HAL
open science

# Making INSPIRE data discoverable and findable through popular search engines.

Abdelfettah Feliachi, Sylvain Grellet, Thierry Vilmus

## To cite this version:

Abdelfettah Feliachi, Sylvain Grellet, Thierry Vilmus. Making INSPIRE data discoverable and findable through popular search engines.. CONFERENCE INSPIRE 2018, Sep 2018, Anvers, Belgium. hal-02009234

**HAL Id: hal-02009234**

**https://brgm.hal.science/hal-02009234v1**

Submitted on 6 Feb 2019

**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.

# MAKING INSPIRE DATA DISCOVERABLE AND FINDABLE THROUGH POPULAR SEARCH ENGINES

## EXPERIMENTATION ON FRENCH GEOCATALOGUE

*A. FELIACHI, S. GRELLET AND T.VILMUS*

INSTITUT CARNOT brgm

Geoscience for a sustainable Earth
brgm

# IT CONTEXT

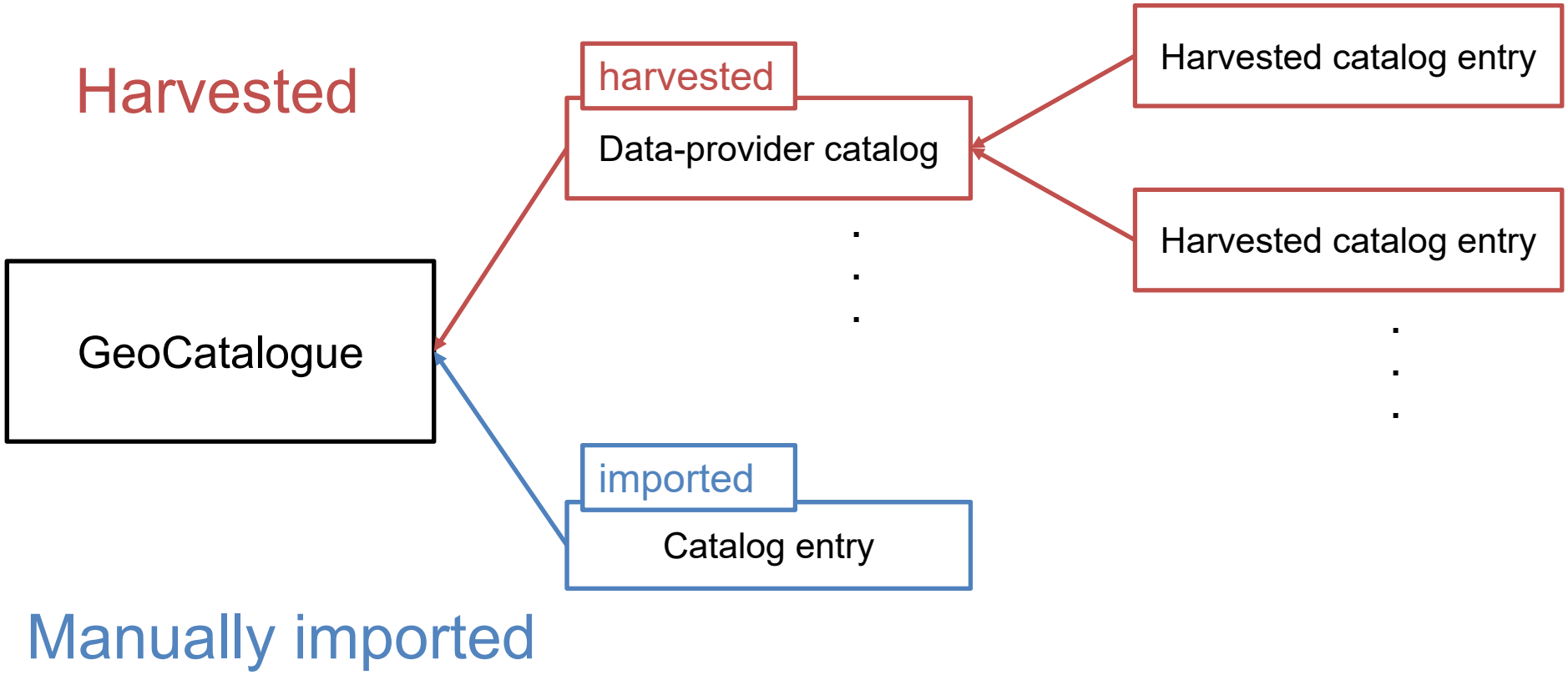
- BRGM, French geological survey, is implementing the national INSPIRE catalogue, named Geocatalogue

- It's hard to find datasets
  - Difficulties to find data through Inspire specialized search engines like Geoportals or Geocatalogues
  - General public even unaware of the existence of such tools

- How to help search engine index those datasets ?
  - Vocabulary : Schema.org
    - Proposed by important search engines Google, Microsoft, Yahoo and Yandex
  - Payload
    - JSON-LD
    - Or embedded in HTML pages

## FRENCH CONTEXT

- Metadata flow into Geocatalogue



# FRENCH CONTEXT

- Testing it on the French GeoCatalogue
  - Identify schema.org tags to be used
    - Building on feedback from previous experience:

      https://www.w3.org/2015/spatial/wiki/ISO_19115_-_DCAT_-_Schema.org_mapping

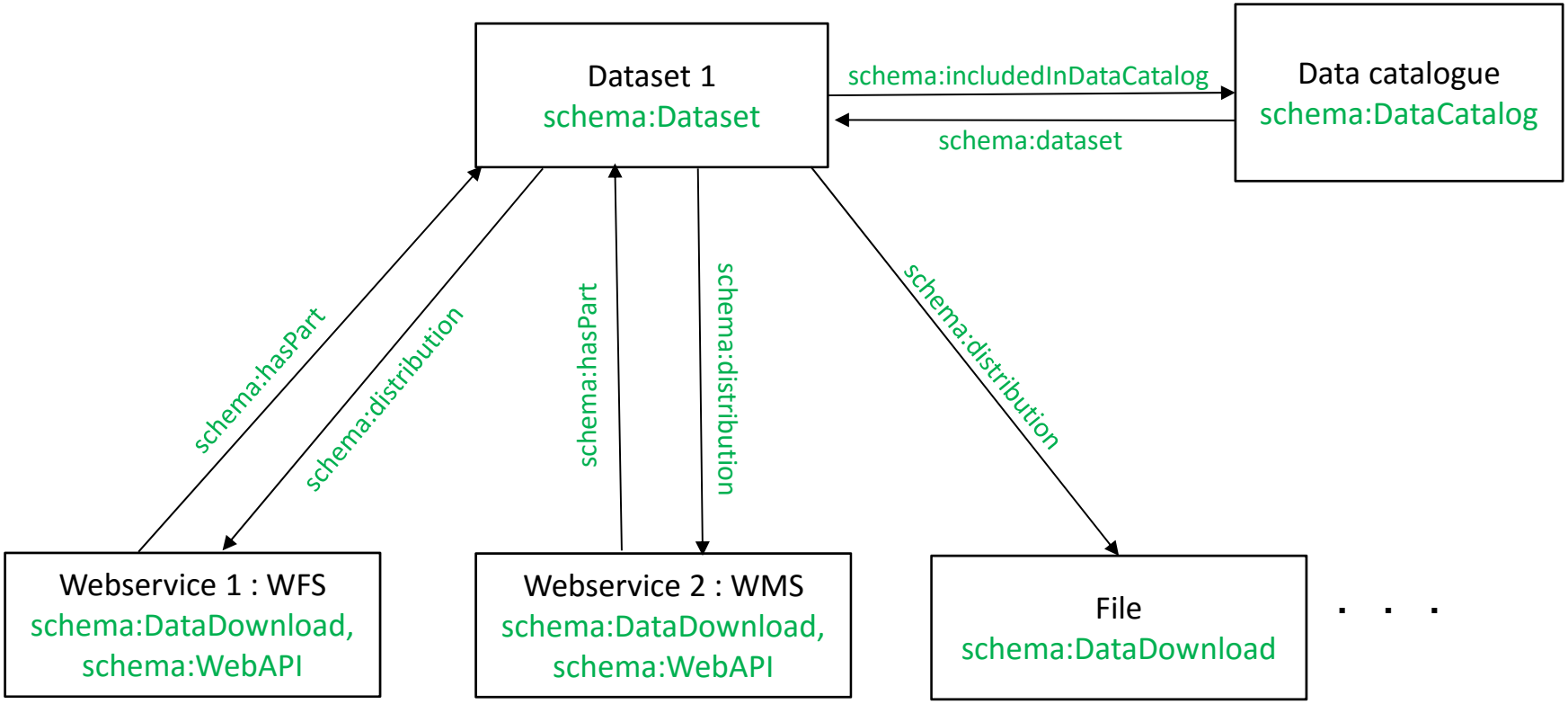
      https://ec-jrc.github.io/dcat-ap-to-schema-org/

      http://geocat.fr/dataset-prop.html

  - Define a national URI architecture
    - The exercise brings back the importance of a clean URI approach (data provider and national level)
    - Taking into account the 3 types of data providers
      - Harvested by the national catalogue
        - with a URI policy (= defined and that resolves)
        - with no URI policy
      - Not having a SDI and which metadata records are manually imported into the national catalogue (thus no data provider URI policy)

# PROPOSED DATA STRUCTURE

- Generic JSON-LD approach
- Use schema:dataset & schema:includedInDatacatalog to link catalogues and datasets
- Use schema:distribution to declare services



# PROPOSED DATA STRUCTURE

- JSON-LD examples (dataset, catalogue & service)

Catalogue

```
{
   "@context": "http://schema.org/",
   "@type": "DataCatalog",
   "@id": "https://data.geoscience.fr/id/catalogue/BRGM",
   "name": {value":"BRGM Data Catalog", "@language": "en"},
   "description": "BRGM metadata catalog",
                 …..
   "dataset":["https://data.geoscience.fr/id/dataset/borehole", … ]
                 …..
   "about": ["https://www.eionet.europa.eu/gemet/en/inspire-
theme/ge", … ],
…..
}
```

Dataset

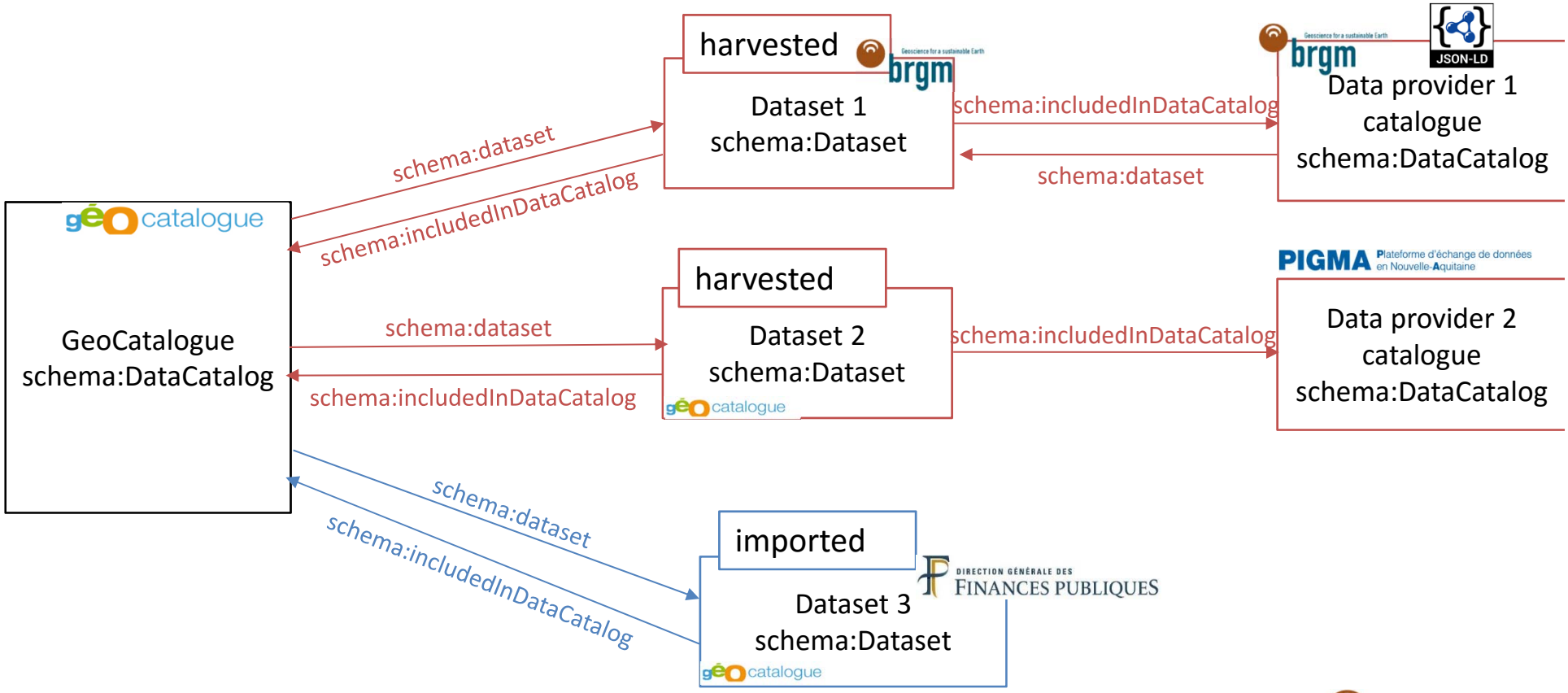
```
{
   "@context": "http://schema.org/",
   "@type": "Dataset",
   "@id": "https://data.geoscience.fr/id/dataset/borehole",
   "includedInDataCatalog": "https://data.geoscience.fr/id/
       catalogue/BRGM",
   "name": {value":"Borehole",  "@language":"en"},
                 …
   "distribution": [
      { "@id":"https://data.geoscience.fr/api/wfs/borehole",
       "@type":["DataDownload","WebAPI"],
       "contentUrl":"http://geoservices.brgm.fr"} … ],
…..
}
```

Service

```
{
   "@context": "http://schema.org/",
   "@id":"https://data.geoscience.fr/api/wfs/borehole",
   "@type":["DataDownload","WebAPI"],
   "name": "Borehole WFS Service",
                 …..
    keywords":[
    {"@value":"Forage", "@language":"fr"},…],
                 ……
"spatialCoverage":{"@type":"Place",
"geo": {"@type":"GeoShape",
"box":[     "-5.79028,41.36493 9.56222,51.09111",
            "-61.7961,15.87 -61.1871,16.5129",
            "-61.2315,14.4028 -60.817,14.8801",
            "-54.6038,2.11347 -51.6481,5.75542",*
            "55.2206,-21.3739 55.8531,-20.8565",
            "45.0392,-12.9925 45.2297,-12.6625"]
            }
 }, ……
```

# PROPOSED DATA STRUCTURE

- Not all data provider have a URI policy that resolves to a well defined JSON-LD representation
- Example below



# URIS IN THE PICTURE

- Use persistent URI to identify catalogues, datasets and services

- Rationale
  - For data provider having a URI policy that resolves in JSON-LD : respect it

  - For the others

  Define a national pattern

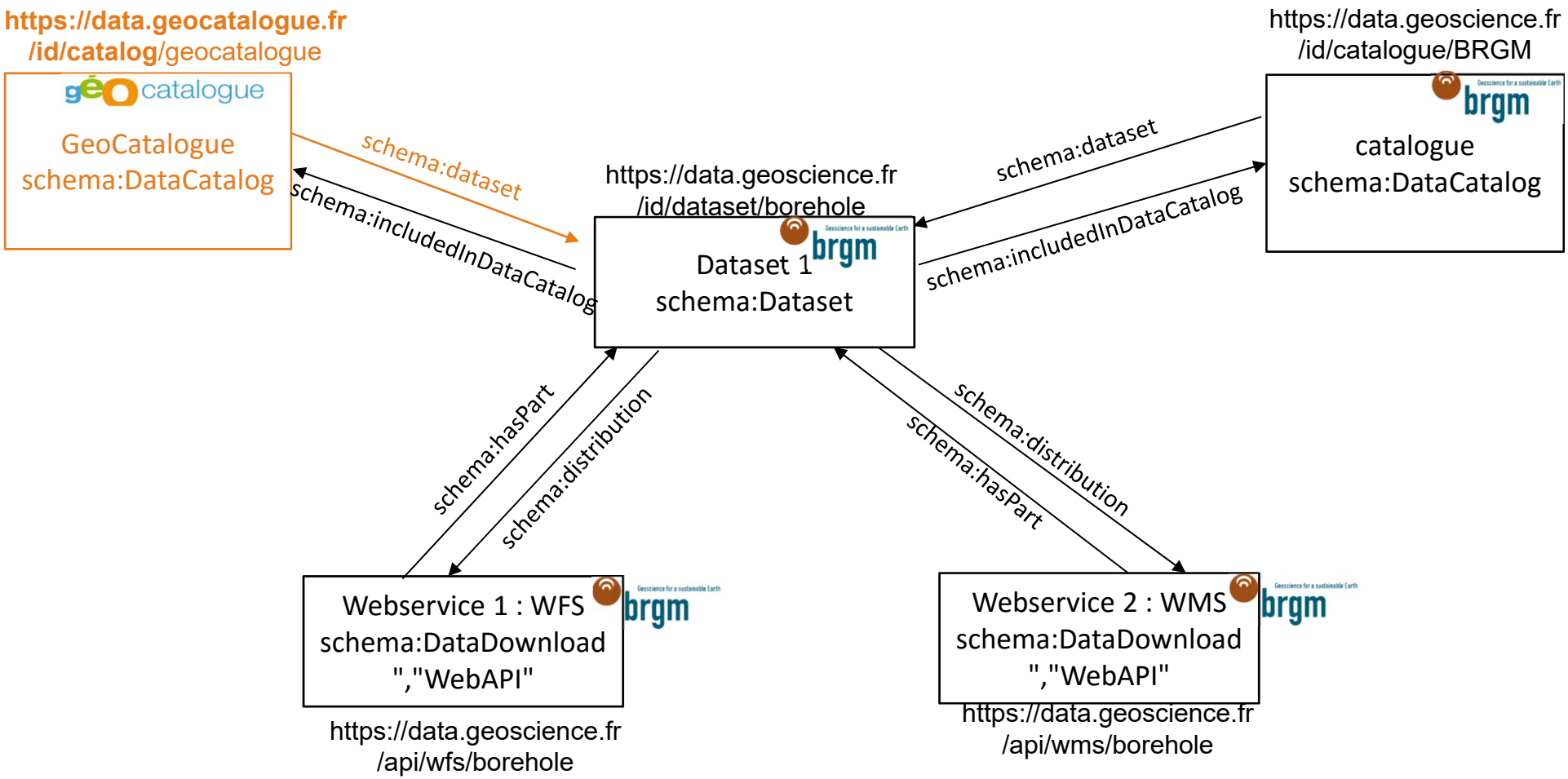
  Data catalogue : https://data.geocatalogue.fr/id/catalog/{data_provider_catalogue_id}

  Dataset : https://data.geocatalogue.fr/id/dataset/{data_provider_id}_{geocatalogue_defined_uuid}

  →When those start having a URI policy that resolves in JSON-LD have a HTTP 301 ('Moved Permanently') from the previous URI to the new one

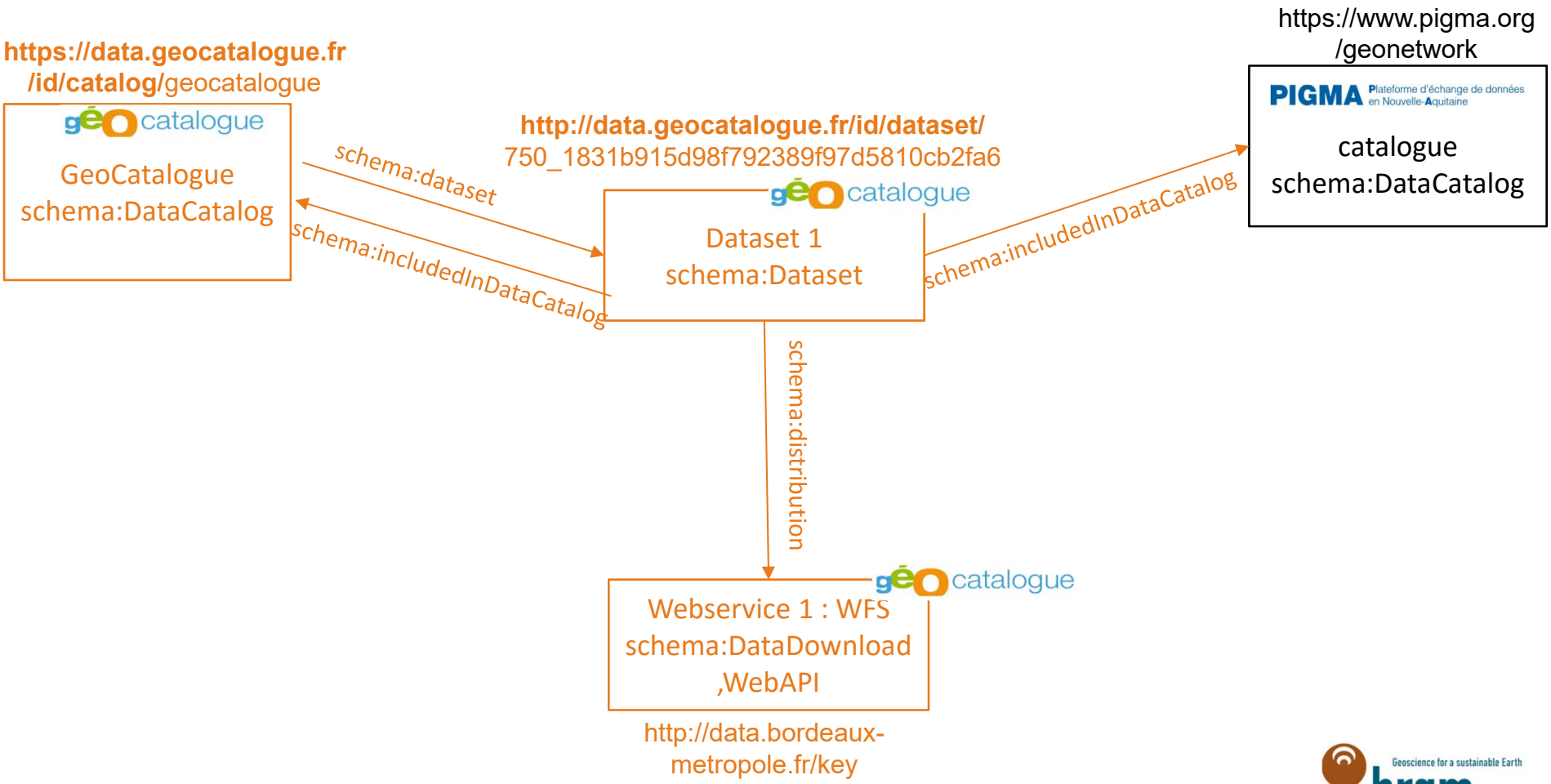
# URIS – APPLIED TO THE DATA STRUCTURE

- Data provider with a URI policy that resolves to JSON-LD
- Comprehensive example on BRGM national borehole dataset



# URIS – APPLIED TO THE DATA STRUCTURE

- Data provider with non URI or a URI policy that does not resolve to JSON-LD
- Comprehensive example on PIGMA platform



# WHAT'S NEXT

- Pending IT aspects
  - How to declare a webservices that is not linked to a specific dataset (ex : WPS) ?
  - Link from catalogue to catalogue ?
  - Follow DCAT2 / schema.org work
  - Possibility to use vocabulary from dcat (ex: dcat:DataService, …) : how is it indexed by search engines

- Implementation
  - Agree on JSON-LD patterns
  - Test indexing by search engines
  - Implement the architecture at national scale
  - Maybe push the solution to open source projects (ex : Geonetwork)

# BENEFITS

- National GeoCatalogue and linked catalogues : increases usability and visibility

- Public : enhances overall search experience, allowing to discover, browse, view and download much more environmental data than before

# THANK YOU

Contact

- Abdelfettah.Feliachi@atos.net
- S.Grellet@brgm.fr
- T.Vilmus@brgm.fr