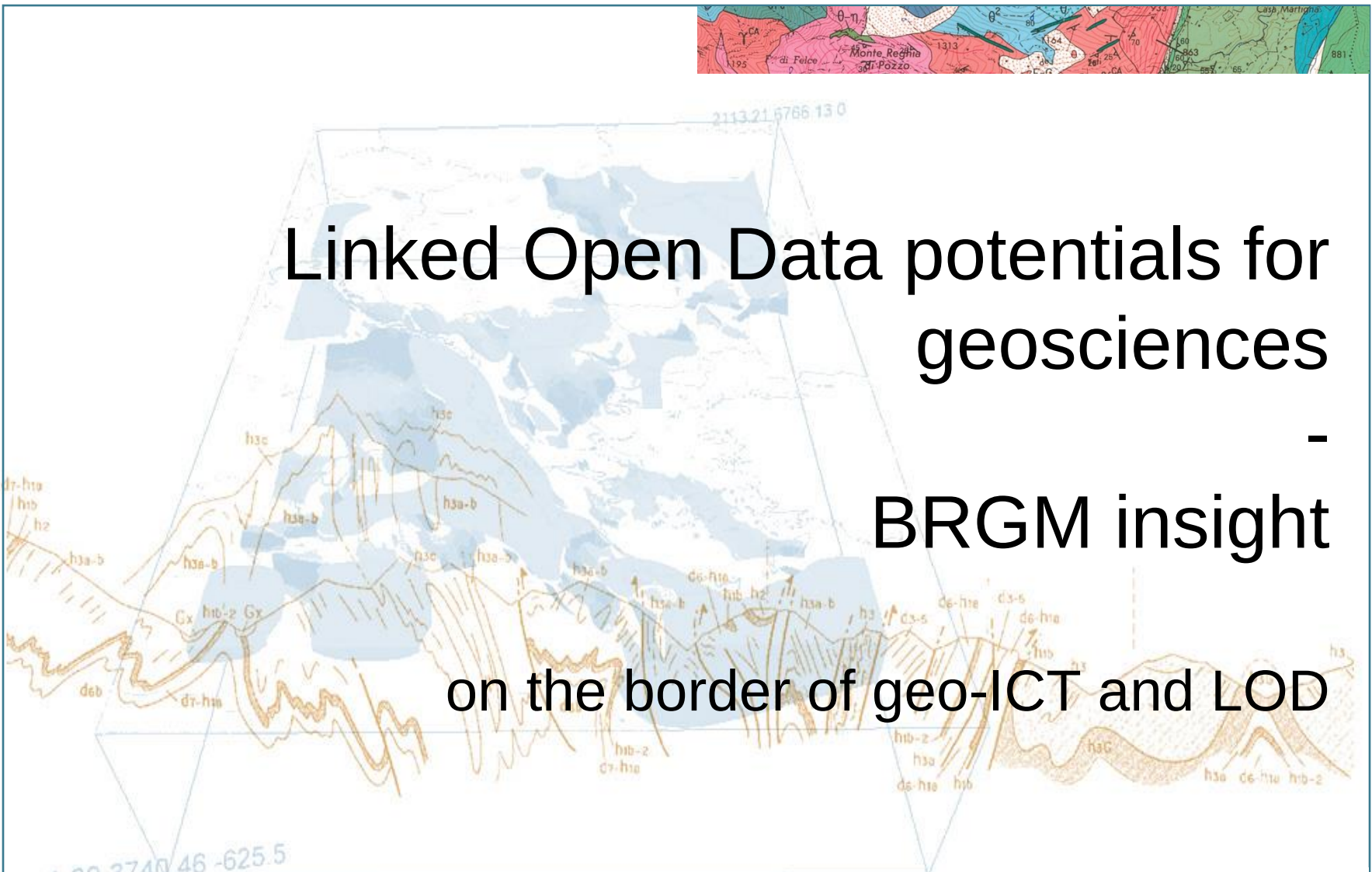




Linked Open Data potentials for geosciences

-
BRGM insight

on the border of geo-ICT and LOD

Starting point

> There is a geoscientific information exchange standardisation process going on

- Widespread : OGC, WMO, IUGS-CGI, Inspire ...
- Leads to exchange standards
 - Legally binding : Inspire, EU environmental directives reportings,
 - Community agreed :
 - International : GeoSciML, EarthResourceML, WaterML2.0, GroundWaterML2.0, generated by EU projects, ...
 - In France: Water Information System, Biodiversity Information system, COVADIS, ...
- Ensure interoperability
 - Semantic (not W3C meaning) : [Maieutics](#) work with domain expert (water, geology, ...) to describe their work (concepts, attributes, associations ...),
 - Technical : the way to transfer information => often OGC webservices (WFS, SOS, WCS, ...)

Starting point

> **Standardisation process**

- Model Driven Approach (MDA)
 - Domain model, independant of the implementation platform
- 0°/ Use Cases
- 1°/ Data models definition
 - With the experts => UML, ISO 19100 serie of standards
- 2°/ UML -> xsd (ISO : 19118, 19136, ...)
- 3°/ XML flows compliant to the xsd available using OGC webservice

> **Where do we put Linked Open Data here ?**

Activities around LOD

> **Standards aspects**

- OGC Geosemantics DWG / W3C joint group
 - How to generate reference ontologies ?
 - Standard methodology
 - Reference repository (ies ?) -> LinkedOpenVocabulary ?

> **Controlled vocabularies aspects**

> **URI**

> **Different representation of the same real world object**

Controlled vocabularies aspects

> Many attributes based on controlled vocab

```
<?xml version="1.0" encoding="UTF-8"?>
<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:sesame="
http://www.openrdf.org/schema/sesame#" xmlns:owl="http://www.w3.org/2002/07/owl#" xmlns:xsd="http://www.w3.org/2001/XMLSchema#" xmlns:fn="
http://www.w3.org/2005/xpath-functions#">
  <rdf:Description rdf:about="http://resource.geosciml.org/classifier/cgi/earth-resource-material-role/host-rock">
    <rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
    <definition xmlns="http://www.w3.org/2004/02/skos/core#" xml:lang="en">A body of rock serving as a host for other rocks or for mineral deposits; e.g.
a pluton containing xenoliths, or any rock in which ore deposits occur. It is a somewhat more specific term than country rock.</definition>
    <notation xmlns="http://www.w3.org/2004/02/skos/core#" rdf:datatype="gtwg:LocalHierarchyKey">03..</notation>
    <prefLabel xmlns="http://www.w3.org/2004/02/skos/core#" xml:lang="en">host rock</prefLabel>
    <source xmlns="http://purl.org/dc/elements/1.1/" xml:lang="en">Glossary of Geology. 5th Edition, AGI - online</source>
    <inScheme xmlns="http://www.w3.org/2004/02/skos/core#" rdf:resource="
http://resource.geosciml.org/classifierscheme/cgi/201403/earth-resource-material-role"/>
  </rdf:Description>
  <rdf:Description rdf:about="http://resource.geosciml.org/classifier/cgi/earth-resource-material-role/">
    <member xmlns="http://www.w3.org/2004/02/skos/core#" rdf:resource="
http://resource.geosciml.org/classifier/cgi/earth-resource-material-role/host-rock"/>
  </rdf:Description>
</rdf:RDF>

<mr-ext:material>
  <!-- Inlined feature 'M4EU.RM_22341' -->
  <mr-ext:RockMaterial gml:id="M4EU.RM_22341">
  </mr-ext:material>
  <mr-ext:proportion xsi:nil="true" nilReason="unpopulated"/>
</mr-ext:EarthResourceMaterial>
```

Controlled vocabularies

> Structured vocabularies

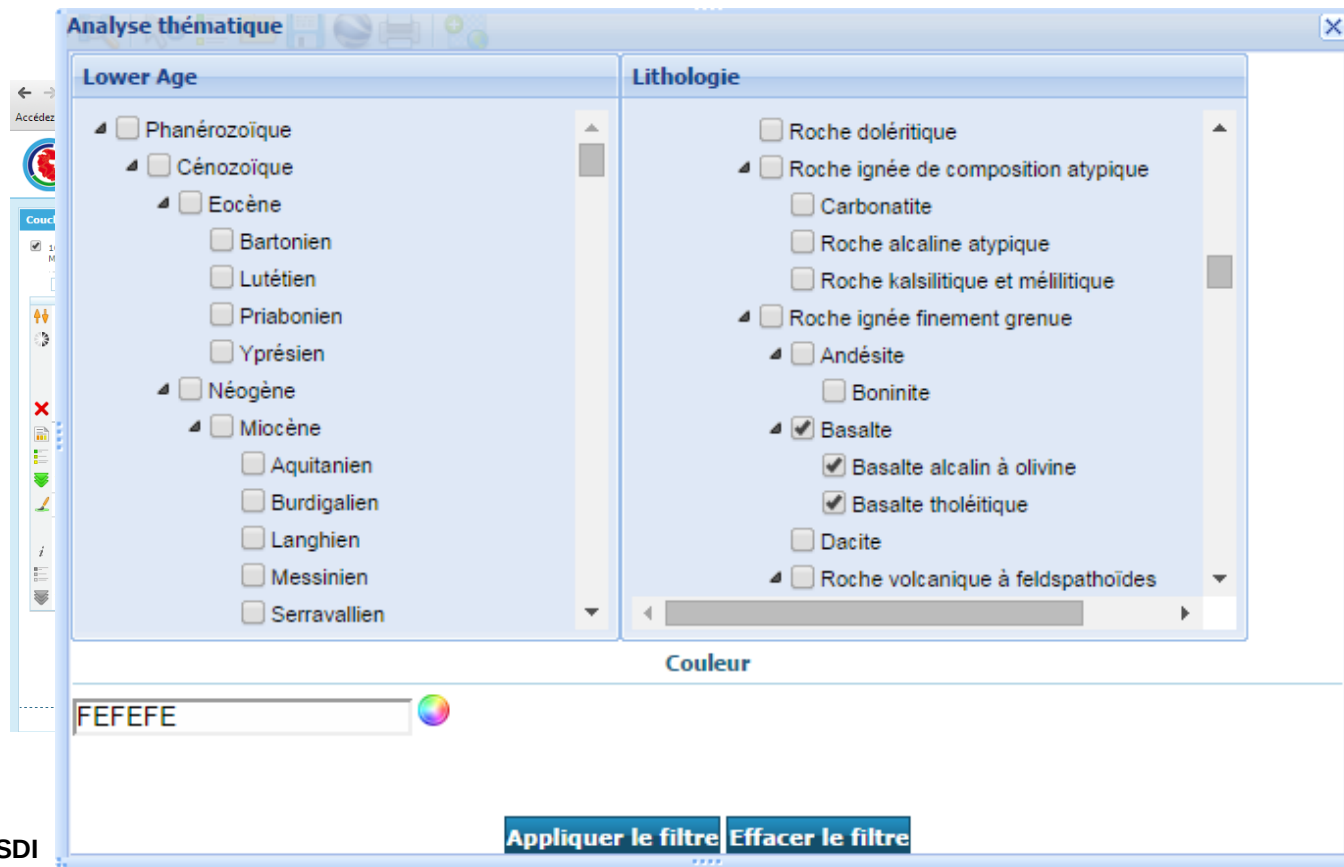
- Ex :Inspire Mineral Resources

```
<rdf:Description rdf:about="http://inspire.ec.europa.eu/codelist/LithologyValue/anthropogenicMaterial">
  <dc:identifier rdf:datatype="http://www.w3.org/2000/01/rdf-schema#Literal">anthropogenicMaterial</dc:identifier>
  <owl:sameAs rdf:resource="http://inspire.ec.europa.eu/codelist/LithologyValue/anthropogenicMaterial:1"/>
  <adms:last rdf:resource="http://inspire.ec.europa.eu/codelist/LithologyValue/anthropogenicMaterial"/>
  <rdf:type rdf:resource="http://purl.org/net/provenance/ns#DataItem"/>
  <rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
  <skos:inScheme rdf:resource="http://inspire.ec.europa.eu/codelist/LithologyValue"/>
  <skos:topConceptOf rdf:resource="http://inspire.ec.europa.eu/codelist/LithologyValue"/>
  <skos:broader rdf:resource="http://inspire.ec.europa.eu/codelist/LithologyValue/compoundMaterial"/>
  <dc:title xml:lang="fr">matériau anthropique</dc:title>
  <dc:description xml:lang="fr">Matériau dont on sait qu'il est d'origine artificielle (liée à l'homme); informations insuffisantes pour permettre une
classification plus précise.</dc:description>
  <skos:prefLabel xml:lang="fr">matériau anthropique</skos:prefLabel>
  <skos:definition xml:lang="fr">Matériau dont on sait qu'il est d'origine artificielle (liée à l'homme); informations insuffisantes pour permettre une
classification plus précise.</skos:definition>
  <adms:status rdf:resource="http://inspire.ec.europa.eu/registry/status/valid"/>
</rdf:Description>
```

Controlled vocabularies

> Use in OneGeology Europe

- Dynamic filters in OneGeology Europe (<http://onegeology-europe.brgm.fr/geoportal/viewer.jsp>) GeoSciML



Controlled vocabularies

> Use in OneGeology Europe

- Filter used to call OGC web services of every EU geological surveys

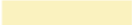
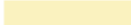
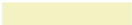
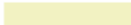
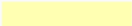
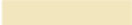
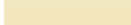
```
<sld:StyledLayerDescriptor version="1.0.0" xmlns="http://www.opengis.net/ogc" xmlns:sld="http://www.opengis.net/sld" xmlns:ogc="http://www.opengis.net/ogc" xmlns:gml="http://www.opengis.net/gml"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://www.opengis.net/sld http://schemas.opengis.net/sld/1.0.0/StyledLayerDescriptor.xsd">
```

```
<sld:NamedLayer>
  <sld:Name>OGE_1M_surface_GeologicUnit</sld:Name>
  <sld:UserStyle>
    <sld:FeatureTypeStyle>
      <sld:Rule>
        <ogc:Filter>
          <ogc:Or>
            <ogc:PropertyIsEqualTo>
              <ogc:PropertyName>gml:MappedFeature/gsml:specification/gsml:GeologicUnit/gsml:composition/gsml:CompositionPart/gsml:lithology@xlink:href</ogc:PropertyName>
              <ogc:Literal>urn:cgi:classifier:CGI:SimpleLithology:201001:basalt</ogc:Literal>
            </ogc:PropertyIsEqualTo>
            <ogc:PropertyIsEqualTo>
              <ogc:PropertyName>gml:MappedFeature/gsml:specification/gsml:GeologicUnit/gsml:composition/gsml:CompositionPart/gsml:lithology@xlink:href</ogc:PropertyName>
              <ogc:Literal>urn:cgi:classifier:CGI:SimpleLithology:201001:alkali-olivine_basalt</ogc:Literal>
            </ogc:PropertyIsEqualTo>
            <ogc:PropertyIsEqualTo>
              <ogc:PropertyName>gml:MappedFeature/gsml:specification/gsml:GeologicUnit/gsml:composition/gsml:CompositionPart/gsml:lithology@xlink:href</ogc:PropertyName>
              <ogc:Literal>urn:cgi:classifier:CGI:SimpleLithology:201001:tholeiitic_basalt</ogc:Literal>
            </ogc:PropertyIsEqualTo>
          </ogc:Or>
        </ogc:Filter>
        <sld:PolygonSymbolizer>
          <sld:Fill>
            <sld:CssParameter name="fill">#fefefe</sld:CssParameter>
          </sld:Fill>
        </sld:PolygonSymbolizer>
      </sld:Rule>
    </sld:FeatureTypeStyle>
  </sld:UserStyle>
</sld:NamedLayer>
</sld:StyledLayerDescriptor>
```

Controlled vocabularies

> Use in OneGeology Europe

- Data exploitation result adapted to the interface language (ex : number of lithology occurrences in a BBOX)

Legend and statistics on BBOX		Légende et statistiques sur BBOX			
		Choisissez un champ : Composition Part ▼			
Symbol	Lithology	Symbole	Lithologie	Proportion	Occu
	Organic bearing mudstone		Pélite riche en matière organique	subordinate	1
	Gravel		Gravier	subordinate	1
	Clay		Argile	subordinate	1
	missing		missing	subordinate	7
	Diamicton		Argiles à blocs	predominant	1
	urn:cgi:classif:CGI:Simple		urn:cgi:classif:CGI:SimpleLithology:201001:biogenic_silica_sedimentary_ro	subordinate	4
	Limestone		Calcaire	predominant	8
	Conglomerate		Conglomérat	subordinate	4
	Dolomite		Dolomie	subordinate	8
	Sand		Sable	subordinate	1
	Silt		Silt	subordinate	1
	Impure limestone		Calcaire impur	subordinate	8

Controlled vocabularies

> A harmonised approach is missing

- Reference vocab service standard (SPARQL end point) and structure (rdf) emerge slowly
- Tendency to
 - Externalise vocabulary from data models (xsd)
 - ⇒ Simplifies xsd maintenance but complexifies xml (GML) validation.
Because requires schematron (ex : EU Air Quality reporting)
 - Have xlink:href and let do the content negotiation
 - ⇒ But how do we validate that the entry being pointed at is still valid ?

> Will exposing information flows only using LOD solve those issues ?

URI

> On what ?

- Document / reports
- Datasets / an object within a dataset (even an observation)
- A controlled vocabulary / an entry within a controlled vocabulary,
- Metadata, feature catalogue, etc...

> Formatting Rules

- Completely structured / human readable VS 'uuid like'
- Do we point at resources using their label (tendency in LOD world) or their unique external identifier (more classical IT system) ?
 - Ex : there are 5 different river stretches called 'Loire' in France; only one is the 1000 km long one !
- If « structured » choosen, do we include versioning ?
 - bad from W3C 10 Rules for Persistent URIs

URI

> InspireId / DOI ?

- We need to expose InspireID : important to only do the job once
- Issue : organisation can change name (thus IT domain)
 - ⇒ Can Digital Object Identifier cover part of our needs ?
 - ⇒ try and test DOI (ISO 26324:2012, Information and documentation -- Digital object identifier system)
- DOI points to a URI in our system
- DOI needs a landing page

URI

> Pattern being tested

<http://dx.doi.org/10.BRGM.providerid/xxxx>

Resolve n°1

[http://unique BRGM entry point/
{complicated but brilliant URI
structure} / xxxx](http://unique BRGM entry point/{complicated but brilliant URI structure} / xxxx)

Resolve n°2

One (more?)
representation of the
ressource

[XXXX](#)

- New issues :
 - what external id do we communicate on ? The DOI or the BRGM one ?
 - on which representation of the ressource do we land on ?

Different representations of the same real world object

> LOD is one representation amongst other

> How do we implement it within our systems ?

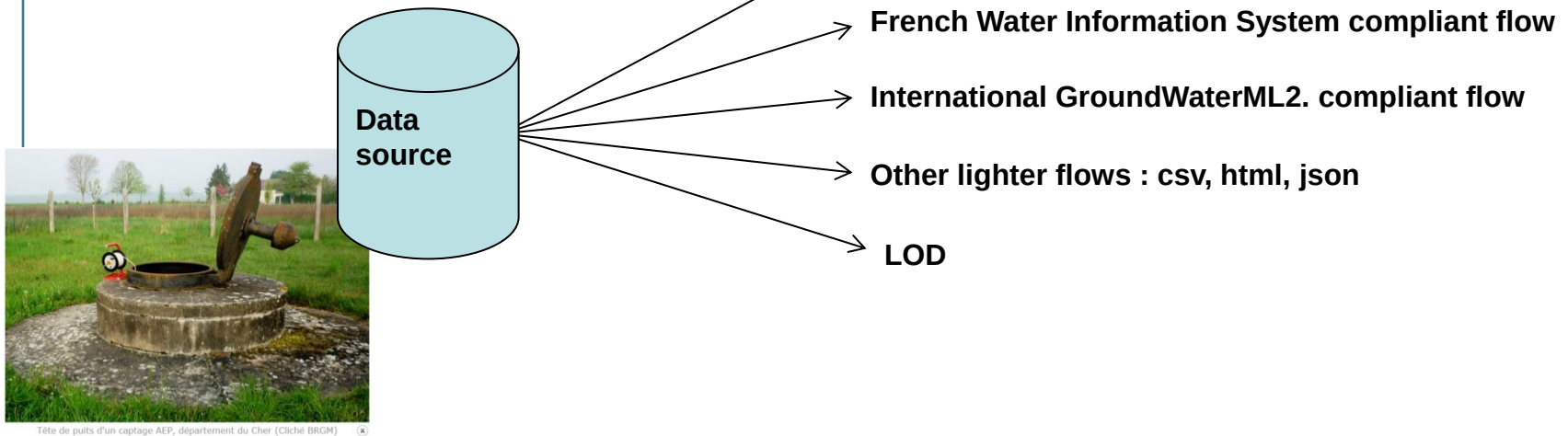
- Without duplicating information between databaseS / triplestores, ...
- Most of the time we have a data storage different from the exchange scenario (national/EU/..., csv/rdf/xml, ...)

> How do we have a precise mapping between ontologies ?

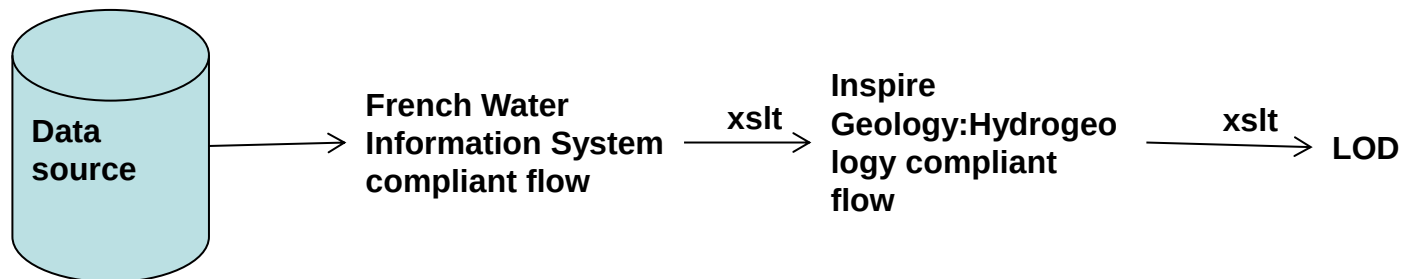
- Devil lies in the detail => is not always 'same as'.
- Need to be able to detail the 'what is not included in the same as'

Different representation of the same real world object

- We want this



- And not a chain like



Conclusion

- > Still on the process of evaluating the potentials of LOD**
- > LOD often**
 - Sold as the “magic black box”, highly flexible
 - Put back to back with the current (rigid?) standardisation approach
- > But organisational aspects remain central**
 - « Reference ontologies », « Ontologie curators »
 - URIs structure, resolving mechanism
- > Both ways can complement each others. We try to have them combined together**

s.grellet@brgm.fr

