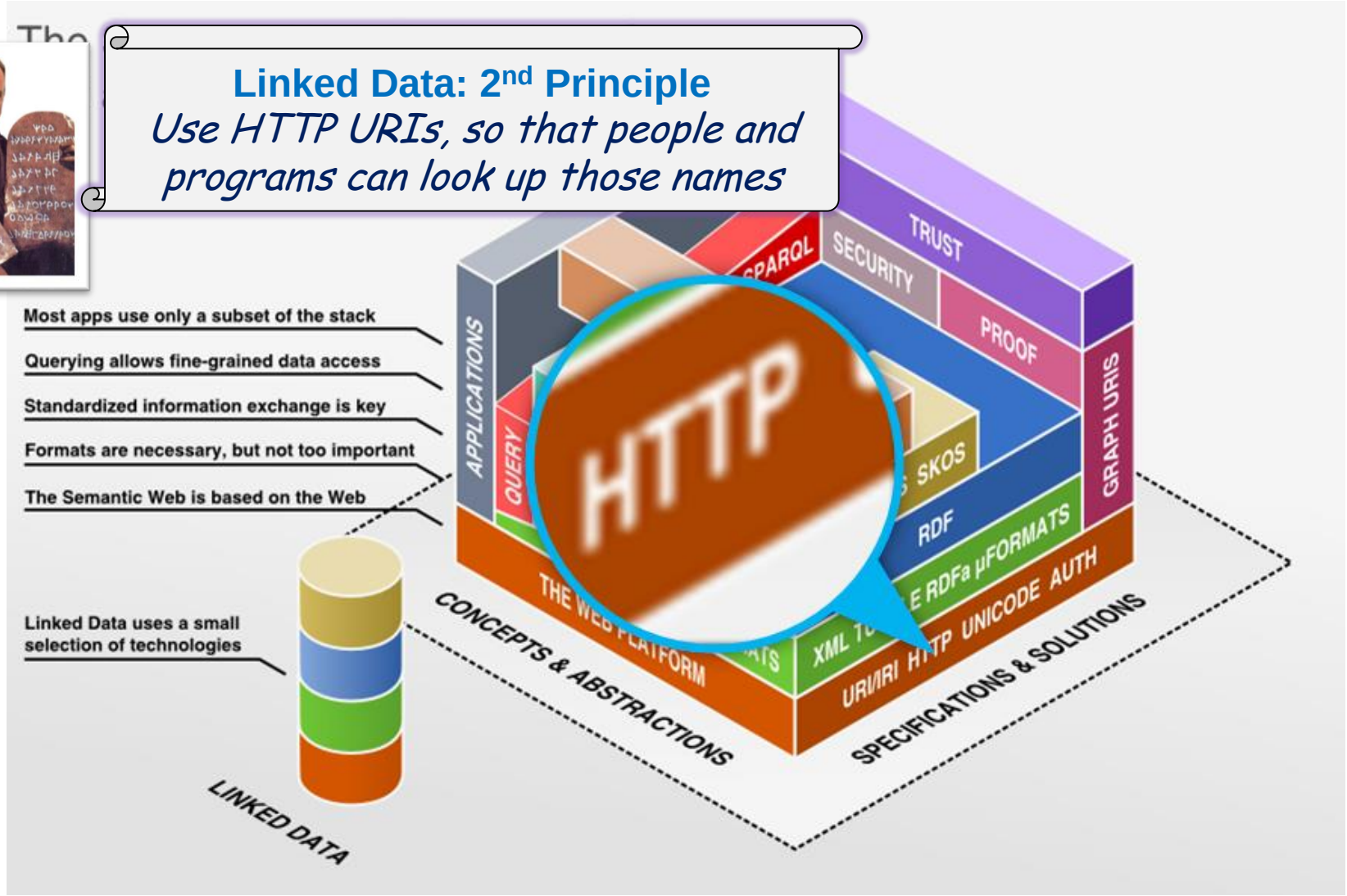
- All these elements are resources identified by an URI

HTTP URIs



Linked Data: 2nd Principle

Use HTTP URIs, so that people and programs can look up those names



http://www.bnode.org/blog/tag/layer%20cake (Benjamin Nowack)

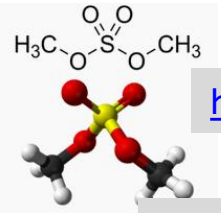
HTTP URIs

- HTTP (Hyper Text Transfer Protocol) protocol is the Web's universal access mechanism.
- HTTP URIs make good names for two reasons*:
  - They provide a simple way to create globally unique names in a decentralized fashion, as every owner of a domain name, or delegate of the domain name owner, may create new URI references.
  - They serve not just as a name but also as a means of accessing information describing the identified
    - HTTP clients can **dereference** (i.e., look up) the URI using the HTTP protocol and retrieve a description of the resource that is identified by the URI.

* Tom Heath and Christian Bizer (2011)
Linked Data: Evolving the Web into a Global Data Space (1st edition).
Synthesis Lectures on the Semantic Web: Theory and Technology, 1:1, 1-136. Morgan & Claypool.
http://linkeddatabook.com/editions/1.0/

HTTP URIs

- Where HTTP URIs identify real-world objects or abstract concepts, it is essential to **not confuse** the objects or concepts themselves with the Web documents that describe them.
  - A real world object or abstract concept can have different representations
  - It allows separate statements to be made about an object and about a document that describes that object.



URI of DBpedia resource representing Dimethyl sulfate

http://dbpedia.org/**resource**/Dimethyl_sulfate

a URN **but not** a URL

http://dbpedia.org/**page**/Dimethyl_sulfate

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML+RDFa
<html xmlns="http://www.w3.org/1999/xhtml"
    xmlns:dbpprop="http://dbpedia.org/property
    xmlns:foaf="http://xmlns.com/foaf/0.1/"
    version="XHTML+RDFa 1.0" xml:lang="en">

<!-- header -->
<head profile="http://www.w3.org/1999/xhtml/v
    <title>About: Georges Brassens</title>
    <link rel="alternate" type="application/rdf+xml" href="http://dbpe
    <link rel="alternate" type="text/rdf+n3" href="http://dbpedia.org/
    <link rel="alternate" type="application/json+rdf" href="http://dbp
    <link rel="alternate" type="application/json" href="http://dbpedia
    <link rel="alternate" type="application/atom+xml" href="http://dbpe
    <link rel="alternate" type="text/plain" href="http://dbpedia.org/da
    <link rel="alternate" href="http://dbpedia.org/sparql?default-graph

    <link rel="alternate" href="http://dbpedia.org/sparql?default-graph

    <link rel="alternate" href="http://dbpedia.org/sparql?default-graph
    <link rel="alternate" href="http://dbpedia.org/sparql?default-graph
    <link rel="alternate" href="http://dbpedia.org/sparql?default-graph
    <link rel="timegate" type="text/html" href="http://mementoarchive.la
    <link rel="stylesheet" type="text/css" href="/statics/style.css" />
```

URL of HTML Page

Description intended to be read by humans

http://dbpedia.org/**data**/Dimethyl_sulfate

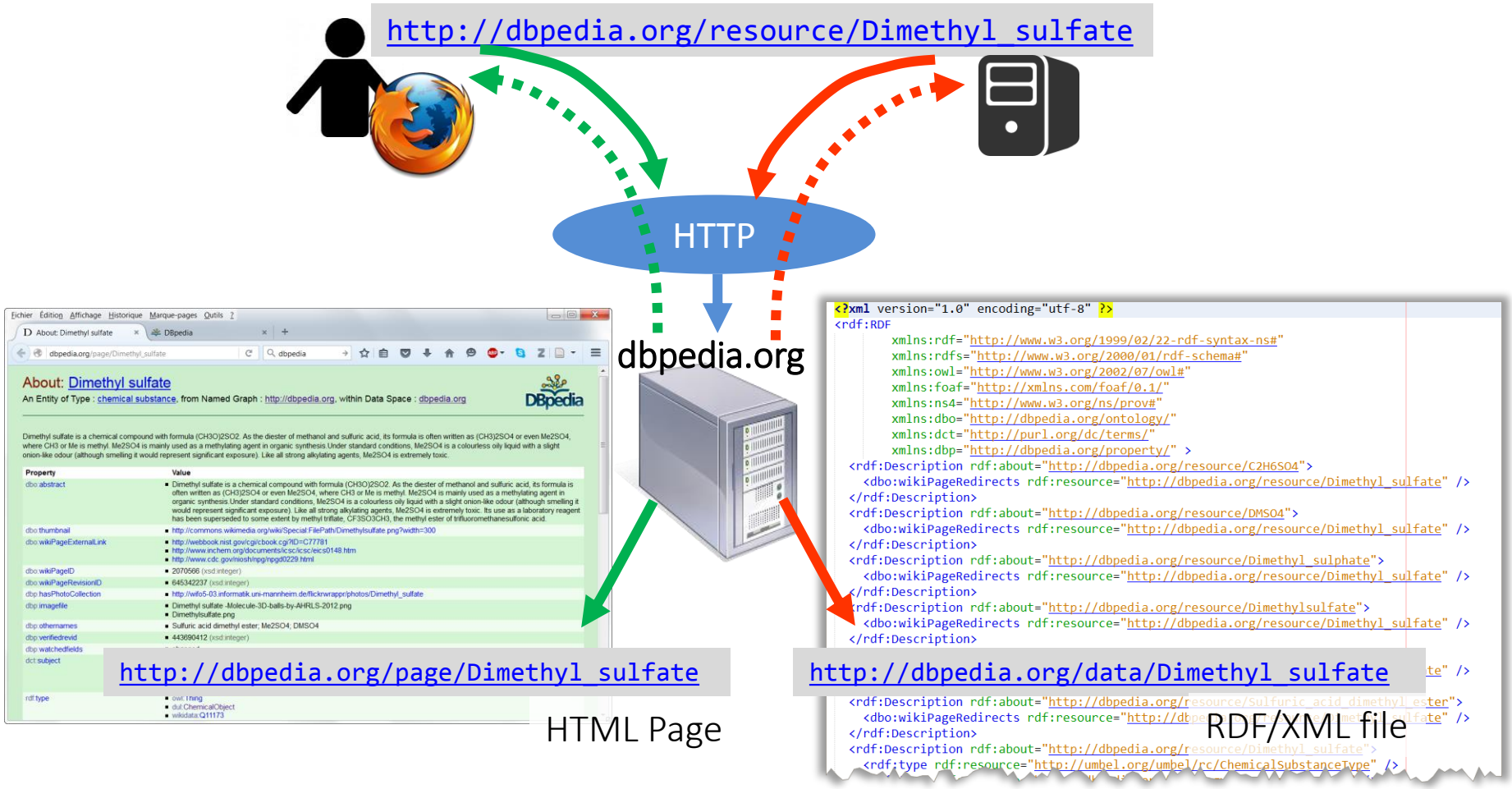
```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF xmlns:dct="http://purl.org/dc/terms/" xmlns:d
xmlns:ns6="http://www.w3.org/ns/prov#" xmlns:dc="h
xmlns:dbp="http://dbpedia.org/property/" xmlns:foaf="
xmlns:owl="http://www.w3.org/2002/07/owl#" xmlns:
schema#" xmlns:rdf="http://www.w3.org/1999/02/22
 - <rdf:Description rdf:about="http://dbpedia.org/resou
      <rdf:type rdf:resource="http://dbpedia.org/class/yago/CausalAgent100007347"/>
      <rdf:type rdf:resource="http://dbpedia.org/class/yago/YagoLegalActor"/>
      <rdf:type rdf:resource="http://dbpedia.org/class/yago/Entertainer109616922"/>
      <rdf:type
          rdf:resource="http://www.ontologydesignpatterns.org/ont/dul/DUL.owl#NaturalPerson"/>
      <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Person"/>
      <rdf:type rdf:resource="http://umbel.org/umbel/rc/MusicalPerformer"/>
      <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#Thing"/>
      <rdf:type rdf:resource="http://dbpedia.org/class/yago/FrenchPoets"/>
      <rdf:type rdf:resource="http://dbpedia.org/class/yago/PeopleFromS%C3%A8te"/>
      <rdf:type rdf:resource="http://dbpedia.org/class/yago/Radical110503452"/>
      <rdf:type rdf:resource="http://www.wikidata.org/entity/Q215627"/>
      <rdf:type rdf:resource="http://dbpedia.org/ontology/Artist"/>
      <rdf:type rdf:resource="http://dbpedia.org/class/yago/FrenchMaleSingers"/>
      <rdf:type rdf:resource="http://schema.org/Person"/>
      <rdf:type rdf:resource="http://schema.org/MusicGroup"/>
      <rdf:type rdf:resource="http://www.ontologydesignpatterns.org/ont/dul/DUL.owl#Agent"/>
      <rdf:type rdf:resource="http://umbel.org/umbel/rc/Artist"/>
```

URL of RDF/XML document

Description intended for consumption by machines

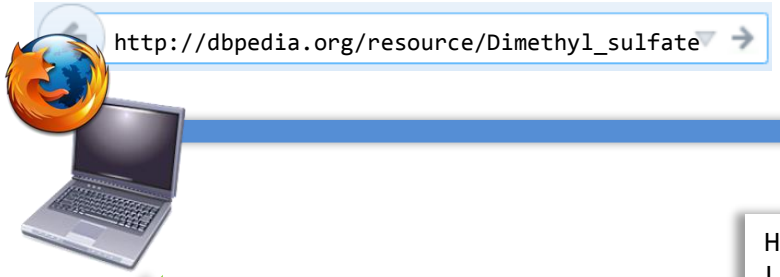
Making URIs Dereferenceable
Content negotiation (303 URIs)

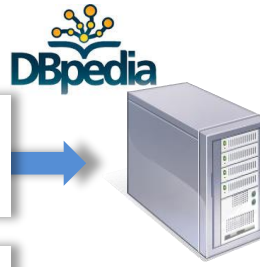
- The same URI can be used to retrieve different representations.



Making URIs Dereferenceable
Content negotiation (303 URIs)

- Contents negotiation uses HTTP headers to a retrieve the resource description

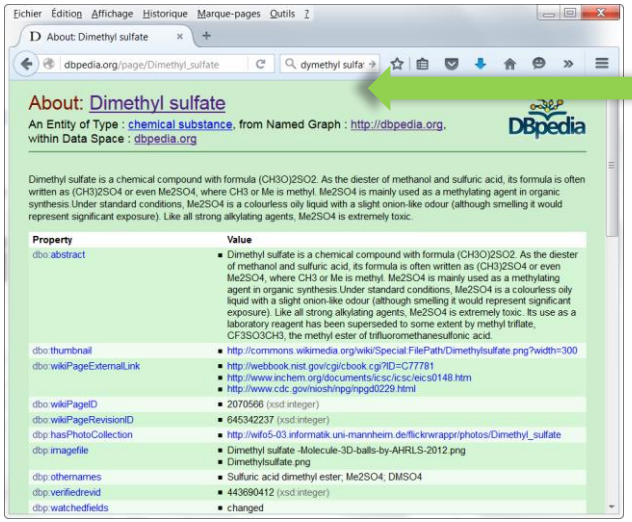





```
GET /resource/Dimethyl_sulfate HTTP/1.1
Host: dbpedia.org
Accept: text/html
```

```
HTTP/1.1 303 See Other
Location: http://dbpedia.org/page/Dimethyl_sulfate
Vary: Accept
```

```
GET /page/Dimethyl_sulfate HTTP/1.1
Host: dbpedia.org
Accept: text/html
```

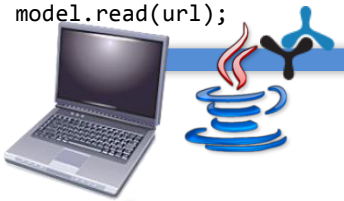


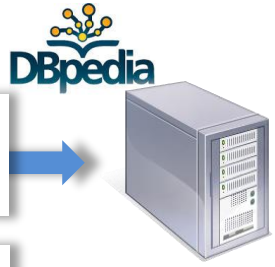

```
HTTP/1.1 200 OK
Content-Type: text/html

<?xml version="1.0" encoding="UTF-8" ?>
<!DOCTYPE html>
   <head>
       <title>About: Dimethyl sulfate</title>
      ...
   </head>
   <body onload="init();">
      <div id="header">
      <h1 id="title">About: Dimethyl sulfate</h1>
       ...
```

Making URIs Dereferenceable
Content negotiation (303 URIs)

- Contents negotiation uses HTTP headers to a retrieve the resource description

```
final String url ="http://dbpedia.org/resource/Dimethyl_sulfate ";
final Model model = ModelFactory.createDefaultModel();
model.read(url);
```






```
GET /resource/Dimethyl_sulfate HTTP/1.1
Host: dbpedia.org
Accept: application/rdf+xml
```

```
HTTP/1.1 303 See Other
Location: http://dbpedia.org/data/Dimethyl_sulfate
Vary: Accept
```

```
GET /data/Dimethyl_sulfate HTTP/1.1
Host: dbpedia.org
Accept: text/html
```

```
HTTP/1.1 200 OK
Content-Type: application/rdf+xml

<?xml version="1.0" encoding="utf-8" ?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"...>
      <rdf:Description rdf:about="http://dbpedia.org/resource/C2H6SO4">
         <dbo:wikiPageRedirects
                 rdf:resource="http://dbpedia.org/resource/Dimethyl_sulfate" />
  </rdf:Description>
   ...
  <rdf:type rdf:resource="http://dbpedia.org/ontology/ChemicalSubstance" />
  <rdf:type rdf:resource="http://dbpedia.org/class/yago/IARCGroup2ACarcinogens" />
  ...
```

Making URIs Dereferenceable
Hash URIs

- An other way to identify real world objects or abstract concepts without creating ambiguity with the document that contains it 's description is to use hash URIs.

Example of a hash URI used by Dbpedia RDF description of Georges Brassens

http://dbpedia.org/data/Georges_Brassens.xml

```
HTTP/1.1 200 OK
Content-Type: application/rdf+xml

<?xml version="1.0" encoding="utf-8" ?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"...>
     <rdf:Description rdf:about="http://dbpedia.org/resource/Dimethyl_sulfate">
     <rdf:type rdf:resource="http://dbpedia.org/class/yago/PeopleFromS%C3%A8te" />
     <rdf:type
        rdf:resource="http://dbpedia.org/class/yago/IARCGroup2ACarcinogens" />
```

Dimethyl Sulfate | is a | IARCGroup2ACarcinogens

term from the RDF vocabulary
to describe the type of a resource

Hash URI `http://www.w3.org/1999/02/22-rdf-syntax-ns#type`

base | fragment identifier

Making URIs Dereferenceable
Hash URIs

When a client wants to retrieve a hash URI, the HTTP protocol requires the fragment part to be stripped off before requesting the URI from the server

http://www.w3.org/1999/02/22-rdf-syntax-ns#type

URL the RDF/XML document containing a description of the whole RDF vocabulary

```
GET /1999/02/22-rdf-syntax-ns
Host: www.w3.org
Accept: application/rdf+xml
```

```
HTTP/1.1 200 OK
Content-Type: application/rdf+xml

<?xml version="1.0"?>
<rdf:RDF   xmlns:rdf=http://www.w3.org/1999/02/22-rdf-syntax-ns#
   xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
   xmlns:owl="http://www.w3.org/2002/07/owl#"
   xmlns:dc="http://purl.org/dc/elements/1.1/">

...
<rdf:Property rdf:about="http://www.w3.org/1999/02/22-rdf-syntax-ns#type">
  <rdfs:isDefinedBy rdf:resource="http://www.w3.org/1999/02/22-rdf-syntax-ns#"/>
  <rdfs:label>type</rdfs:label>
  <rdfs:comment>The subject is an instance of a class.</rdfs:comment>
  <rdfs:range rdf:resource="http://www.w3.org/2000/01/rdf-schema#Class"/>
  <rdfs:domain rdf:resource="http://www.w3.org/2000/01/rdfschema#Resource"/>
</rdf:Property>
...
<rdfs:Class rdf:about="http://www.w3.org/1999/02/22-rdf-syntax-ns#Bag">
     <rdfs:isDefinedBy rdf:resource="http://www.w3.org/1999/02/22-rdf-syntax-ns#"/>
     <rdfs:label>Bag</rdfs:label>
     <rdfs:comment>The class of unordered containers.</rdfs:comment>
```

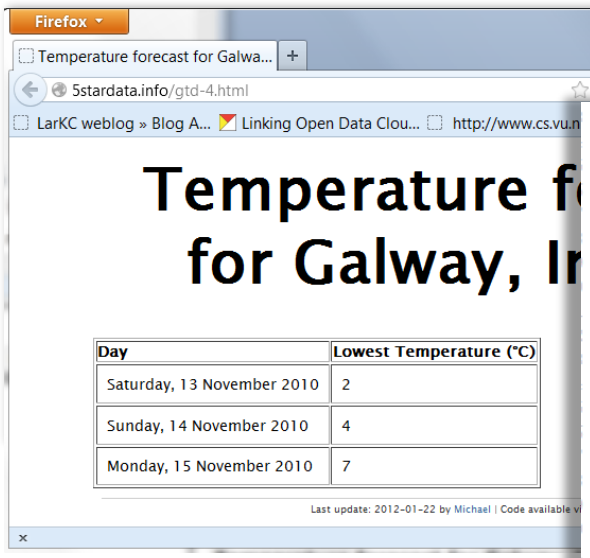
Client is in charge of extracting the information associated to the fragment

★★★★ Linked data

1. Use URIs as names for things.
2. Use HTTP URIs, so that people can look up those names.
3. When someone looks up a URI, **provide useful information, using the standards**(RDF, SPARQL).

RDFa = RDF in attributes

- to add metadata (RDF annotations) in (X)HTML documents
- use existing attributes (e.g. `href`, `src`) et introduce new ones (`vocab`, `typeof`, `property`, `resource`, `prefix`)




```
<h1 property="dcterms:title">Temperature forecast for Galway, Ireland</h1>

<div id="data" about="#Galway" typeof="meteo:Place">
<table border="1px">
    <tr>
        <th>Day</th>
        <th>Lowest Temperature (&deg;C)</th>
    </tr>
    <tr rel="meteo:forecast" resource="#forecast20101113">
        <td>
            <div about="#forecast20101113">
                <span property="meteo:predicted" content="2010-11-13T00:00:00Z" datatype="xsd:dateTime">Saturda
            </div>
        </td>
        <td rel="meteo:temperature">
            <div about="#temp20101113">
                <span property="meteo:celsius" datatype="xsd:decimal">2</span>
            </div>
        </td>
    </tr>
```

W3C XHTML + RDFa

Michael Hausenblas http://5stardata.info/

RDF Data Model

- With RDF, knowledge is represented by a set of assertions (statements)
- All RDF statements follow a simple structure composed of three parts :
  - the thing the statement describes
  - the properties of the thing the statement describes
  - the values of those properties the statement describe

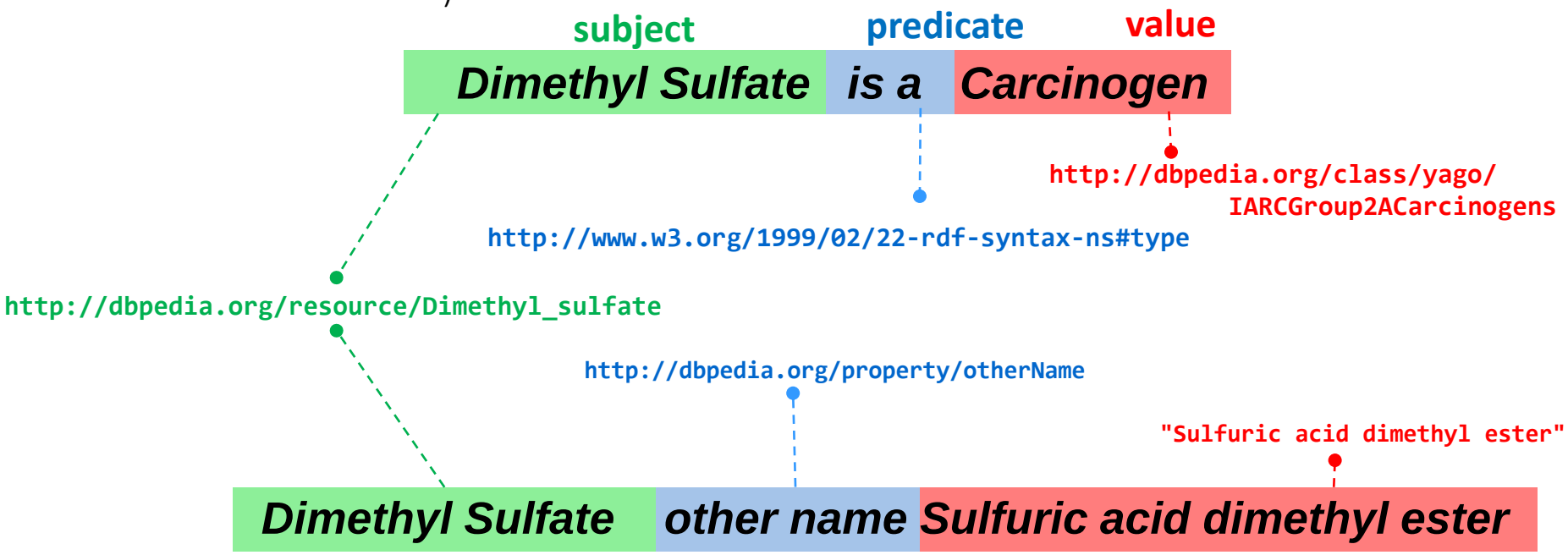
the thing described	property	value
Dimethyl Sulfate	***is a***	***Carcinogen***

RDF Data Model

- RDF Statements are *triples*

Subject	Predicate	Object
URI	URI	URI/Literal

- the subject and the predicate are resources : RDF uses **URIs** (Universal Resource Identifiers) for **uniquely identifying** them
- object can be a **resource** (URI) or a **literal** (constants that don't have other attributes that describe them)

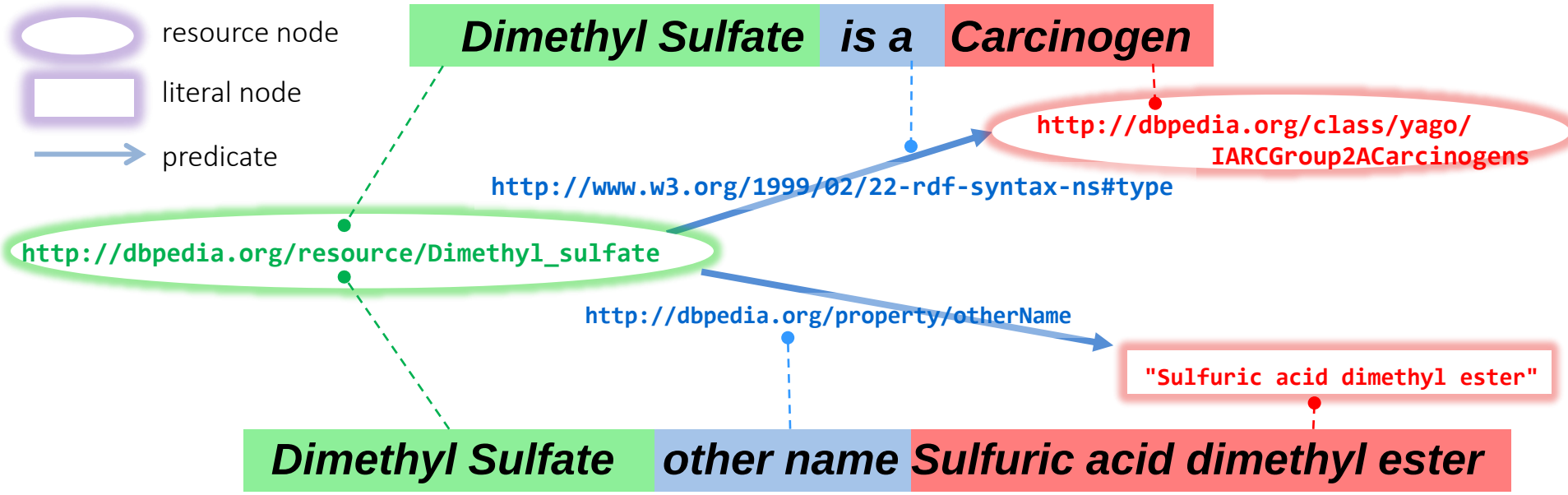


RDF Data Model

- RDF Statements are ***triples***

Subject	Predicate	Object
URI	URI	URI/Literal

- RDF data can be viewed as a directed labeled graph
  - subjects and objects are nodes (vertices)
  - predicates are oriented edges (arcs)



Water Chemical Data

- tabular data (excel file)

Region

Measure Descriptor

	A	D	E	F	G	H	I	J	K	L	AA	AB	AC	AD	AE	AF	AG
1	Banks Isl																
2		MG/L	MG/L	MG/L	MG/L	MG/L	MG/L	MG/L	MG/L	MG	/L	uS	Celsius	m asl			L=Lake
3		Na	K	SO4	Cl	Al	Ba	Fe	Li	M	U pH	COND	TEMP	ELEV	LAT	LONG	P=Pond
28	BKY	1,7	1,0	9,30	2,64	0,990	0,0341	1,130	0,0030	0,02	50 7,6	48,0	6,0	2	74 30.69N	121 41.08W	L
29	BKZ	232,0	10,8	68,70	404,00	0,110	0,0303	0,644	0,0140	0,14	50 8,1	1160,0	8,0	0	74 27.82N	122 34.55W	P
30	BKAA	3,3	0,3	2,20	12,20	0,020	0,0181	0,230	**0,0005**	0,02	0 7,6	83,0	5,0	8	74 21.46N	124 33.92W	P
31	BKAB	2,7	0,7	**0,10**	8,46	0,010	0,0292	0,068	0,0010	0,0	0 8,1	89,0	7,0	20	[illegible]	124 12.52W	P
32	BKAC	22,1	3,1	13,40	72,80	1,230	0,1250	4,940	0,0090	0,2	0 8,4	333,0	8,0	0	72 21.13N	125 24.43W	P
33	BKAD	1,9	0,7	34,40	3,91	0,220	0,0159	0,508	0,0020	0,0	0 7,8	137,0	3,0	122	71 43.79N	123 28.94W	L
34	BKAE	3,5	0,5	32,90	17,10	0,010	0,0139	0,049	**0,0005**	0,00	0 8,4	216,0	7,5	169	71 34.48N	122 30.69W	P
35	BKAF	0,9	0,3	5,00	2,52	**0,005**	0,0126	0,034	0,0010	0,00	50 7,8	109,0	3,5	175	72 14.22N	121 29.96W	L
36	BKAG	1,1	0,2	9,40	2,35	0,010	0,0103	0,075	**0,0005**	0,00	00 7,9	105,0	8,0	105	72 30.96N	119 56.11W	P
37	BKAH	0,4	0,3	1,60	0,84	0,010	0,0072	0,054	**0,0005**	0,02	0 7,7	41,0	3,0	131	73 35.57N	119 35.01W	L
38	BKAI	0,5	0,4	4,70	1,22	0,170	0,0061	0,304	0,0010	0,02	0 7,9	65,0	4,0	137	73 20.79N	116 46.23W	L
39	BKA[illegible]	1,1	[illegible]	26,[illegible]	[illegible]	[illegible]	[illegible]	080	0010	[illegible]	[illegible] 8,5	137,0	7,0	195	73 29.[illegible]	115 41.[illegible]5W	P

Measure

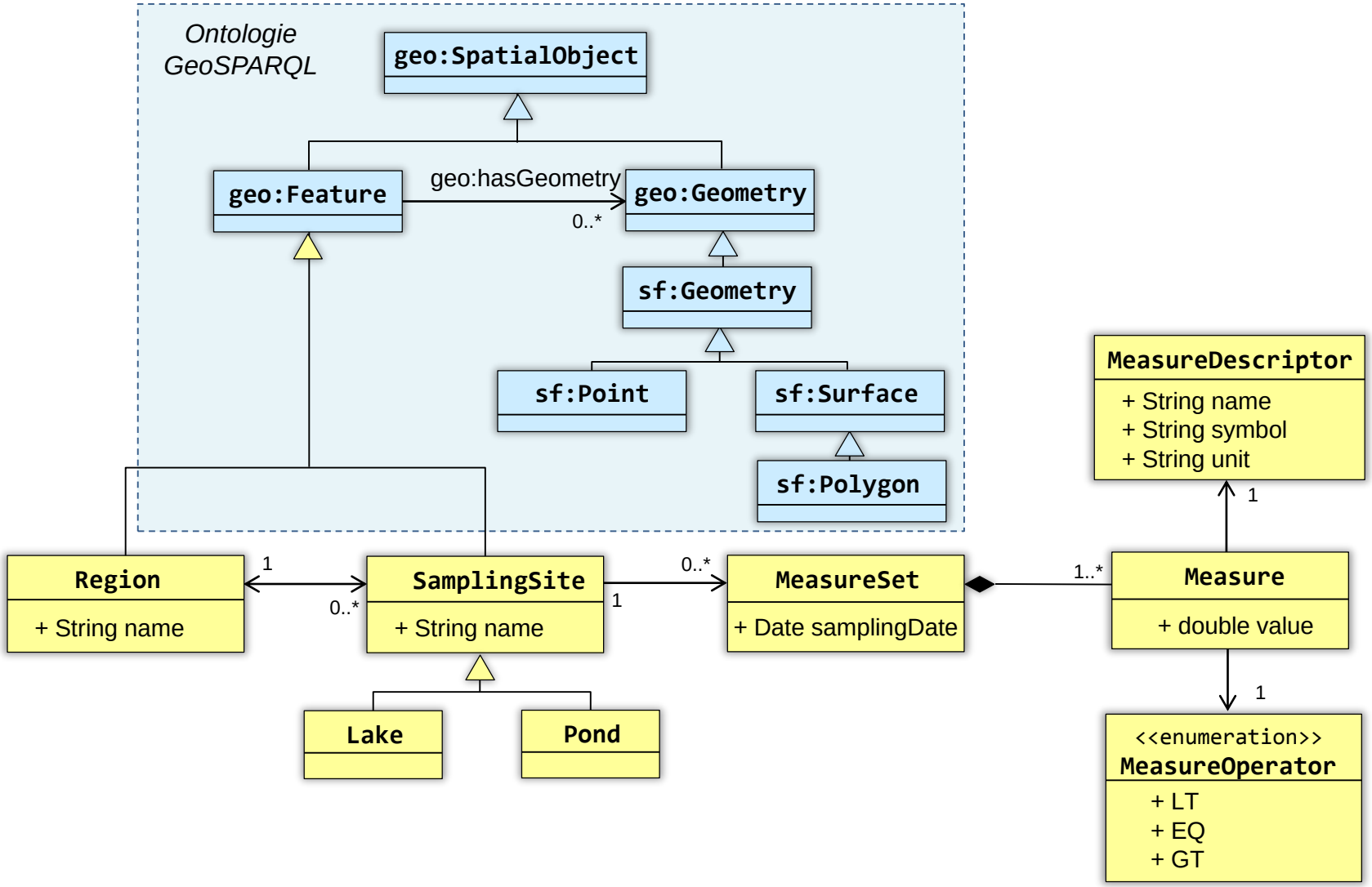
Site

Measures Set

location

site type

Data Model

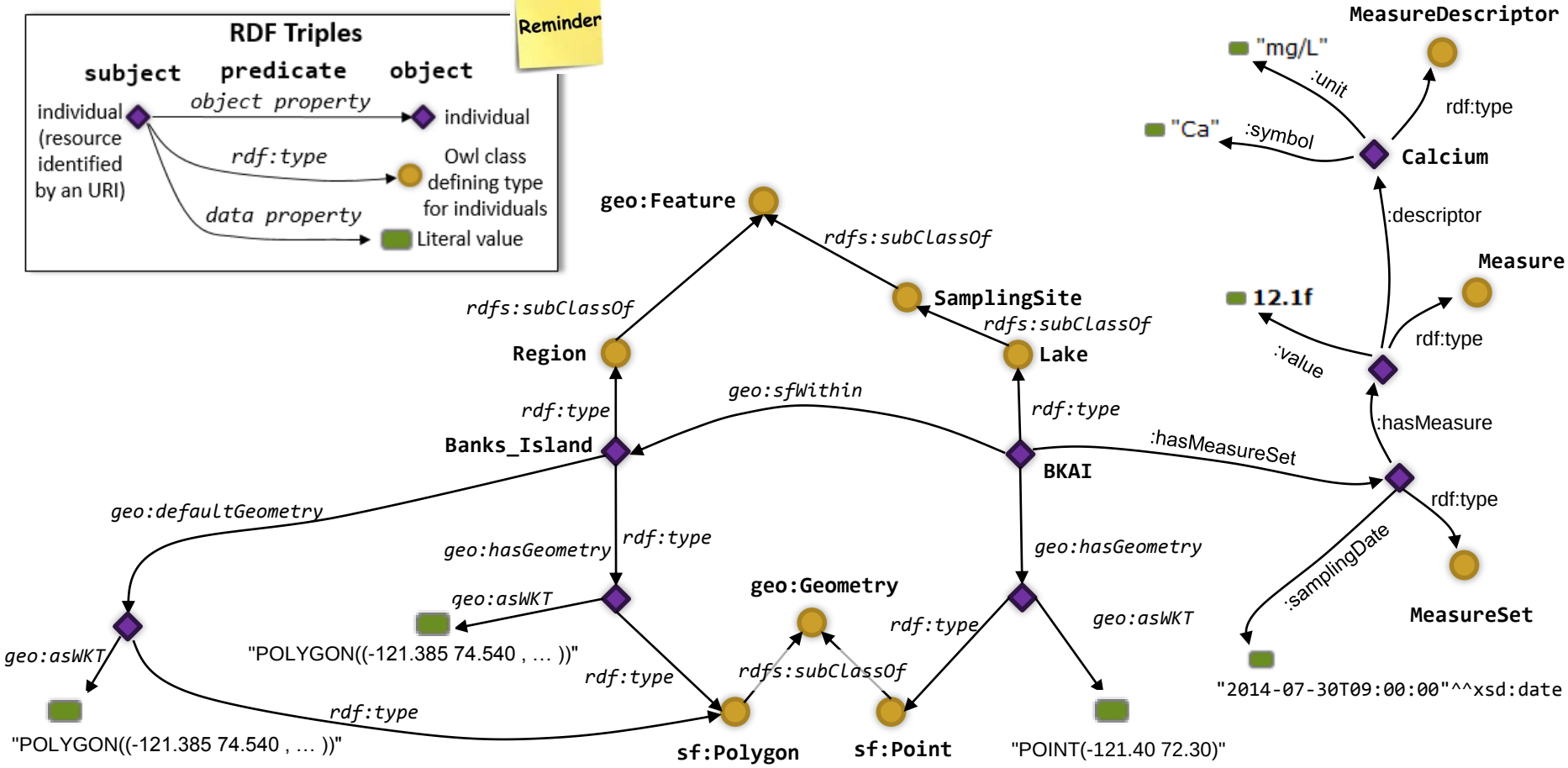

Ontologie
GeoSPARQL
geo:SpatialObject
geo:Feature
geo:hasGeometry
0..*
geo:Geometry
sf:Geometry
sf:Point
sf:Surface
sf:Polygon
MeasureDescriptor
+ String name
+ String symbol
+ String unit
1
Region
+ String name
1
0..*
SamplingSite
+ String name
0..*
1
MeasureSet
+ Date samplingDate
1..*
Measure
+ double value
1
Lake
Pond
<<enumeration>>
MeasureOperator
+ LT
+ EQ
+ GT

RDF Graph

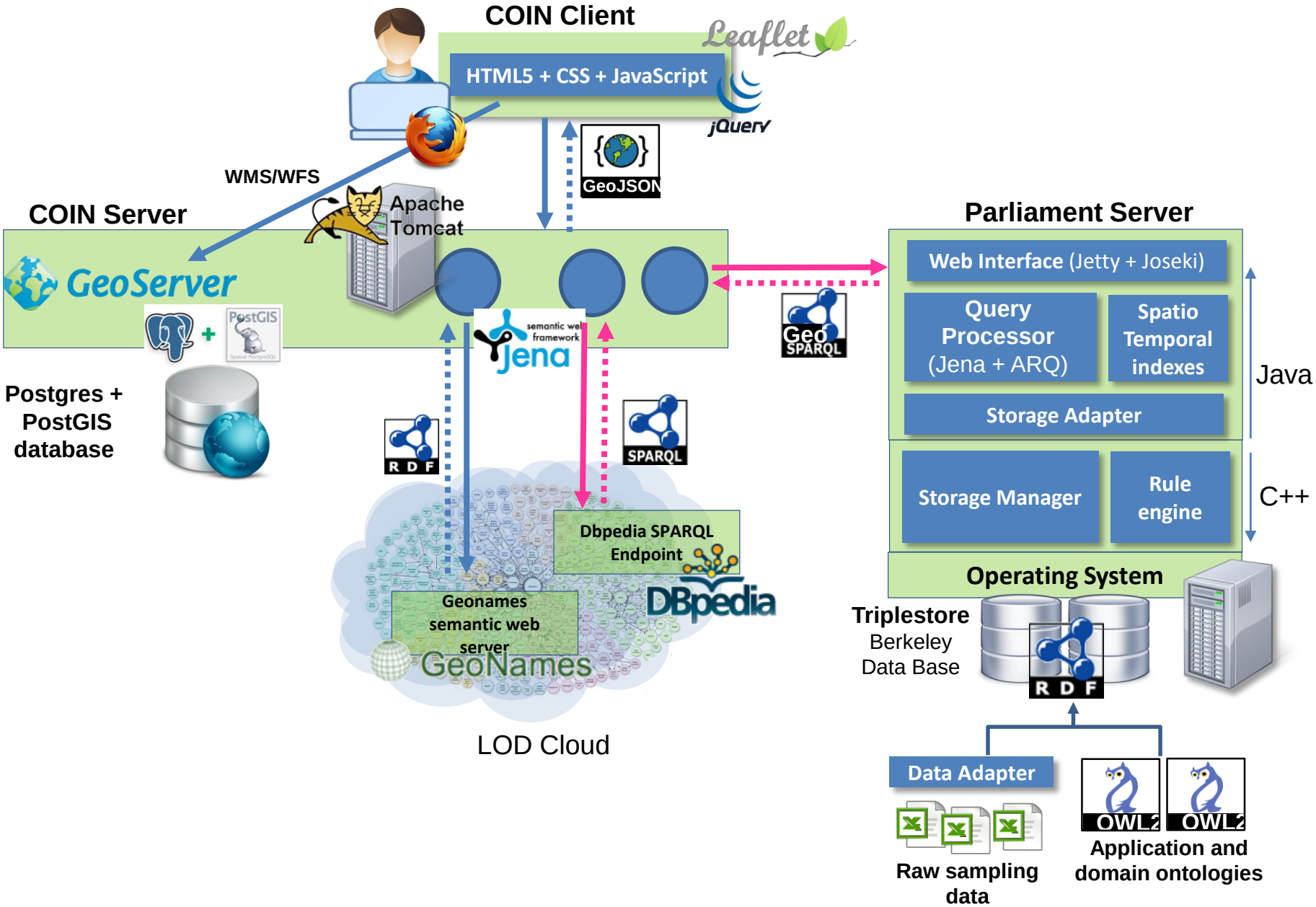
RDF Resource Description Framework (W3C standard)

Banks Isl	MG/L Na	MG/L K	MG/L SO4	MG/L Ca	MG/L Al	MG/L Ba	MG/L Fe	MG/ L	pH	uS COND	Celsius TEMP	m asl ELEV	LAT	LONG	L=Lake P=Pond
BKAI	0,5	0,4	4,70	12,10	0,170	0,0061	0,304	0,001	7,9	65,0	4,0	137	73 20.79N	116 46.23W	L

Reminder



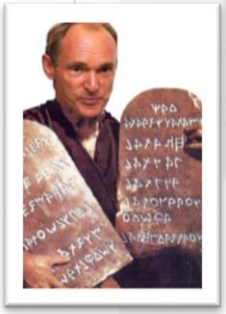
Architecture of COIN semantic component



TripleStore Parliament

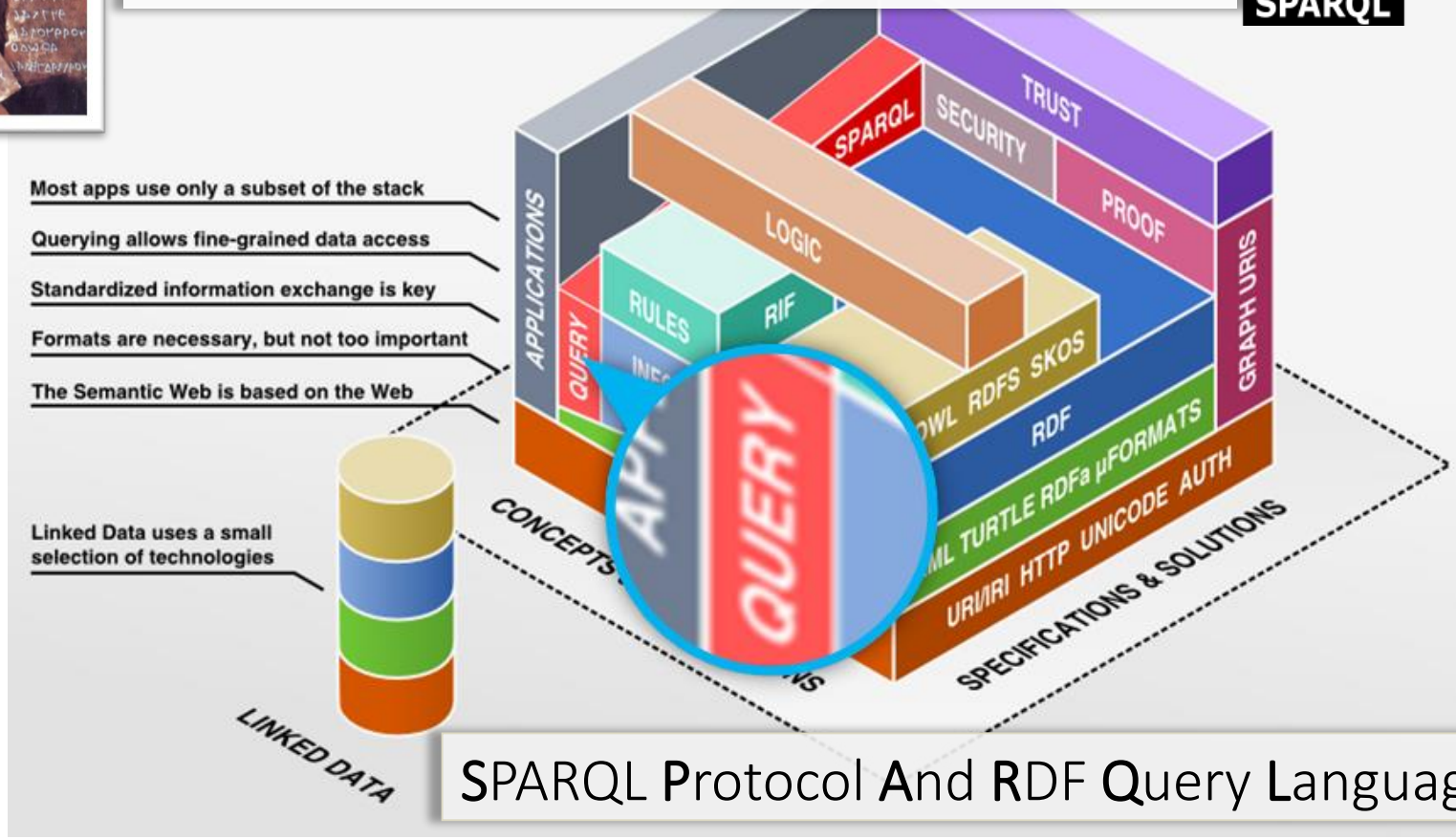
- Storage Engine
- Developed by Raytheon BBN Technologies (Dave Kolas)
- First implementation of GeoSPARQL
  - Geometries represented using typed literals
    - WKT & GML serializations supported
  - Three families of topological functions exposed
    - OGC-SFA
    - Egenhofer
    - RCC8
  - Multiple CRS support

Querying Linked Data with SPARQL



Linked Data: 3rd Principle

*When someone looks up a URI, provide useful information, using the standards (RDF, **SPARQL**).*



SPARQL Protocol And RDF Query Language

SPARQL : introduction

- RDF (Resource Description Framework)
  - Flexible and extensible way to represent information about resources of the web
- SPARQL (SPARQL Protocol And RDF Query Language)
  - A W3C standard
    - SPARQL 1.0 recommendation - January 2008,
    - SPARQL 1.1 recommendation – March 2013 http://www.w3.org/TR/2013/REC-sparql11-overview-20130321/
  - a request language to access a RDF graph (SPARQL Query Language Specification) inspired from SQL
  - a protocol to submit request through HTTP GET, HTTP POST or SOAP (SPARQL protocol for RDF Specification)
  - an XML format for the results (SPARQL Query XML Results Format), and now JSON

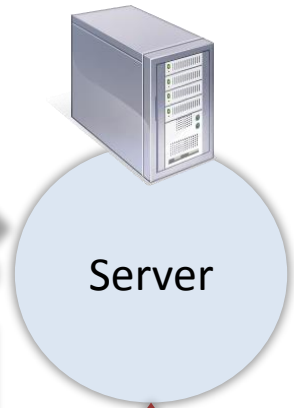
SPARQL Protocol And RDF Query Language
overview

Client

a protocol to SPARQL Request over HTTP
(**SPARQL protocol for RDF Specification**)

SPARQL Protocol Layer

HTTP



Server

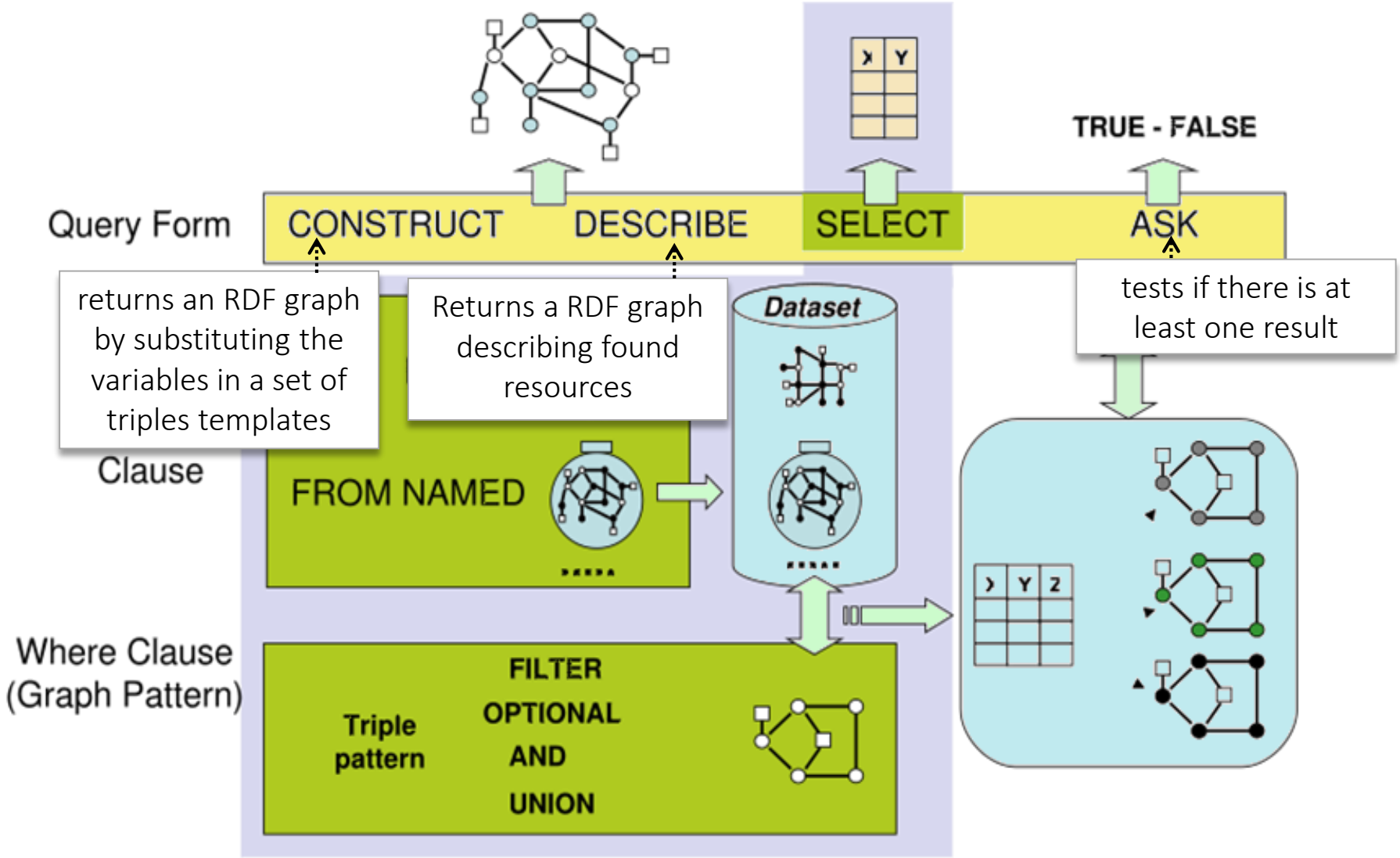
An XML output format (and now also JSON, ...) for SPARQL queries results (**SPARQL Query XML Results Format**)

User Interface

A request language to query RDF graphs
(**SPARQL Query Language**)

SPARQL Endpoint

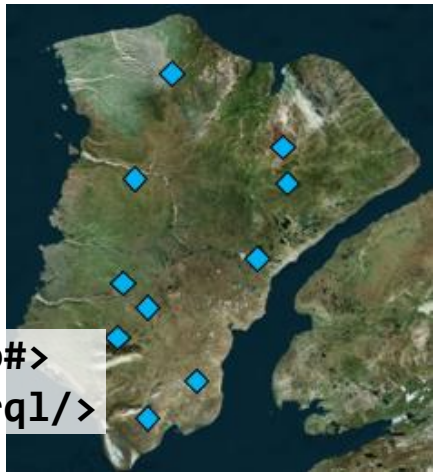
SPARQL Query Language: types of requests



from: Pérez, Arenas and Gutierrez, Chapter 1: On the Semantics of SPARQL, Semantic Web Information Management: A Model Based Perspective, Springer 2010

SPARQL Query Language
Select

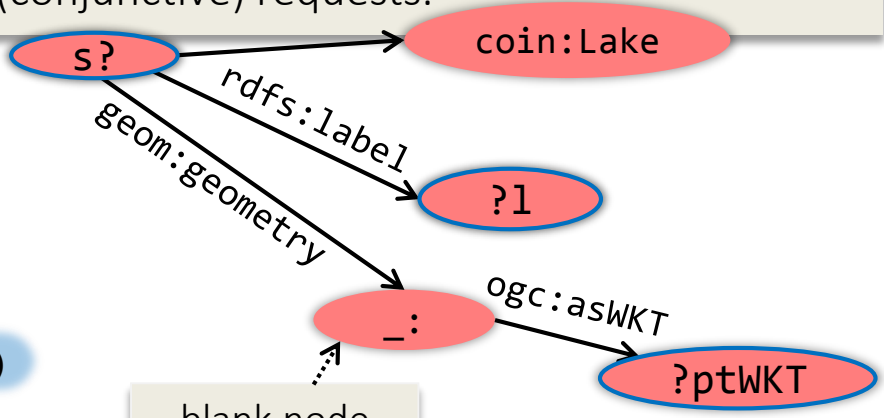
- find all the sampling sites of type lake that are inside Banks Island



```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX geo: <http://www.opengis.net/ont/geosparql#>
PREFIX coin: <http://coin.yukonwaterboard.ca/ont/coinOnto#>
PREFIX geof: <http://www.opengis.net/def/function/geosparql/>
SELECT ?s, ?name
WHERE
{
    ?s a coin:Lake;
       rdfs:label ?l ;
       geo:hasGeometry [
           geo:asWKT ?ptWKT
       ].
    ?region rdfs:label "Banks Island";
       geo:hasGeometry [
           geo:asWKT ?rGeom
       ].
    FILTER (geof:sfWithin(?ptWKT,?rWKT))
}
```

variables

Graph Pattern: RDF triple containing one or more variables at any position (subject, predicate, object). Graph patterns can be combined to construct complex (conjunctive) requests.



blank node

Filter on ?ptWKT variable

Integrative Environmental data needs

- Analyses require a wide range of data
  - Broad scales: geospatial, temporal, biological (micro-macro)
  - Diverse topics:
    - Predicting impact
    - Documenting effects of climate change
- Large amounts of relevant data…
- But researchers struggle to …
  - Discover relevant datasets / information for a study
  - And combine these into an integrated product to analyze
- Need consistency and rigor in terminology
- Standardized protocols, methods when possible
  - Interoperability (syntax)
  - Comparability (semantics)
- Minimally, need a "shared community vocabulary"
  - For hydrologists--- WATERML?
  - metadata and keywords
  - controlled vocabularies:

Shared Knowledge models

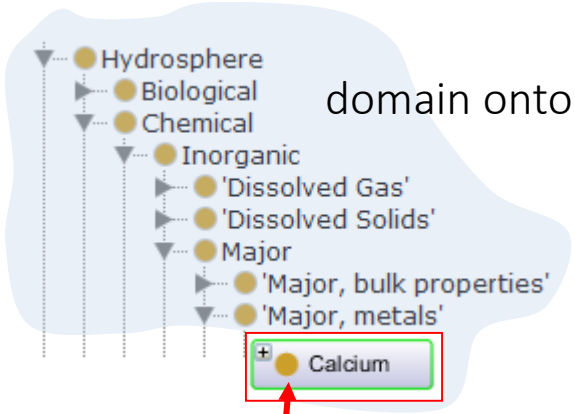
- Ontologies provide a “shared vocabulary”
  - Common “external” definitions (namespaces)
  - for explicating relationships among terms
  - describing data schemas (observations)
  - for machine-assisted discovery, reasoning, integration

- Standard technologies for creating and operating on ontologies:
  - Syntaxes: RDF, SKOS, OWL
  - FOSS applications and frameworks: Jena, Protégé
  - Standard Reasoners: Pellet, FaCT++, Racer

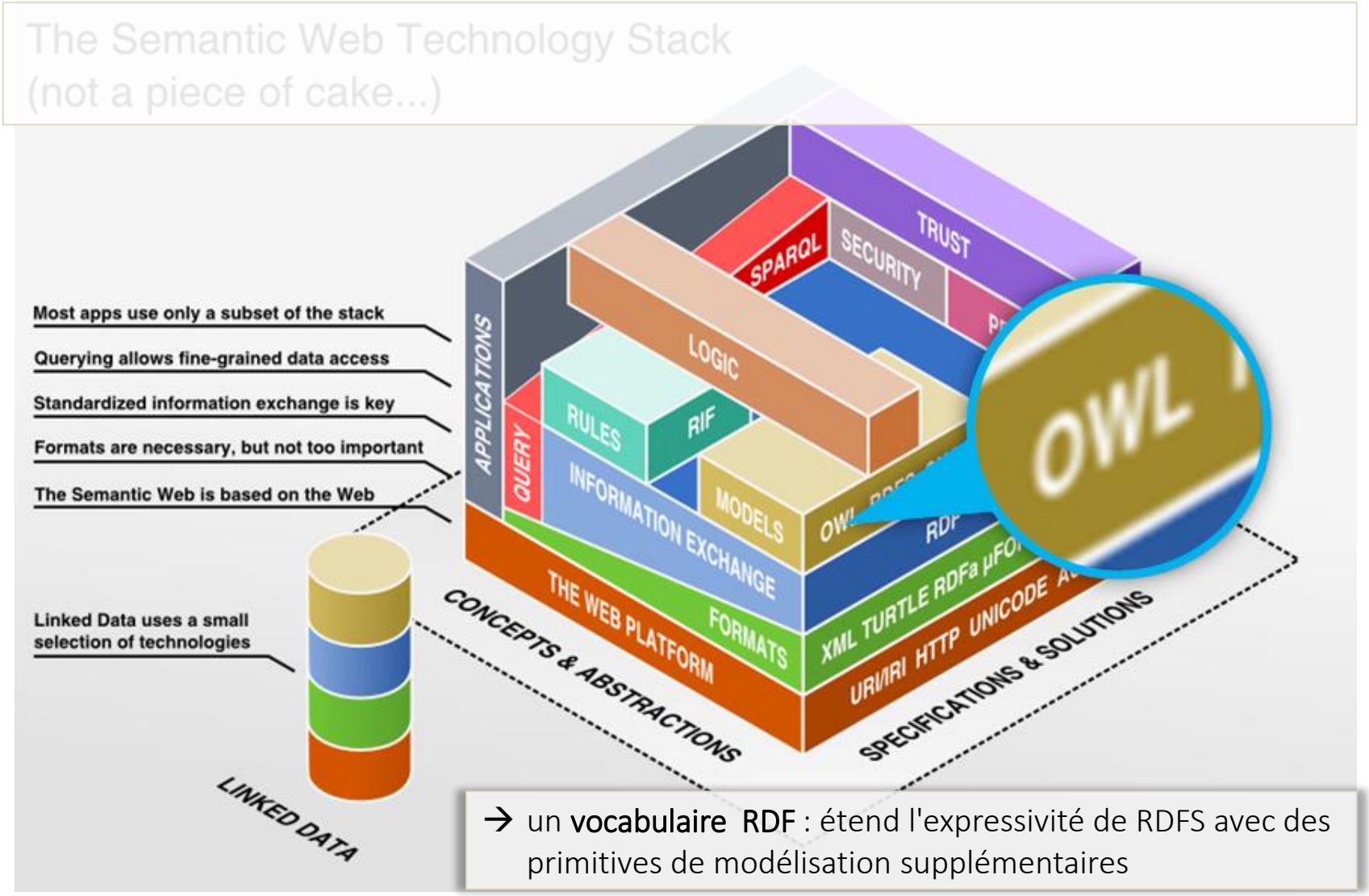
Shared Knowledge ontologies

- GeoSPARQL Components
- **Hydrologic Concepts Hierarchy**
- Observations and measurements ontology

Consortium of Universities for the Advancement of Hydrologic Science http://www.cuahsi.org



Web Ontology Language (OWL)



→ un **vocabulaire RDF** : étend l'expressivité de RDFS avec des primitives de modélisation supplémentaires

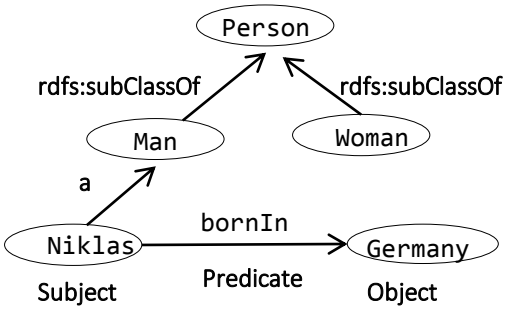
Web Ontology Language (OWL)

- a W3C standard
  - OWL 1 : W3C recommendation 10 Feb. 2004 http://www.w3.org/TR/owl-features/
  - OWL 2 : W3C recommendation 11 Dec. 2012 http://www.w3.org/TR/owl2-overview/

- OWL vocabulary : a set of primitives described in RDF that extends RDFS vocabulary
  - namespace
    http://www.w3.org/2002/07/owl# ⇔ **owl:**

- Far more expressive than RDFS
  - Classes can be describe by union, intersection, complement, properties restrictions.
  - notions of classes or properties equivalence, resources equality,
  - notions of inverse, symmetric, transitive … properties
  - properties cardinality…
- Formal specification (based on Description Logics)
  → support for automated reasonning

RDFS (RDF Schema):
Vocabulary for defining hierarchical classes and properties (taxonomies)

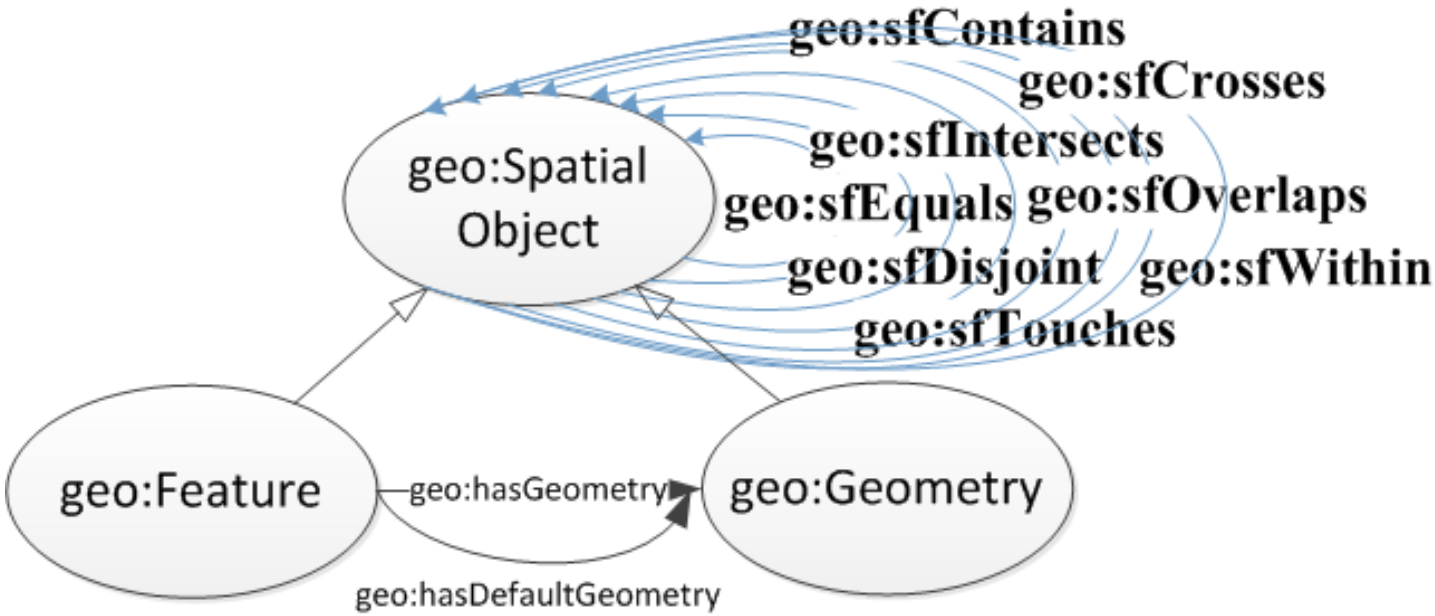


OWL (Web Ontology Language):
Ontology language based on DL (subset of 1rst Order Logic) to describe classes

$$\texttt{Person} \equiv \texttt{Man} \cup \texttt{Woman}$$
$$\texttt{Man} \cap \texttt{Woman} \equiv \varnothing$$

GeoSPARQL Topology Vocabulary Extension

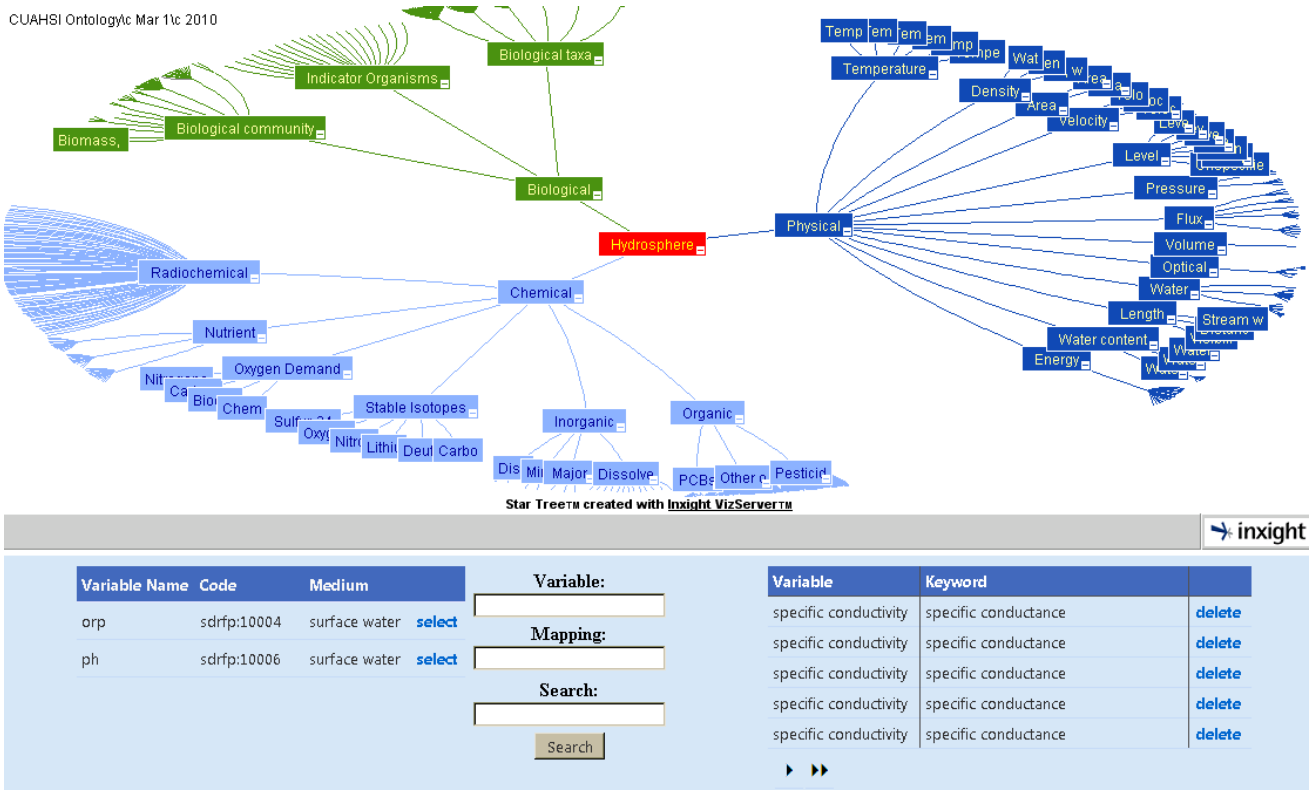
- The extension is parameterized by the family of topological relations supported.
  - Topological relations for simple features



  - The Egenhofer relations e.g., **`geo:ehMeet`**
  - The RCC8 relations e.g., **`geo:rcc8ec`**

Domain Ontology

- Keyword Hierarchy used to categorize and assist in the discovery of monitored variables.
- Each Variable is “tagged” to a keyword concept.



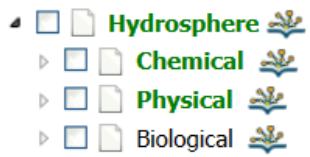
Water Quality Sampling data

- Data enhancement: Linking sampling variables with CUASHI ontology

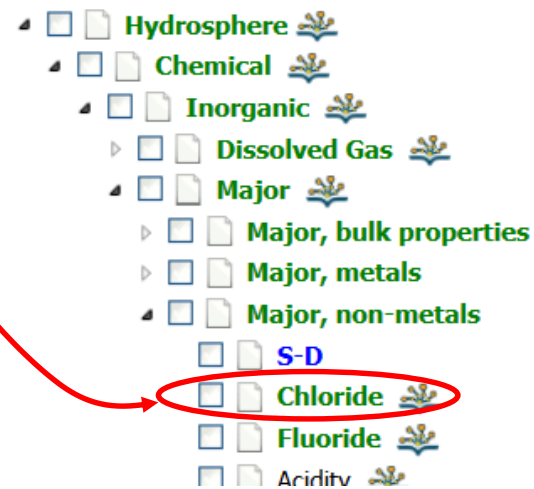
138	Site-A	01/01/2014	Ca-D	Calcium, dissc	Ca-D	met	6,55	mg/L	63,8167	-139,205
139	Site-A	01/01/2014	Cd-D	Cadmium, dis	Cd-D	met	0,00003	mg/L	63,8167	-139,205
140	Site-A	01/01/2014	Chloride	Chloride	Chlord	inorg	<0.05	mg/L	63,8167	-139,205
141	Site-A	01/01/2014	Chrysene	Chrysene	Chrysene	org	0,299	ng/L	63,8167	-139,205
142	Site-A	01/01/2014	Co-D	Cobalt, dissol	Co-D	met	0,0002	mg/L	63,8167	-139,205

CUAHSI
universities allied for water research

Consortium of Universities for the Advancement of Hydrologic Science
http://www.cuahsi.org



More than 4000 concepts organized in a hierarchy (taxonomy)

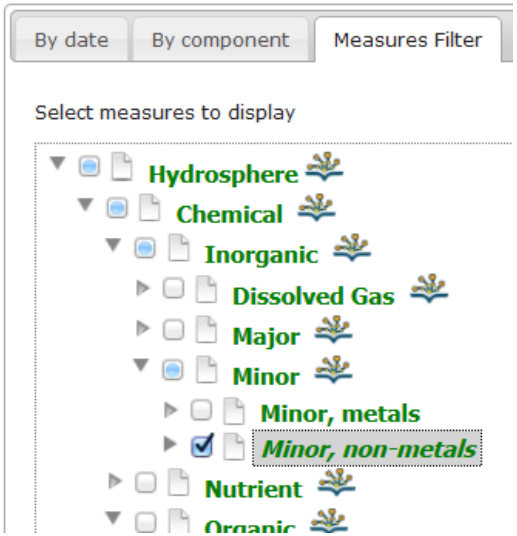


- 252 variables used by YWB WQ sampling data
  - 200 can be classified directly inside the CUSASHI ontology
- YWB ontology a subset of CUASHI ontology with addition of new concepts for unclassified variables
- Hierarchy of 350 concepts used by YWB ontology

Canadian Council of Ministers of the Environment CCME

Water Quality Sampling data

- The CUASHI ontology can be used to filter data by selecting various levels of abstraction



Name	Description	Code	Class	Result	Units
Alk-Bicrb	Alkalinity, Bicarbonate HCO3	Alk-B	phy	=158	mgCaCO3/L
Alk-Carb	Alkalinity, Carbonate CO3	Alk-C	phy	<6	mgCaCO3/L
Alk-OH	Alkalinity, Hydroxide OH	Alk-OH	phy	<5	mgCaCO3/L
Alk-T	Alkalinity, Total	Alk-T	phy	=129	mgCaCO3/L
As-D	Arsenic, dissolved	As-D	met	=0.00019999999494757503 3	mg/L
As-T	Arsenic, total	As-T	met	<9.99999974737875164e-05	mg/L
B-D	Boron, dissolved	B-D	met	=0.0199999995529651642	mg/L
Sb-D	Antimony, dissolved	Sb-D	met	=0.00019999999494757503 3	mg/L
Sb-T	Antimony, total	Sb-T	met	=0.00015000000712461769 6	mg/L
Se-D	Selenium, dissolved	Se-D	met	=0.00079999999797903001 31	mg/L

Showing 1 to 10 of 20 records

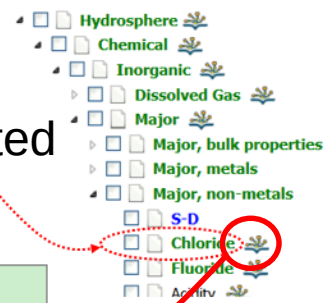
Pages: Previous 1 2 Next

Only the relevant records are displayed

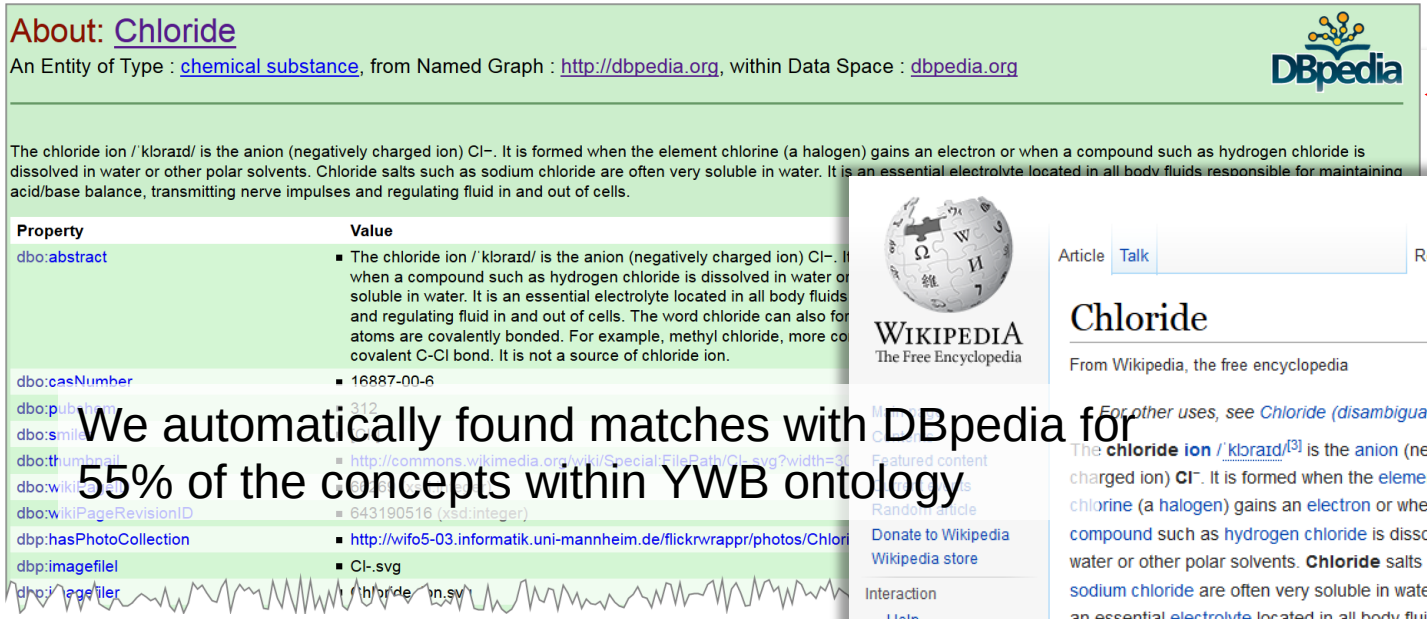
Water Quality Sampling data

- Data enhancement: Linking YWB ontology with DBpedia (and Wikipedia) to get extra information about the variables used for sampling

138	Site-A	01/01/2014	Ca-D	Calcium, dissc	Ca-D	met	6,55	mg/L	63,8167	-139,205
139	Site-A	01/01/2014	Cd-D	Cadmium, dis	Cd-D	met	0,00003	mg/L	63,8167	-139,205
140	Site-A	01/01/2014	Chloride	Chloride	Chlord	inorg	<0.05	mg/L	63,8167	-139,205
141	Site-A	01/01/2014	Chrysene	Chrysene	Chrysene	org	0,299	ng/L	63,8167	-139,205
142	Site-A	01/01/2014	Co-D	Cobalt, dissol	Co-D	met	0,0002	mg/L	63,8167	-139,205



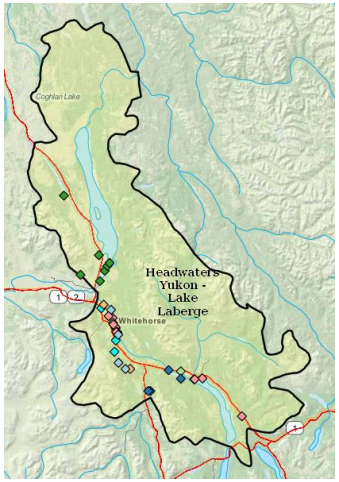
DBpedia is structured data (accessible by programs) extracted from Wikipedia (accessible by human)



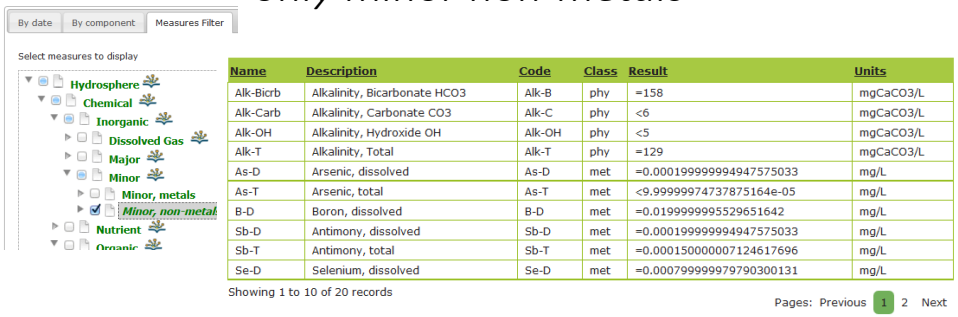
We automatically found matches with DBpedia for 55% of the concepts within YWB ontology

Geospatial and Semantic Queries

- Geospatial queries
  - Using spatial information to filter data
  - Ex. : *find all the Water licences inside a watershed*



- Semantic queries
  - Using some ontology to filter data
  - Ex : *for water quality data, display only minor non-metals*

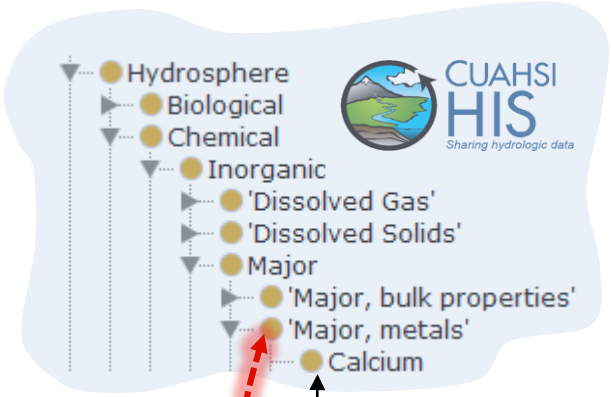


Name	Description	Code	Class	Result	Units
Alk-Bicrb	Alkalinity, Bicarbonate HCO3	Alk-B	phy	=158	mgCaCO3/L
Alk-Carb	Alkalinity, Carbonate CO3	Alk-C	phy	<6	mgCaCO3/L
Alk-OH	Alkalinity, Hydroxide OH	Alk-OH	phy	<5	mgCaCO3/L
Alk-T	Alkalinity, Total	Alk-T	phy	=129	mgCaCO3/L
As-D	Arsenic, dissolved	As-D	met	=0.00019999999947575033	mg/L
As-T	Arsenic, total	As-T	met	<9.99999974737875164e-05	mg/L
B-D	Boron, dissolved	B-D	met	=0.019999999529651642	mg/L
Sb-D	Antimony, dissolved	Sb-D	met	=0.00019999999947575033	mg/L
Sb-T	Antimony, total	Sb-T	met	=0.00015000000712461769 6	mg/L
Se-D	Selenium, dissolved	Se-D	met	=0.00079999997979030013 1	mg/L

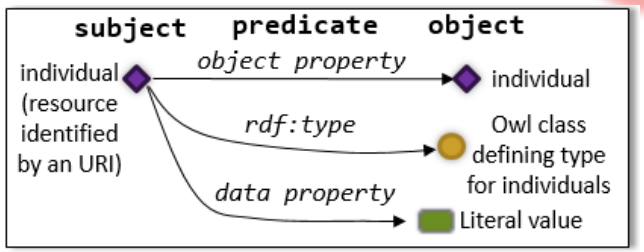
- Geospatial semantic queries
  - Using spatial **and** semantic information to filter data
  - Ex. : *find the water quality stations in a given watershed that have a value for minor non-metals above a given threshold*

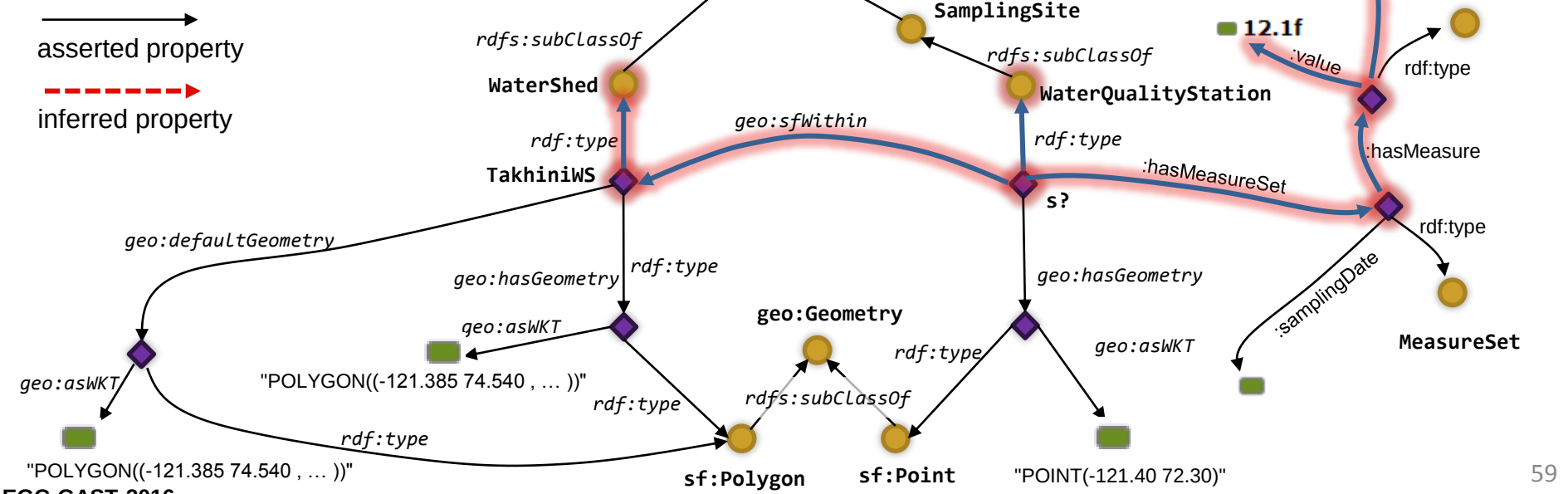
Geospatial and Semantic Queries

- *find the water quality stations in a given watershed that have a value for a major metal above a given threshold*



Takhini								
	MG/L Na	MG/L K	MG/L SO4	MG/L Ca	MG/L Al	MG/L Ba	MG/L Fe	MG/L Li
TAKA1	0,5	0,4	4,70	12,10	0,170	0,0061	0,304	0,0016

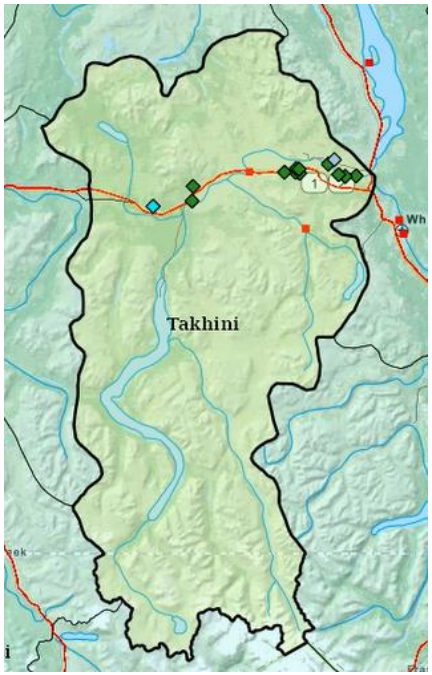




Geospatial and Semantic Queries

- *find the water quality stations in a given watershed that have a value for major metal above a given threshold*

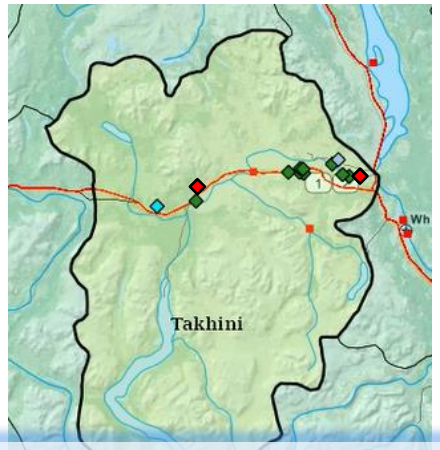
```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX geo: <http://www.opengis.net/ont/geosparql#>
PREFIX geof: <http://www.opengis.net/def/function/geosparql/>
PREFIX coin: <http://coin.yukonwaterboard.ca/ont/coinOnto#>
PREFIX cuahsiOnto: <http://coin.yukonwaterboard.ca/ont/cuashiOnto#>
SELECT ?s, ?name
WHERE {
     ?s a coin:WaterQualityStation;
        rdfs:label ?l ;
        geo:hasGeometry [
              geo:asWKT ?ptWKT
           ].
     ?ws a coin:WaterShed;
         rdfs:label "Takhini";
          geo:hasGeometry [
               geo:asWKT ?rGeom
          ].

     FILTER (geof:sfWithin(?ptWKT,?rWKT))
}
```



Geospatial and Semantic Queries

- *find the water quality stations in a given watershed* ***that have a value for a major metal above a given threshold***

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX geo: <http://www.opengis.net/ont/geosparql#>
PREFIX geof: <http://www.opengis.net/def/function/geosparql/>
PREFIX coin: <http://coin.yukonwaterboard.ca/ont/coinOnto#>
PREFIX cuahsiOnto: <http://coin.yukonwaterboard.ca/ont/cuashiOnto#>
SELECT ?s, ?name
WHERE {
     ?s a coin:WaterQualityStation;
        rdfs:label ?l ;
        geo:hasGeometry [
              geo:asWKT ?ptWKT
           ].
     ?ws a coin:WaterShed;
         rdfs:label "Takhini";
          geo:hasGeometry [
               geo:asWKT ?rGeom
          ].
     FILTER (geof:sfWithin(?ptWKT,?rWKT))


}
```



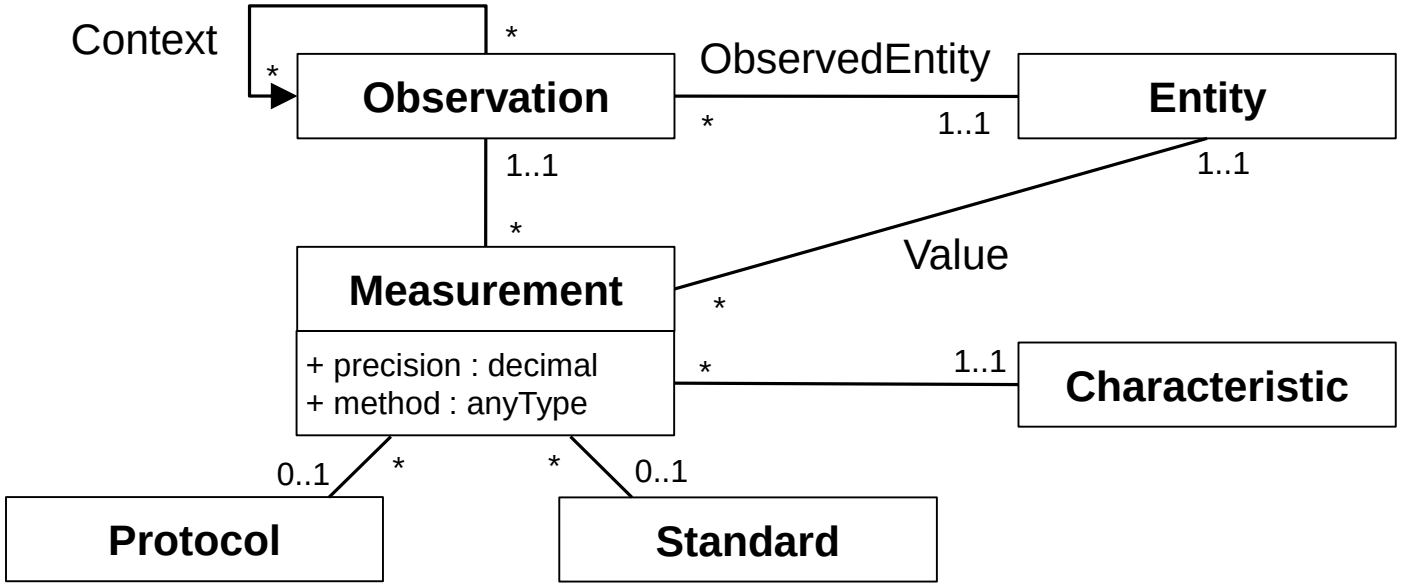

```
?s coin:hasMeasureSet ?ms.
?ms coin:hasMeasure ?m.
?m coin:mDescriptor ?md.
?md rdf:type cuahsiOnto:MajorMetal.
?m coin:mValue ?val.

FILTER (?val >=10.0)
```

Several observation models…

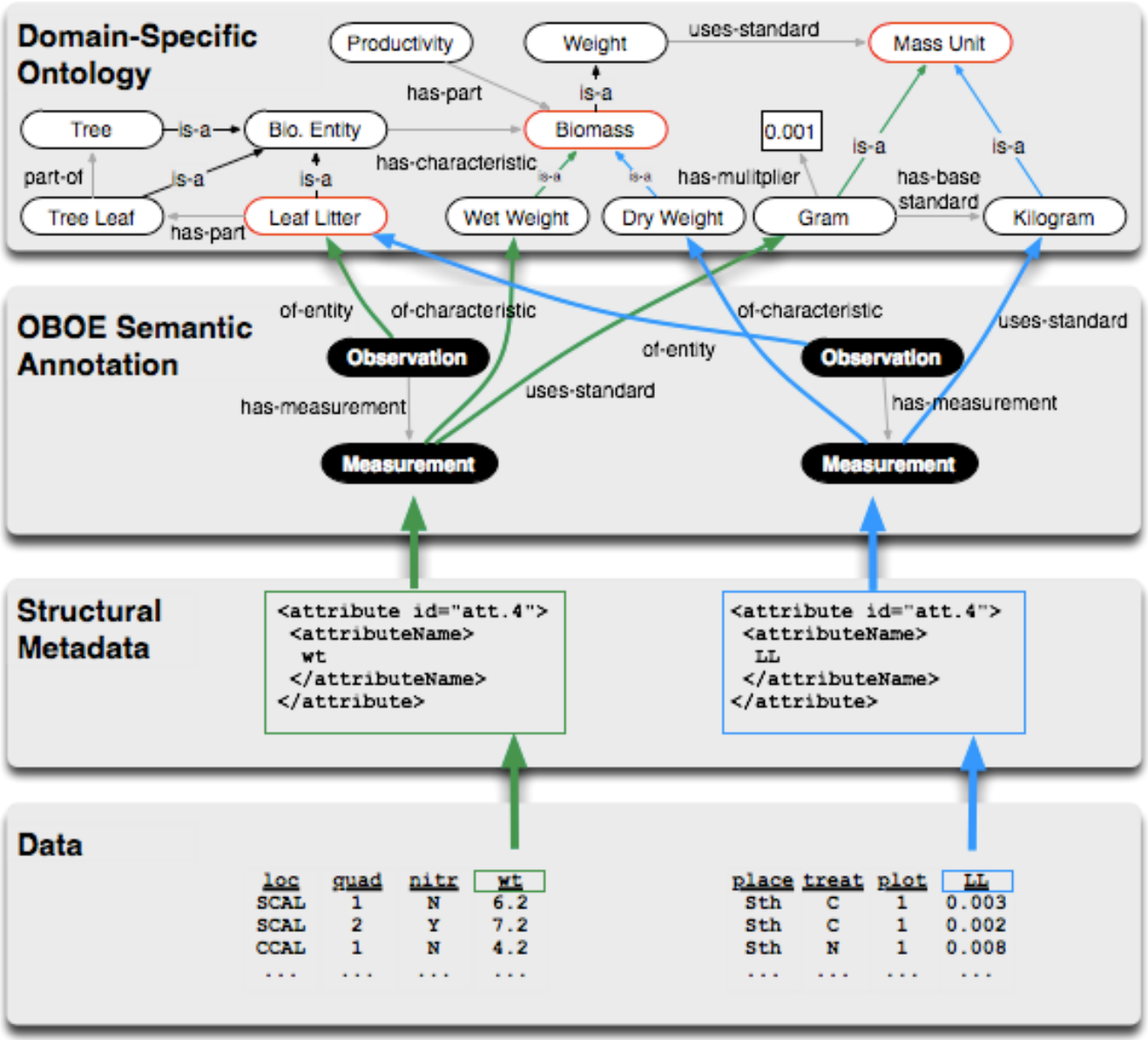
Project	Domain	Observational data model
VSTO	Atmospheric sciences	Ontologies for interoperability among different meteorological metadata standards and other atmospheric measurements
SERONTO	Socioecological research	Ontology for integrating socio-ecological data
OGC's O&M	Geospatial	Observations and Measurements standard for enhancing sensor data interoperability
SEEK's OBOE	Ecology	Extensible Observation Ontology for describing data as observations and measurements
PATO's EQ	Phenotype/Evolution	Underlying model for describing phenotypic traits to link with genomic data

Observational Data Model

- Implemented as an OWL-DL ontology
  - Provides basic concepts for describing observations
  - Specific "*extension points*" for domain-specific terms



Semantic annotation



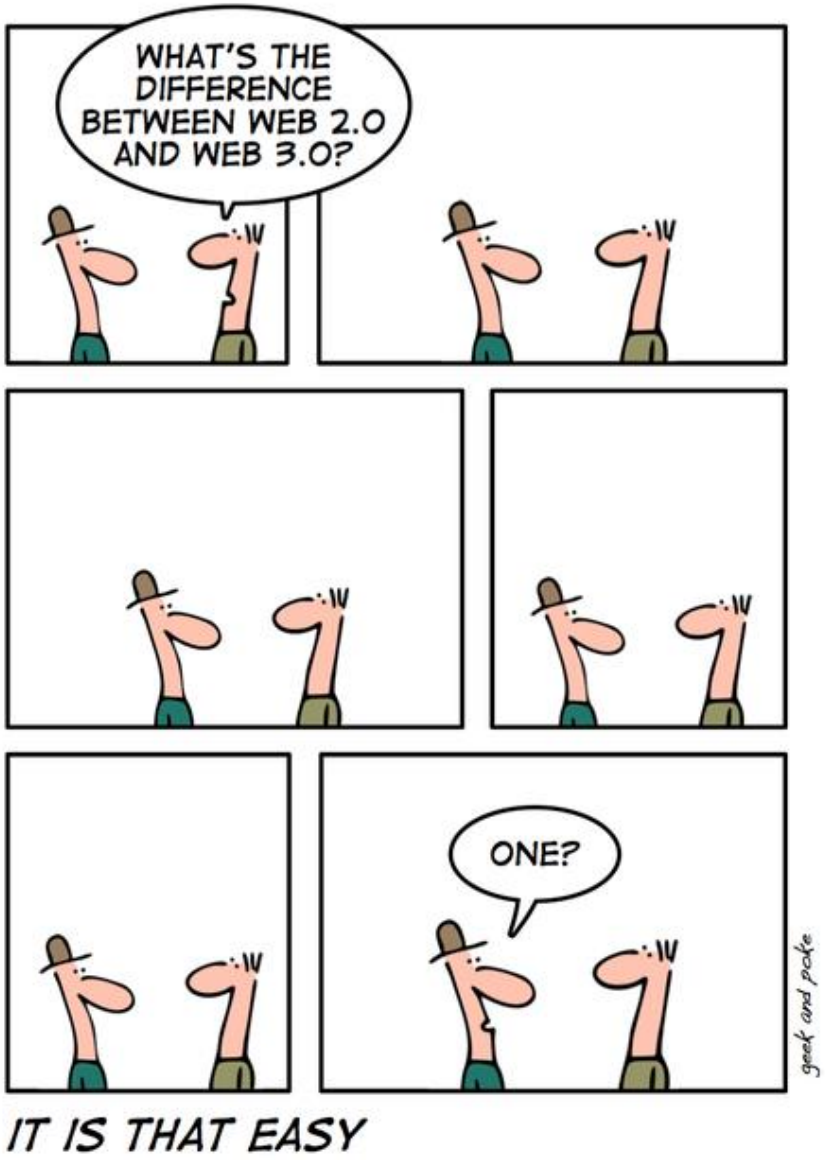
Several prospective observation models…

Project	Domain	Observational data model
VSTO	Atmospheric sciences	Ontologies for interoperability among different meteorological metadata standards and other atmospheric measurements
SERONTO	Socioecological research	Ontology for integrating socio-ecological data
OGC's O&M	Geospatial	Observations and Measurements standard for enhancing sensor data interoperability
SEEK's OBOE	Ecology	Extensible Observation Ontology for describing data as observations and measurements
PATO's EQ	Phenotype/Evolution	Underlying model for describing phenotypic traits to link with genomic data

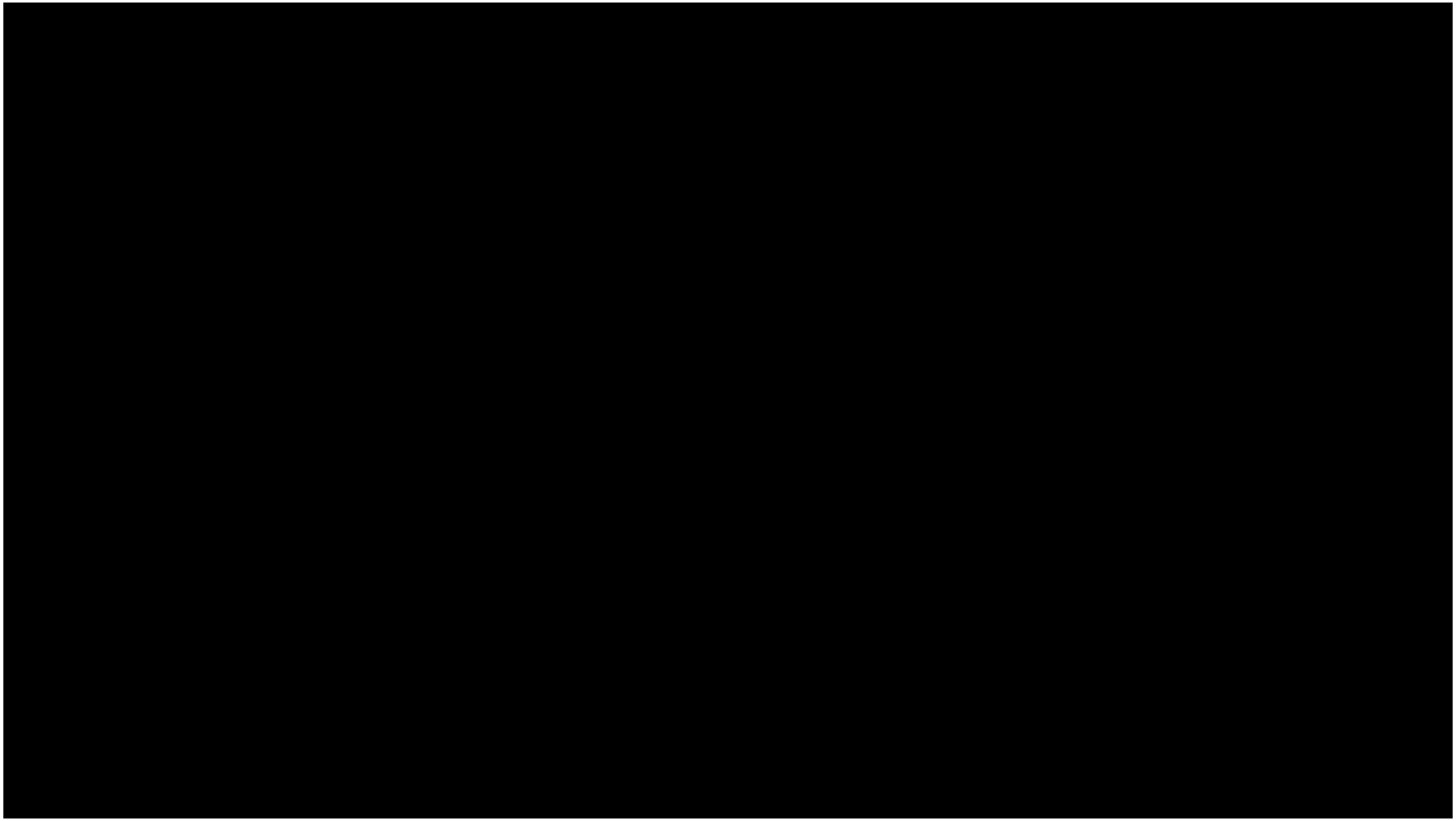
Conclusion

- Developed but not yet public
  - Preproduction server at Grenoble University
- Developed
  - Interactive map of Yukon
  - Display water quality data in a table
  - Display water quality data in a graph
  - Search water quality data by chemical
  - Search water quality data by chemical in a geographic area
  - Filter map layers to anything within a geographic area
  - Display YWB Water licence information
  - Display water flow direction
  - Publishing YWB Water licence data to the GeoYukon site
- Future
  - Finding and displaying more data
  - Setting standards for water quality data received by the YWB
  - Better user interface
  - Deploy on a production server
  - Convergence with GeoYukon tools

Merci de votre attention

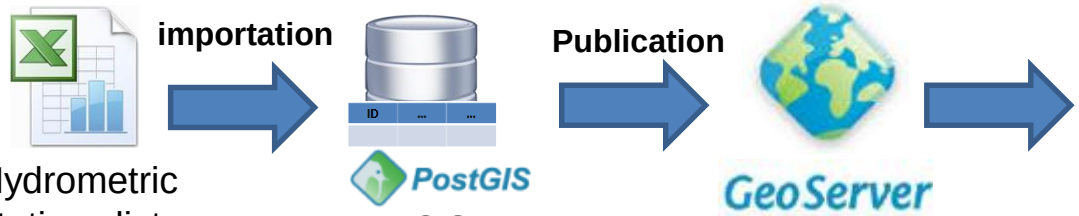


COIN Web application



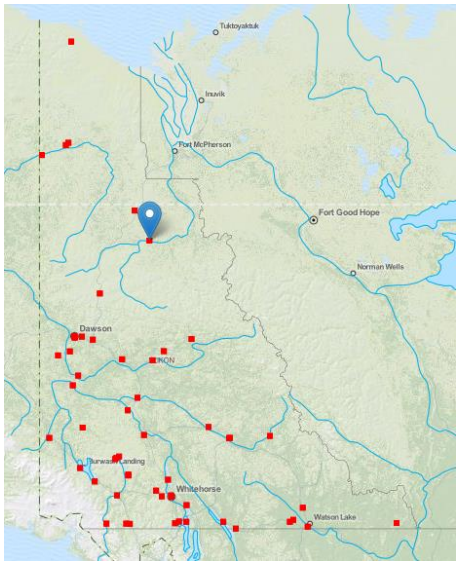
Hydrometric stations

- Data source:
  - Federal hydat
- Data publication

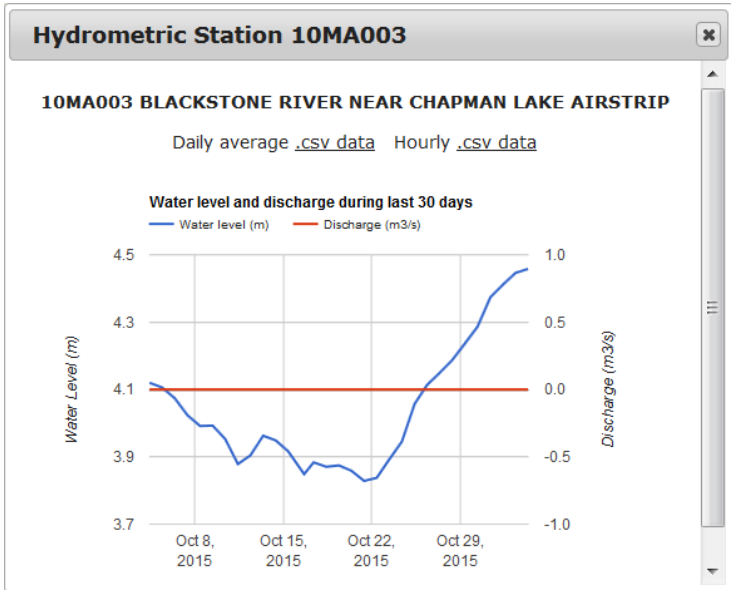


Hydrometric stations list extracted from federal hydat csv file

GIS Database

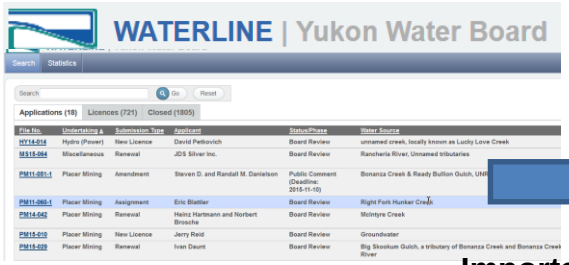


- Data visualization
  - Direct access to federal hydat servers (http://dd.weather.gc.ca/hydrometric/) to get measures for the last 30 days



Water Licences

- Data source: YWB Waterline
- Data publication

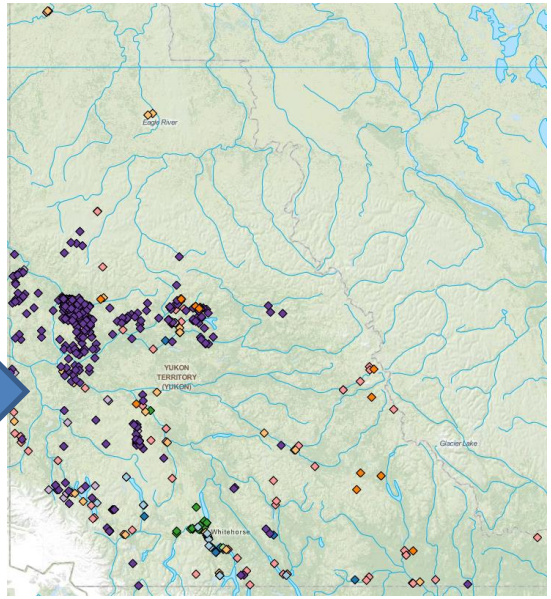


Importation (update every week)

PostGIS
GIS Database

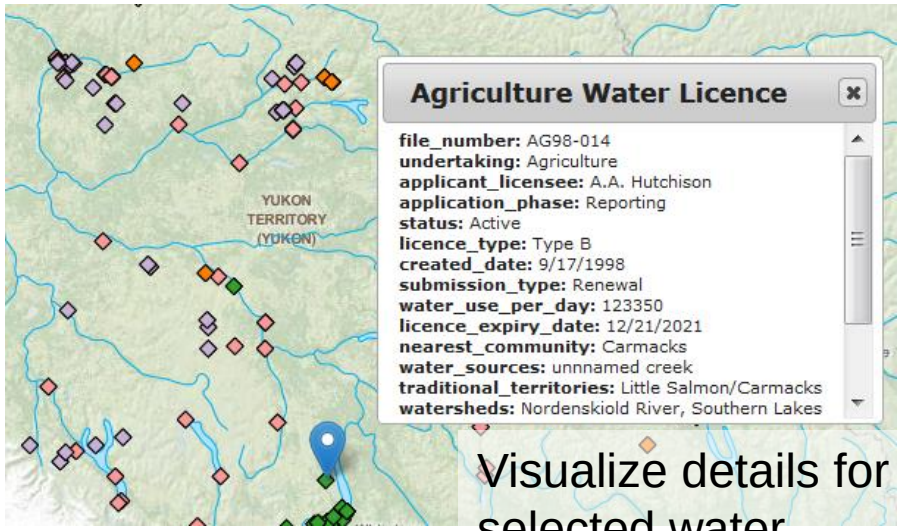
Publicatio

GeoServer



- Data visualization

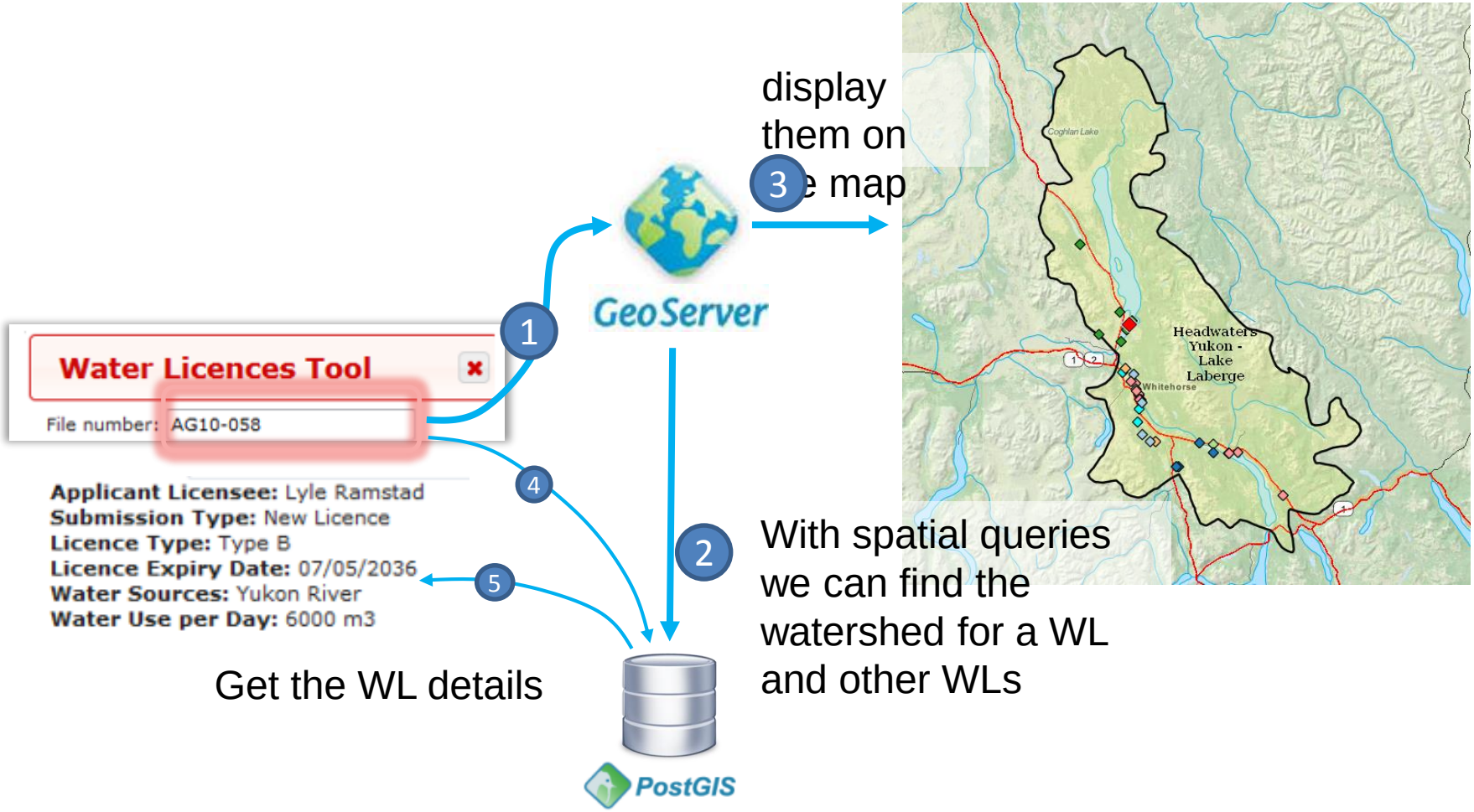
Filter according undertaking type



Visualize details for a selected water

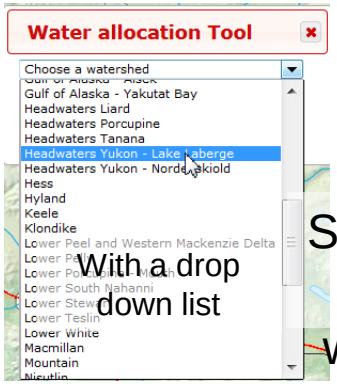
Water Licences

- Water Licence tool



Water Licences

- Water Allocation Tool



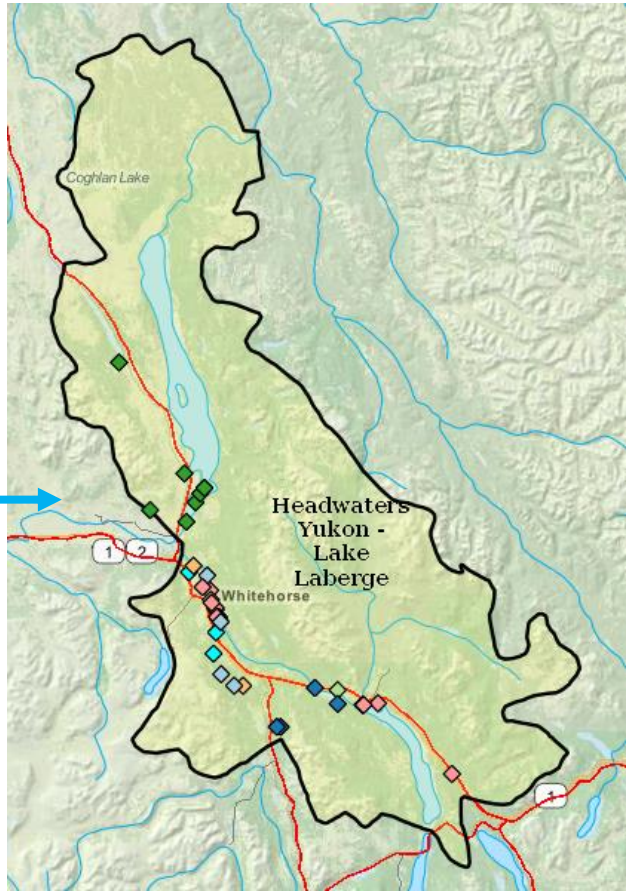
With a drop down list

Directly on the map

Selection of a watershed 1

- Number of Water Licences: **36**
- Total Water Allocation per day: **428322 m^3**

3 display them on the map



Calculate the total amount of water allocated for that WS 4

details

- **Agriculture** count: **8** total: **166326 m^3**
- **Conservation** count: **3** total: **193000 m^3**
- **Hydro (Power)** count: **4** total: **15176 m^3**
- **Industrial** count: **3** total: **28100 m^3**
- **Miscellaneous** count: **11** total: **4610 m^3**
- **Municipal** count: **3** total: **20160 m^3**
- **Recreational** count: **4** total: **950 m^3**

2 With spatial queries we can find all WLs in the WS