

‘Linked data’ the glue within interoperable information systems

Sylvain Grellet, Abdelfettah Feliachi

▶ **To cite this version:**

Sylvain Grellet, Abdelfettah Feliachi. ‘Linked data’ the glue within interoperable information systems. INSPIRE Conference 2020, Jun 2020, Dubrovnik, Croatia. hal-03499535

HAL Id: hal-03499535
https://brgm.hal.science/hal-03499535

Submitted on 21 Dec 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



‘LINKED DATA’ THE GLUE WITHIN INTEROPERABLE INFORMATION SYSTEMS

ABDELFETTAH FELIACHI
SYLVAIN GRELLET

Geoscience for a sustainable Earth
brgm

CONTEXT

‘Previous generation’ of OGC APIs (ex: WFS, SOS)

based on design from the 90's, 2000 (RPC, SOAP)

Brought SDIs to the web but

- Are opaque to the web (and crawlers)
  - need to know the catalogue where to search for the API end-point
  - need to know how to interact with the such catalogue
  - when the API end-point is found, need to RTFM of the API spec

=> We expose(d) data in a siloed way

Idea n°1: Data is naturally linked
-> why not linking data content ?

(1) ? LINKING DATA CONTENT ?

Still using ‘Previous generation’ OGC APIs

- Assigning HTTP URIs
  - Applying W3C best practices (Data on the web, Spatial Data on the web)

- Links
  - Between features
    - Station to river,
    - AquiferSystem to its part, …
  - To controlled vocabularies

(1) ? LINKING DATA CONTENT ?

- Example

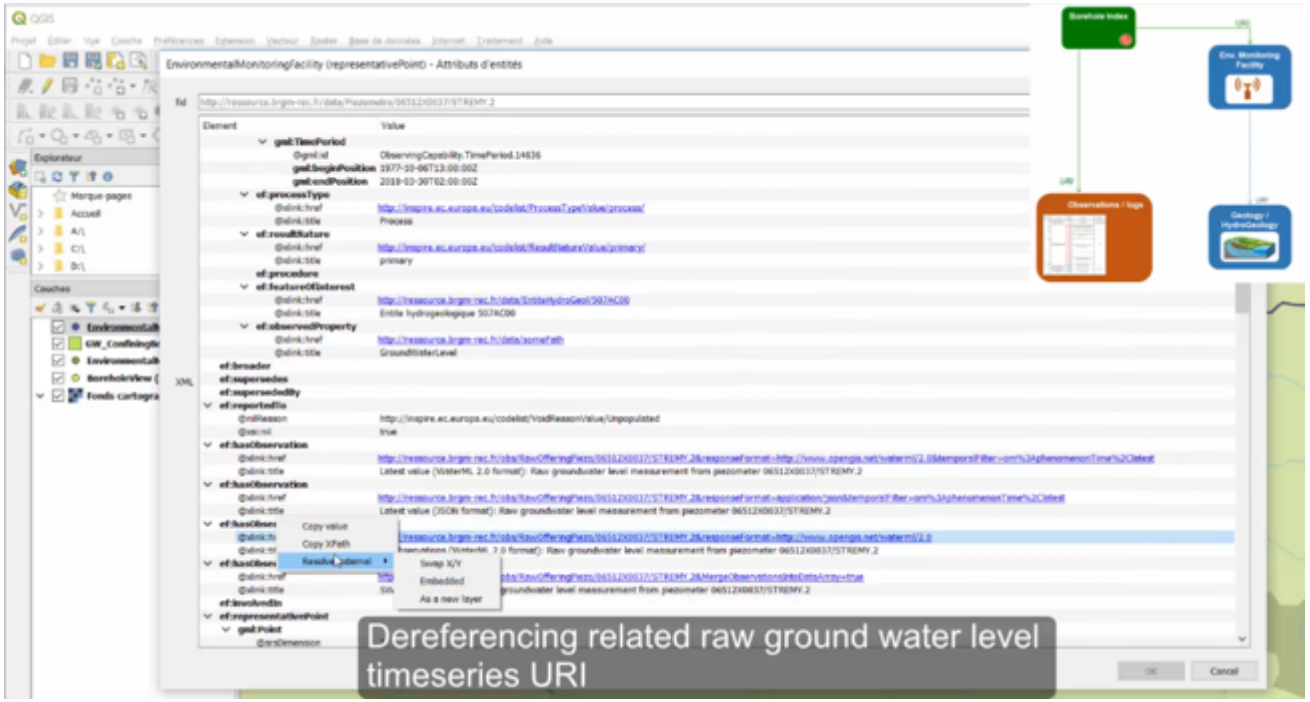
```
<gwml2:GW_AquiferSystem gml:id="EntiteHydroGeol.121AS">
  <gml:description>Aquifer from national BD LISA (Base de Données des Limites des Systèmes Aquifères) database</gml:description>
  <gml:identifier codeSpace="http://www.ietf.org/rfc/rfc2616">https://data.geoscience.fr/id/hydrogeounit/121AS</gml:identifier>
  <gml:name codeSpace="http://id.eaufrance.fr/saq/2/">Craie du Sénonien au Turonien inférieur du Bassin Parisien du Neubourg au Saint-André - bassin versant de l'Eure, Iton, Avre (bas
  <gsmlb:observationMethod>
    <swe:Category>
  </gsmlb:observationMethod>
  <gsmlb:occurrence>
  <gsmlb:purpose>instance</gsmlb:purpose>
  <gsmlb:geologicUnitType xlink:href="http://resource.geosciml.org/classifier/cgi/geologicunittype/hydrogeologic_unit" xlink:title="Hydrogeologic Unit"/>
  <gsmlb:rank xlink:href="http://id.eaufrance.fr/nsa/352#2" xlink:title="Régional"/>
  <gwml2:gwUnitMedia xlink:href="http://id.eaufrance.fr/nsa/353#9" xlink:title="Triple porosité : matricielle et karstique et de fractures"/>
  <gwml2:gwAquiferSystem xlink:href="https://data.geoscience.fr/id/hydrogeounit/121" xlink:title="Grand système multicouche du Campanien au Turonien (Séno-Turonien) du Bassin Pari
  <gwml2:gwAquiferSystemPart xlink:href="https://data.geoscience.fr/id/hydrogeounit/121AS01" xlink:title="Craie du Séno-Turonien du Bassin Parisien du Neubourg au Saint-André - bass
Seine-Normandie)"/>
  <gwml2:gwAquiferSystemPart xlink:href="https://data.geoscience.fr/id/hydrogeounit/121AS30" xlink:title="Craie marneuse et marnes du Turonien inférieur du Bassin Parisien du Neubou
l'Eure, Iton, Avre (bassin Seine-Normandie)"/>
  </gwml2:GW_AquiferSystem>
```

URI to Feature

URI to codeList

(1) ? LINKING DATA CONTENT ?

- From this you can already build nice things



See :
https://github.com/BRGM/gml_application_schema_toolbox/tree/master/presentations/2018_INSPIRE_conference

Even between ‘Previous generation’ OGC APIs that’s already a better experience

(1) ? LINKING DATA CONTENT ?

Still using ‘Previous generation’ OGC APIs

- Data content starts to open up
  - -> start to cross silos
  - -> that’s already a better experience

- But
  - API are still opaque to the web
  - Data semantics unaccessible to the web of data

Idea n°2: Why not serving data in a web oriented approach ?

(2) ? SERVING DATA IN A WEB ORIENTED APPROACH ?

New OGC APIs

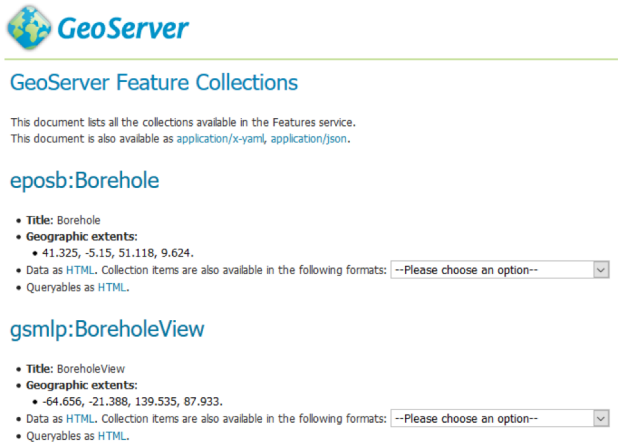
- Important turn/change in OGC specs
  - Web APIs, sometimes referred to as RESTful services.
  - Each API instance described by a documentation that is meaningful for the Web
  - No retrocompatibility is not an issue
  - Targetting « JSON family » (often (Geo)JSON) representation

- Hypermedia (HATEOAS)
  - Within API operation
  - Within data structure returned: perfect example of ST API response
  - -> A client could work by following links between resources, that's the way of the ~~modern~~ Web

(2) ? SERVING DATA IN A WEB ORIENTED APPROACH ?

New OGC APIs

- OGC - API - Features




```
{
  type: "FeatureCollection",
  features: [
    {
      type: "Feature",
      id: "0001000001",
      geometry: {
        type: "Point",
        coordinates: [
          [redacted],
          [redacted]
        ]
      },
      properties: {
        "@featureType": "Borehole",
        identifier: {
          value: "BSS000AAAA",
          "@codeSpace": http://www.ietf.org/rfc/rfc2616
        },
        type: {
          "@href": http://vocabulary.odm2.org/samplingfeaturetype/borehole/,
          "@title": "borehole"
        },
        sampledFeature: [
          {
            "@href": https://sweet.jpl.nasa.gov/2.3/realmEarthReference.owl#EarthLithosphere,
            "@title": "Lithosphere"
          }
        ],
        shape: {
          "@dataType": "shape",
          Point: {
            type: "Point",
            coordinates: [
              [redacted],
              [redacted]
            ]
          }
        },
        custodian: {
          "@dataType": "Custodian",
          custodian: {
            "@dataType": "CI_ResponsibleParty",
            organisationName: "BRGM",
            role: {
              "@codeListValue": "custodian",
              "@codeList": http://www.isotc211.org/2005/resources/Codelist/gmxCodelists.xml
            }
```

(2) ? SERVING DATA IN A WEB ORIENTED APPROACH ?

New OGC APIs

- SensorThings API


```
{
  "@iot.nextLink": "https://sensorthings-wq.brgm-rec.fr/FROST-Server/v1.0/Datastreams?$skip=100",
  "value": [
    {
      "name": "Dichloropropène-2,3 in Eau on Eau brute at LA GRANDE NOUE D'HERMÉ A GOUAIX 1 with method Qualité de l'eau - Dosage par chromatographie en phase gazeuse d'un certain nombre d'hydrocarbures aromatiques monocycliques, du naphtalène et de divers composés chlorés par dégazage, piégeage et désorption thermique (T90-129/NF EN ISO 15680 01/2004)",
      "description": "Dichloropropène-2,3 in Eau on Eau brute at LA GRANDE NOUE D'HERMÉ A GOUAIX 1 with method Qualité de l'eau - Dosage par chromatographie en phase gazeuse d'un certain nombre d'hydrocarbures aromatiques monocycliques, du naphtalène et de divers composés chlorés par dégazage, piégeage et désorption thermique (T90-129/NF EN ISO 15680 01/2004)",
      "properties": {↔},
      "observationType": "http://www.opengis.net/def/observationType/OGC-OM/2.0/OM_Measurement",
      "observedArea": {
        "type": "Point",
        "coordinates": [
          48.457967041,
          3.296382646
        ]
      },
      "phenomenonTime": "2010-01-04T00:20:00.000Z/2011-11-02T11:55:00.000Z",
      "Sensor@iot.navigationLink": "https://sensorthings-wq.brgm-rec.fr/FROST-Server/v1.0/Datastreams(1142553)/Sensor",
      "Thing@iot.navigationLink": "https://sensorthings-wq.brgm-rec.fr/FROST-Server/v1.0/Datastreams(1142553)/Thing",
      "Observations@iot.navigationLink": "https://sensorthings-wq.brgm-rec.fr/FROST-Server/v1.0/Datastreams(1142553)/Observations",
      "unitOfMeasurement": {↔},
      "ObservedProperty@iot.navigationLink": "https://sensorthings-wq.brgm-rec.fr/FROST-Server/v1.0/Datastreams(1142553)/ObservedProperty",
      "@iot.id": 1142553,
      "@iot.selfLink": "https://sensorthings-wq.brgm-rec.fr/FROST-Server/v1.0/Datastreams(1142553)"
```

(2) ? SERVING DATA IN A WEB ORIENTED APPROACH ?

New OGC APIs

- APIs endpoint can be automatically discovered (by machines)
- But
  - Data semantics unaccessible to the web of data

Idea n°3: Why not expressing data semantics differently ?

(3) ? EXPRESSING DATA SEMANTICS DIFFERENTLY ?

Using web of data oriented semantics

- OGC ELFIE / SELFIE (https://opengeospatial.github.io/ELFIE/)
  - 2018 – 2019 ELFIE: OGC and W3C standards with the goal of establishing a best practice for exposing links between and among environmental domain and sampling features in a highly adaptible standards compliant way that is compatible with modern web search technology.
  - 2019 – 2020 SELFIE: designing and vetting Web-resource model and network behavior for cross-domain linked feature data that compliments and uses new OGC APIs as a building block.

- JSON-LD serialization/model of choice

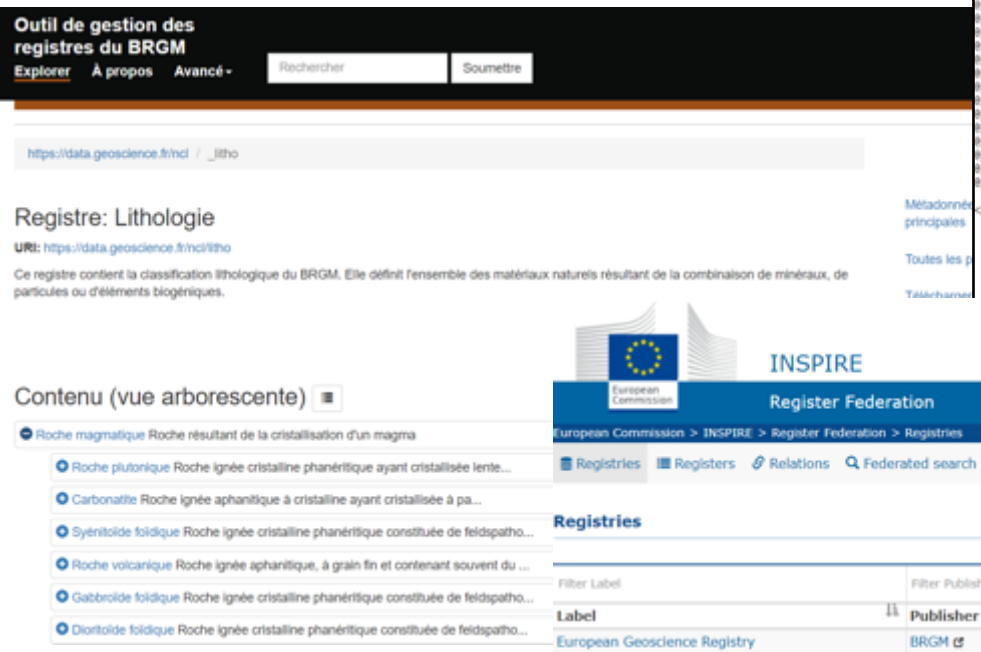
- Semantics
  - Schema.org for indexing by search engines
  - Complemented by ontologies stemming from OGC, INSPIRE semantics, W3C, …

- Content negotiation on HTTP URIs

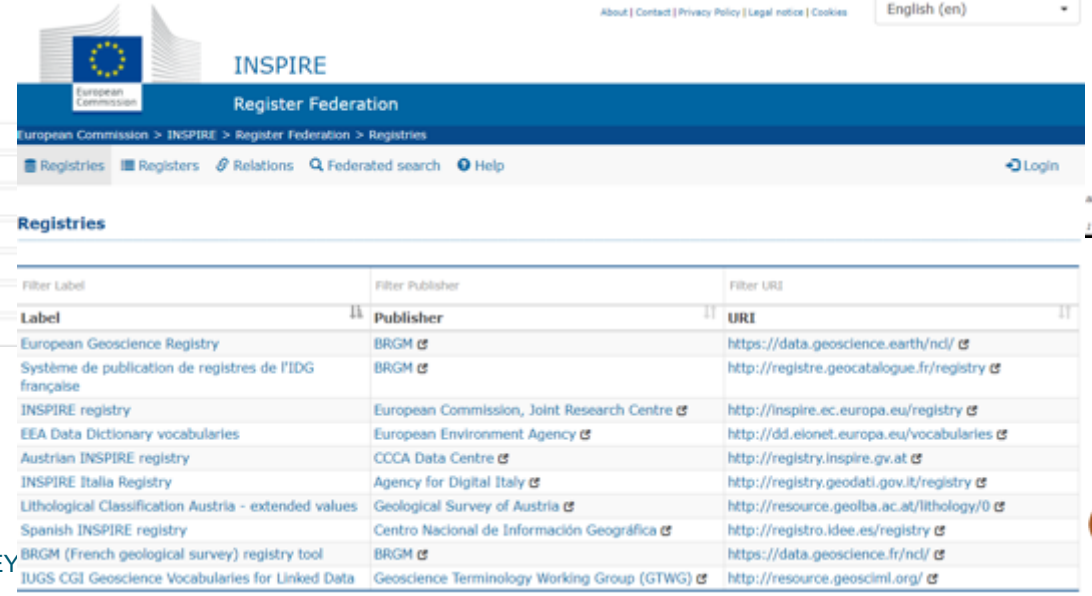
(3) ? EXPRESSING DATA SEMANTICS DIFFERENTLY ?

Using web of data oriented semantics

- Controlled vocabulary Registry
  - Being federated at the EU level







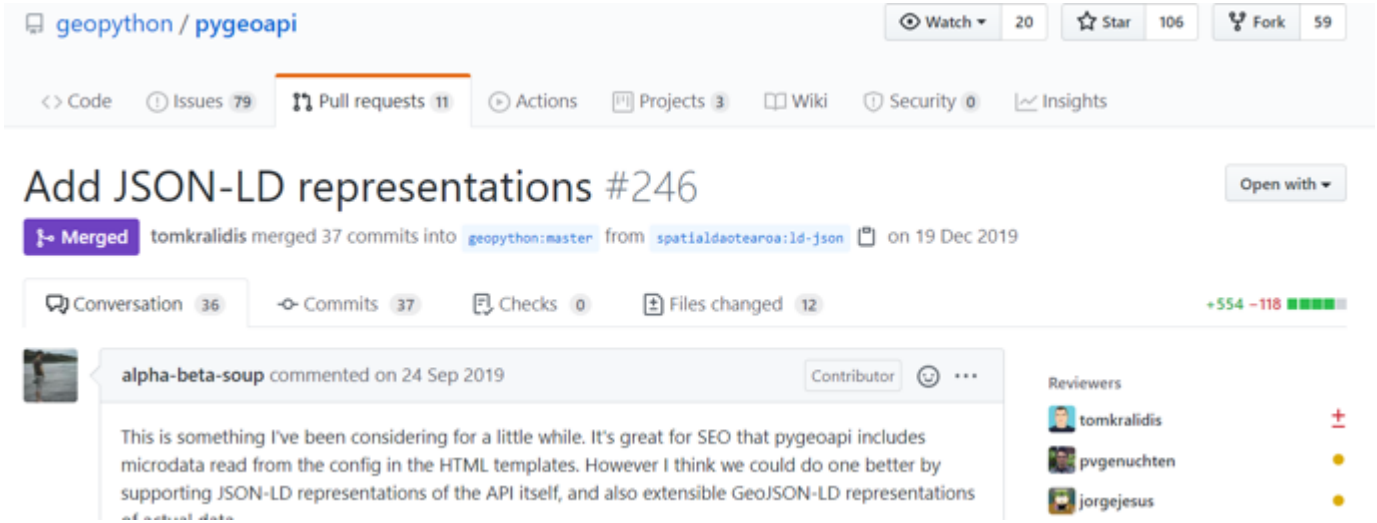
Geoscience for a sustainable Earth
brgm

(3) ? EXPRESSING DATA SEMANTICS DIFFERENTLY ?

Using web of data oriented semantics

- Feature API support



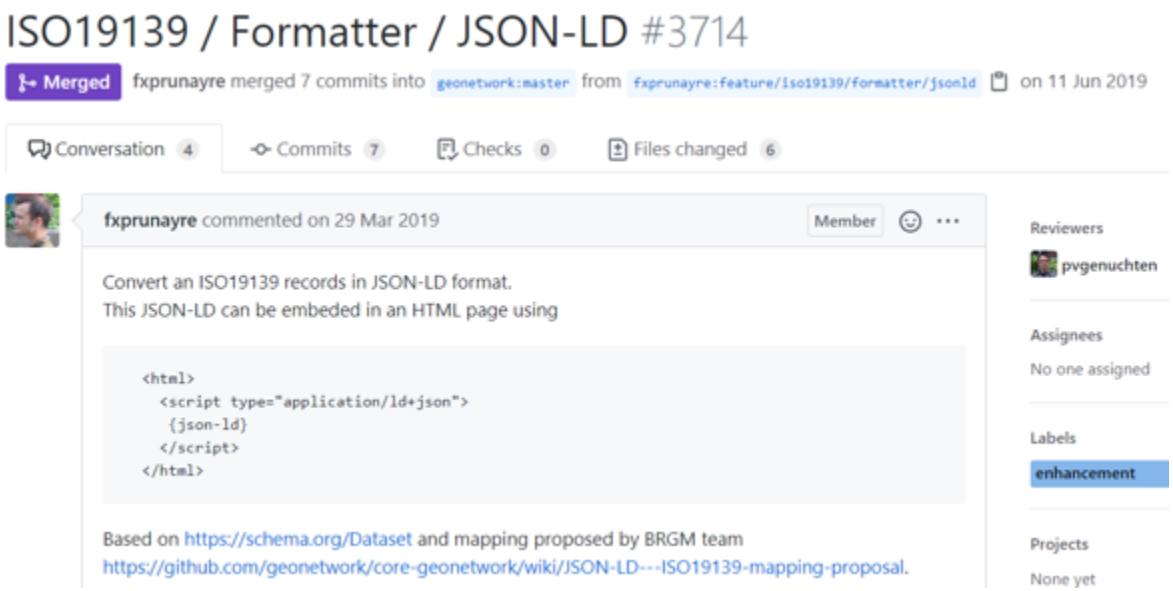


brgm
Geoscience for a sustainable Earth

(3) ? EXPRESSING DATA SEMANTICS DIFFERENTLY ?

Using web of data oriented semantics

- Catalogue API support

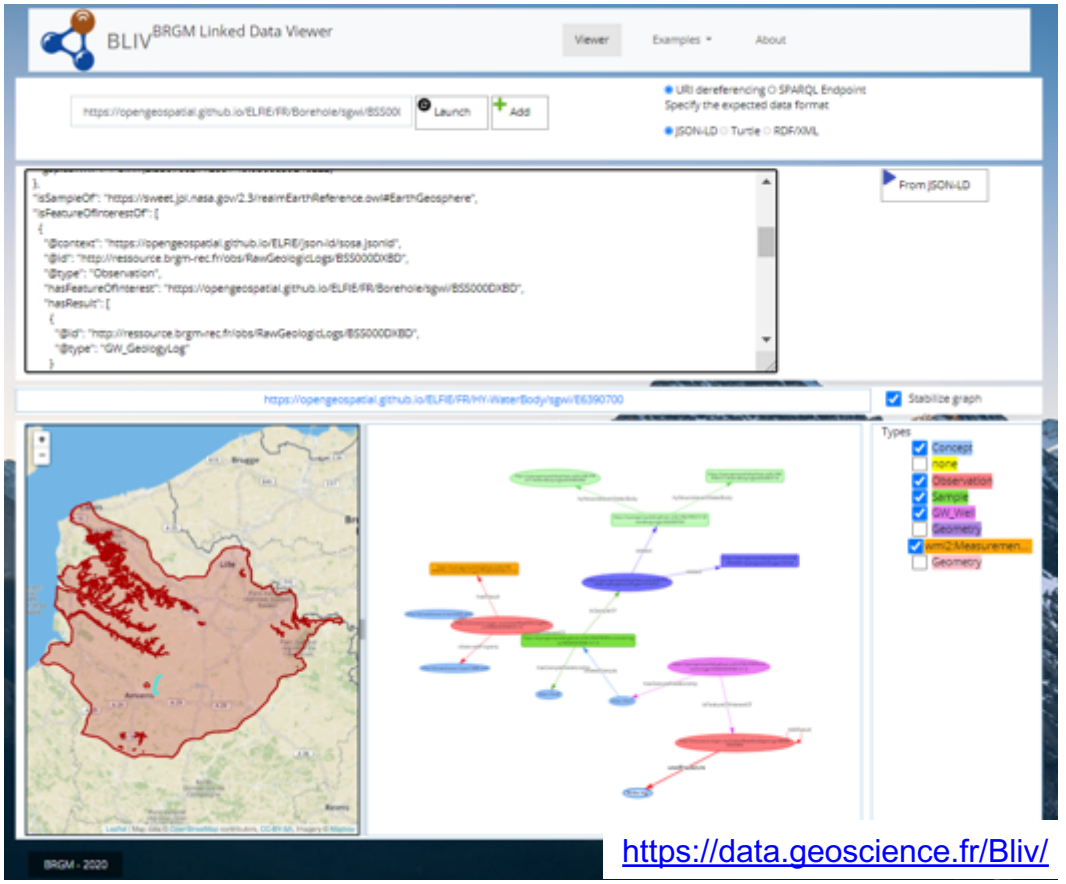


- SensorThings API
  - Discussed in SELFIE

(3) ? EXPRESSING DATA SEMANTICS DIFFERENTLY ?

Using web of data oriented semantics

- Linked data client



https://data.geoscience.fr/Bliv/

- Benefits
  - Data instances not hidden anymore behind the API (as with previous OGC APIs)

  -> data instances can be indexed by search engines

  -> more visibility (data back on the 1st line), more reuse of public environmental data/assets

GENERALISATION

What a journey !

- Previous era/decade was bringing SDI's to the Web
  - INSPIRE dynamics abide by this (previous OGC APIs, INSPIRE network services guidelines)
- Now SDI's (or environmental IS) are starting to contribute to the web of data
  - Thanks to Linked Data & SemanticWeb
- Note : could have detailed "Metadata" dynamics as well OGC API Record, DCAT
  - But other presentations in that session speak about this
- OGC, INSPIRE related ontologies
  - Generated during (S)ELFIE, being published by OGC-NA
- HTTP URIs on everything
  - Features, codeList entries, APIs, dataset … to link everything

CONCLUSION

Introducing LinkedData and SemanticWeb in Information Systems

- Data providers still provide their data
- in a way that allows both APIs and data to be discovered and linked to

-> more reuse

-> less loss of time searching for, trying to understand data instead of producing new domain content

-> more domain experts doing their work



LINKS COMPENDIUM

- W3C Best practices
  - Data on the Web Best Practices : https://www.w3.org/TR/dwbp/
  - Spatial Data on the Web Best Practices : https://www.w3.org/TR/sdw-bp/
- QGIS GMLAS Toolbox
  - https://github.com/BRGM/gml_application_schema_toolbox
- OGC Linked data Interoperability Experiments
  - ELFIE : https://opengeospatial.github.io/ELFIE/
  - SELFIE: https://github.com/opengeospatial/SELFIE
- Registry tool - UkGovLd
  - https://github.com/UKGovLD
- INSPIRE register federation
  - https://inspire-regadmin.jrc.ec.europa.eu/ror/
- GeoServer JSON-LD community plugin
  - https://docs.geoserver.org/latest/en/user/community/json-ld/index.html
- PyGeoApi – JSON-LD
  - https://github.com/geopython/pygeoapi/pull/246
- Geonetwork – JSON-LD
  - https://github.com/geonetwork/core-geonetwork/pull/3714

THANK YOU



a.feliachi@brgm.fr
s.grellet@brgm.fr

Géosciences pour une Terre durable
brgm