

Plateformes de données nouvelles API OGC et LINKED DATA

Sylvain Grellet

▶ **To cite this version:**

Sylvain Grellet. Plateformes de données nouvelles API OGC et LINKED DATA. AfiGeo : Réunion du GT Open Data – 13 mai 2019, AfiGeo, May 2019, SAINT-MANDÉ, France. hal-03955829

HAL Id: hal-03955829
https://brgm.hal.science/hal-03955829

Submitted on 25 Jan 2023

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.



PLATEFORMES DE DONNEES
NOUVELLES API OGC ET LINKED DATA

S.GRELLET

AFIGEO GT OPENDATA - 13/05/2019

INSTITUT CARNOT brgm

Géosciences pour une Terre durable
brgm

CONTEXTE

20 ans de mise en place de grandes architectures de données

Près de 20 ans de standardisation technique (CSW, WMS, WFS, SOS etc…) pour la diffusion de données mis en place dans un silo (celui de l'OGC)

- Avantages
  - Reprises aux niveaux internationaux, européens (ex: INSPIRE), nationaux, régionaux par de grandes infrastructures de données
  - Meilleures réussites : WMS (sûr), CSW (poids réglementaire)
  - Moins bonnes (WFS, SOS : coût d'entrée mental élevé (technicité, appropriation des modèles de données externes)

- Inconvénients
  - Une approche 'élitiste' hors des pratiques du web (ex : Swagger API, linked data). Gros besoin d'appropriation par les développeurs des systèmes clients
    - des différentes spécifications de services web OGC (API)
    - et des modèles de données
  - Des données cachées derrières les API
    - Les crawlers des moteurs d'indexation ne requêtent pas les API, ils suivent des liens

⇒ Des gisements de données visibles par un faible nombre d'utilisateurs potentiels (ceux qui comprennent/maîtrise les standards et leurs implémentation)

CONTEXTE

Un changement de paradigme en cours

Collaboration W3C / OGC depuis 2014 suite workshop à l'initiative UK Government Linked Data

- Conséquences
  - Publications des recommandations sur la publication des données et données spatiales en ligne
    - Data on the web best practices
    - Spatial data on the web best practices
  - Publication d'un White paper OGC sur les API reconnaissant le besoin de fournir plus d'API REST (travaux déjà initiés depuis 2011) et de migrer vers de architectures orientées ressources auxquelles sont affectées des URI (Uniform Resource Identifier)

  - Une augmentation de la volonté d'utilisation d'URI sur les objets diffusés par les API
  - Une volonté de considérer également les pratiques du web (donc du web des données) dès les spécifications

  - /!\ version courte volontairement peu technique /!\

⇒ Passage d'une approche SOAP/XML silotée à une approche plus RESTful fournissant des données dans des formats tels JSON (et ses dérivés) et liées au maximum

⇒ Une revisite des spécifications des API de diffusion de données

LA NOUVELLE RECETTE ÉMERGENTE

Focus sur l'exposition des données

- Version courtes
  - Des API qui s'autodécrivent: basées sur OpenAPI (anciennement Swagger) et/ou protocole OASIS OData
  - Des hypermédias en réponse

- Une nouvelle « Architecture OGC »
  - API WFS 3: pour les objets métiers (cours d'eau, commune, …)
  - SensorThings API: pour les observations associées
  - URI pour identifier, lier
    - les objets entre eux
    - mais également exposer les vocabulaires contrôlés (dynamique de registres -> INSPIRE register federation) et lier les descriptions d'objets vers les codeList utilisées
  - Métadonnées
    - Exposées en JSON-LD (schema.org / DCAT-AP): pour meilleure indexation
  - Format (sérialisation)
    - Favoriser les formats dérivés de JSON(-LD) et surtout une approche hypermédia permettant de lier les données et d'interagir avec l'API

OGC SENSORTHINGS API

Exposition des données d'observation

La 'version REST du SensorObservationService V2.0'

- V1 depuis 2015:
  - http://docs.opengeospatial.org/is/15-078r6/15-078r6.html
  - V 1.1 en cours

- Philosophie
  - API: basée sur le protocole OASIS OData and les conventions d'URL associées
  - Sémantique des données: respectant le standards ISO/OGC « Observations & Measurements »

- Implémentations techniques
  - Nombreuses implémentations open source
  - Travaux en cours pour le faire reconnaître comme service de téléchargement INSPIRE

- Usage
  - En cours de déploiement dans de nombreuses plate formes en remplacement du SOS
  - Exemple de déploiement en cours dans le pôle de recherche BRGM / AFB 'INSIDE' : https://github.com/INSIDE-information-systems/SensorThingsAPI

OGC SENSORTHINGS API

Exemple

La ‘version REST du SensorObservationService V2.0’

Non sécurisé | sensorthings.brgm-rec.fr/SensorThingsGroundWater/v1.0/Observations(1752377)/Datastream

Appel REST


```
// 20190513140752
// http://sensorthings.brgm-rec.fr/SensorThingsGroundWater/v1.0/Observations(1752377)/Datastream

{
  "name": "GroundWaterLevel Piezo 00497X0018/S1/PZ/2",
  "description": "GroundWaterLevel Piezo 00497X0018/S1/PZ/2",
  "observationType": "http://www.opengis.net/def/observationType/OGC-OM/2.0/OM_Measurement",
  "unitOfMeasurement": {
    "name": "m",
    "symbol": "m",
    "definition": "http://www.qudt.org/qudt/owl/1.0.0/unit/Instances.html#Meter"
  },
  "phenomenonTime": "2016-05-02T22:00:00.000Z/2026-05-02T22:00:00.000Z",
  "resultTime": "2026-05-02T22:00:00.000Z/2026-05-02T22:00:00.000Z",
  "Sensor@iot.navigationLink": "http://sensorthings.brgm-rec.fr/SensorThingsGroundWater/v1.0/Datastreams(631)/Sensor",
  "ObservedProperty@iot.navigationLink": "http://sensorthings.brgm-rec.fr/SensorThingsGroundWater/v1.0/Datastreams(631)/ObservedProperty
  "Thing@iot.navigationLink": "http://sensorthings.brgm-rec.fr/SensorThingsGroundWater/v1.0/Datastreams(631)/Thing",
  "Observations@iot.navigationLink": "http://sensorthings.brgm-rec.fr/SensorThingsGroundWater/v1.0/Datastreams(631)/Observations",
  "@iot.id": 631,
  "@iot.selfLink": "http://sensorthings.brgm-rec.fr/SensorThingsGroundWater/v1.0/Datastreams(631)"
}
```


Approche hypermedia:
Liens activables pour traverser le contenu de l’API

OGC WFS 3

Exposition des objets métiers

Un cours d'eau, un forage, une commune, ...

- Une revisite complète du standard et de sa méthode de spécification
  - V 3.0.0 'draft' : en ligne et ouverte à commentaire depuis Juillet 2018 -> finalisation mi-2019
    - https://github.com/opengeospatial/WFS_FES

- Philosophie
  - API: un cœur simple visant des réponses en JSON, disponible également en OpenAPI. Des méthodes de requêtages plus simples
  - Sémantique des données: assouplie

- Implémentations techniques
  - Nombreuses implémentations open source serveurs et clients, listées sur le GitHub (en cours de mise à jour).

  Clients : dont GDAL, openlayers et leaflet …

OGC WFS 3

- Usage
  - Nombreux tests internationaux
  - Exemple de déploiement en cours dans le pôle de recherche BRGM / AFB ‘INSIDE’ : https://github.com/INSIDE-information-systems/WFS3

- Exemples d’appel
  - Exploration de collection

```
GET /collections
```

  - Recherche par emprise

```
GET /collections/{collectionId}/items?bbox=160.6,-55.95,-170,-25.89
```

  - Accès par identifiant

```
GET /collections/{collectionId}/items/{featureId}
```

OGC WFS 3

- Exemple de réponse


```
{
  "type": "FeatureCollection",
  "features": [
    {
      "type": "Feature",
      "featureType": "Borehole",
      "id": "0001000001",
      "geometry": {
        "type": "Point",
        "coordinates": [
          1.4298,
          51.0684
        ]
      },
      "properties": {
        "identifier": {
          "value": "https://data.geoscience.fr/id/Borehole/BSS000AAAA",
          "@codeSpace": "http://www.ietf.org/rfc/rfc2616"
        },
        "bholeHeadworks": [
          {
        ],
        "type": {
          "@dataType": "Reference",
          "@href": "http://vocabulary.odm2.org/samplingfeaturetype/borehole/",
          "@title": "borehole"
        },
        "startPoint": {
          "@dataType": "Reference",
          "@href": "http://resource.europe-geology.eu/vocabs/StartPoint/naturalLandSurface",
          "@title": "naturalLandSurface"
        },
        "status": {
          "@dataType": "Reference",
          "@href": "http://id.eaufrance.fr/nsa/146#3",
          "@title": "Opérationnel"
        },
        "locationMeasurementMethod": {
          "@dataType": "Reference",
          "@href": "http://id.eaufrance.fr/nsa/917#7",
          "@title": "DGPS"
        },
        "length": {
          "value": 82,
          "@uom": "http://qudt.org/vocab/unit/M"
        },
```


Approche hypermedia: URI vers autre contenu (liens activables plus bas dans la réponse)

L'EXPOSITION DES MÉTADONNÉES

Complément de l'exposition 'à la INSPIRE'

Utilisation du JSON JSON-LD en utilisant les vocabulaires schema.org, DCAT_AP

Mapping des éléments de métadonnées vers ces vocabulaires

- Implémentation BRGM au niveau du Geocatalogue national (geocatalogue.fr)
  - Passage en production récent
  - Suivi de l'indexation par les moteurs de recherche et affinage

- Reprise par les communautés
  - Besoin 'outillage' au plus près du propriétaire de la donnée

  ⇒ Reversement vers la communauté OpenSource (cible Geonetwork V3.8.0)

  - Mais également réflexion nécessaires par les producteurs sur l'identifications des ressources

  ⇒ URI

L'EXPOSITION DES MÉTADONNÉES

Quelques exemples

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",
"bbox":[     "-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"]
             }
 }, ……
```


Géosciences pour une Terre durable
brgm

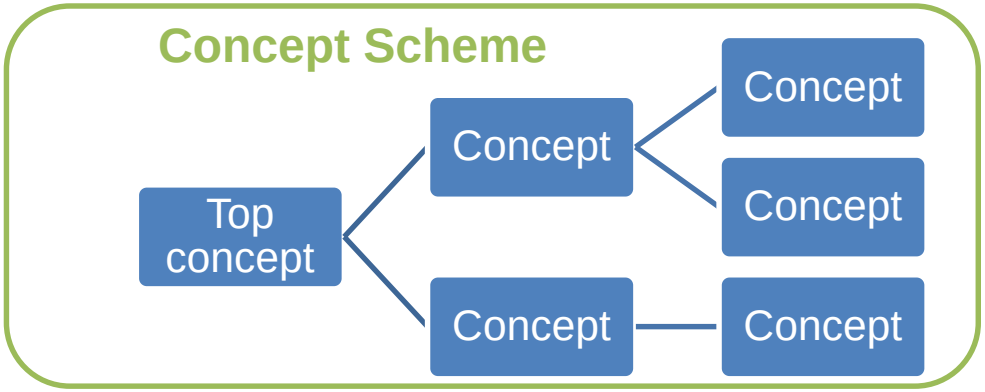
L'EXPOSITION DE LISTES DE CODES

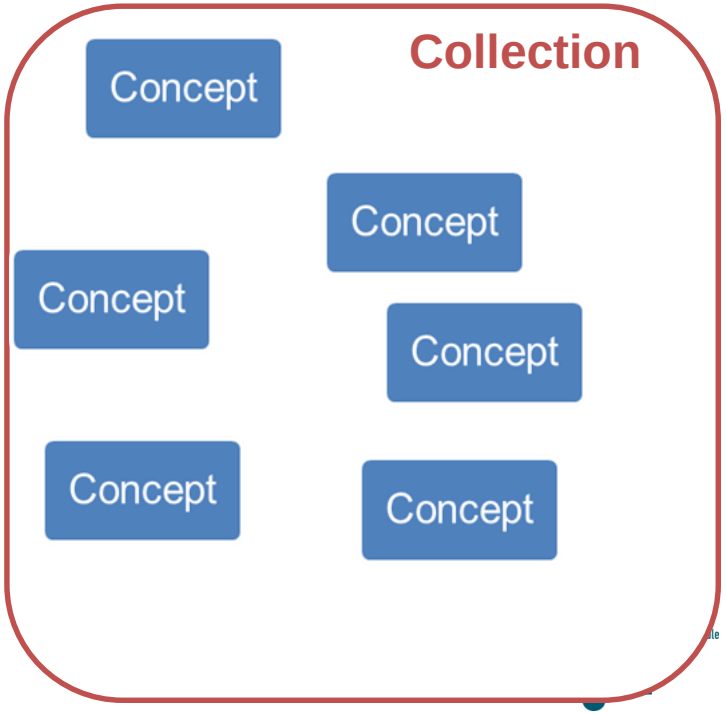
Notion de registre de codeList

Ensemble de termes homogènes, codifiés et répertoriés (nomenclature, thesaurus, …) chacun possédant un identifiant unique externe stable sous la forme d'une URI. Ces termes sont cohérents entre eux selon un domaine donné. Exemple : (le registre de la lithologie validée du BRGM : https://data.geoscience.fr/ncl/litho).

Une Dynamique européenne (INSPIRE register federation) et une implémentation Française autour du Geocatalogue national

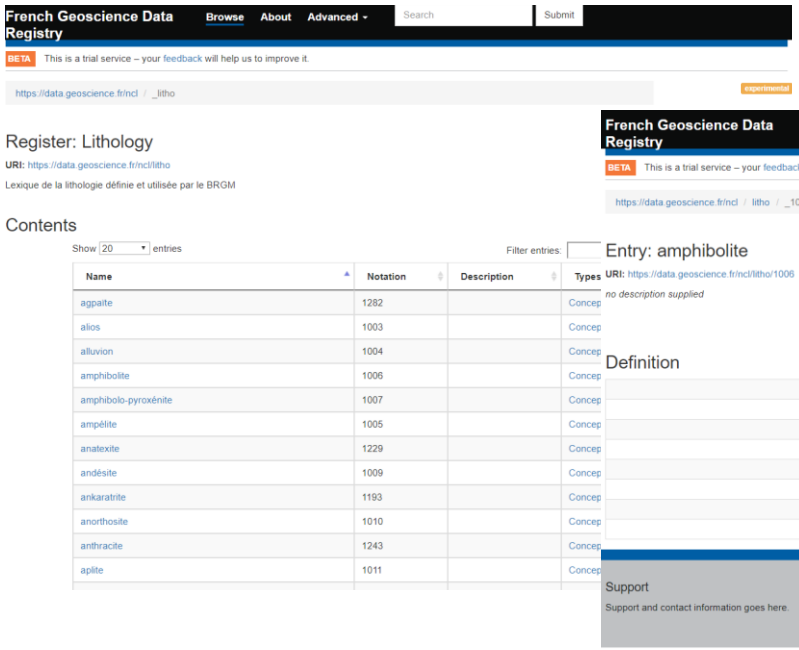
- Formalisation du contenu des codeliste , nomenclature, taxonomie:
  - Utilisation de l'ontologie SKOS
  - Entités skos:Concept
  - Conteneur hiérarchique (ou pas) skos:ConceptScheme
  - Conteneur non hiérarchique skos:Collection

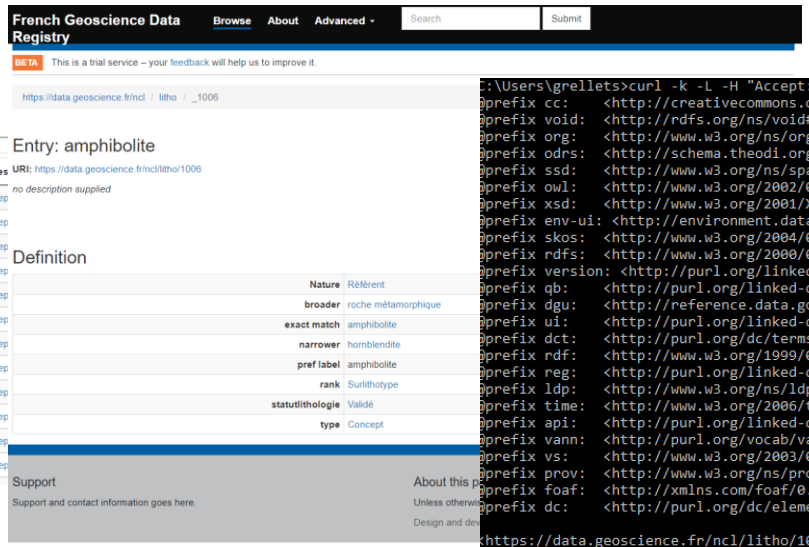


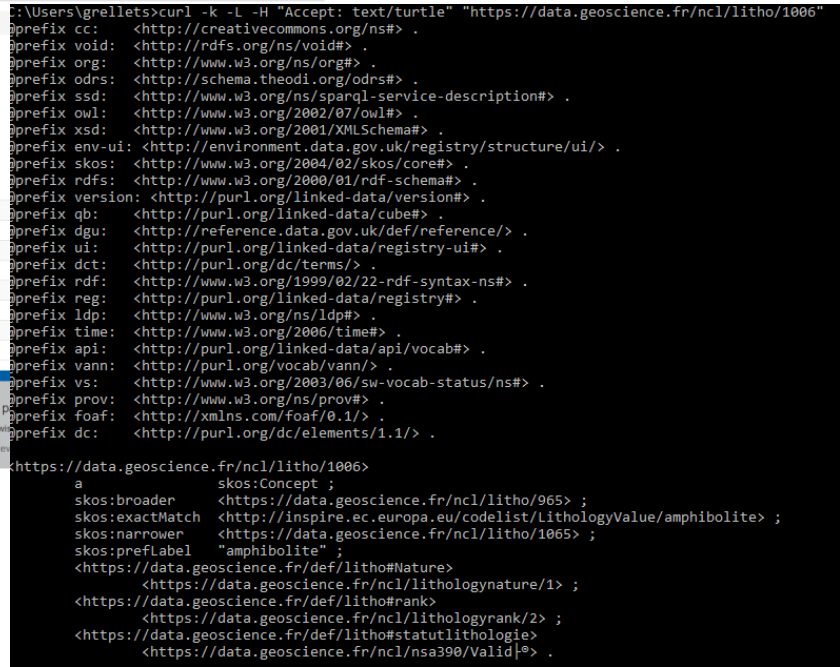


L'EXPOSITION DE LISTES DE CODES

- Une approche linked data basée sur les bonnes pratiques du Web:
  - Des URI pour identifier les registres et les entités des registres
  - Des vocabulaires standards (ou recommandés) pour structurer le contenu (skos, rdfs, foaf, dcterms,…
  - Retourner de différentes représentations (structurés, non structurées)
  - ⇒ négociation de contenu (Accès pour les thématiciens ET les machines)
  - Inclure dans la description des entités des liens vers des registres externes







STANDARDS OGC ET DONNÉES LIÉES

Appliquer ces principes dès les spécifications

Une suite de tests OGC autour de la notion d'objets environnementaux liés (Environmental Linked Features)

- ELFIE-1 : Environmental Linked Features Interoperability Experiment (2018)
  - Définition d'approches utilisant JSON-LD (Linked Data) pour décrire et lier des descriptions d'objets environnementaux. Ex : une station suivant le niveau d'une nappe phréatique
  - Application des bonnes pratiques W3C tout en respectant les modèles de données métiers
  - JSON-LD -> Indexation par les crawlers des moteurs de recherche (test de schema.org)
  - Démo/engineering report: https://opengeospatial.github.io/ELFIE/ (deux exemples en France : 'Surface-Ground Water Networks Interaction Demo' et 'Ground Water Monitoring Demo')

- SELFIE : Second ELFIE / ELFIE-2 (2019)
  - Intégration des conclusions d'ELFIE dans les standards OGC (surtout WFS3 et SensorThings API)
  - https://www.opengeospatial.org/projects/initiatives/selfie

LES SUITES

Des changements de pratiques

- Gérer le biseau entre les deux paradigmes
  - Ces nouvelles API ne vont pas éteindre les précédentes en un jour
  - Migration des infrastructures à gérer. Certaines solutions permettent les deux (ex : Geoserver)
  - Dénominateur commun aux deux: mettre des URI sur les ressources diffusées (objets, codeList)

- Recaler les démarches d'APIfication déjà lancées (mais non standards)
  - De nombreuses plateformes ont déjà des API ad'hoc donc non interopérables
  - ⇒ Prévoir de les 'mettre à jour' pour bénéficier de effets de l'interopérabilité
  - ⇒ Sans perdre les communautés créées

- Côté spécifications
  - WFS3: officialisation du cœur, intégration du JSON-LD ensuite
  - SensorThingsAPI : optimisation dans la V1.1, intégration du JSON-LD ensuite

LES SUITES

Un déploiement plus large

- Côté technique
  - Amélioration des solutions serveurs
  - Généralisation des clients

- Généralisation aux autres méthodes de diffusion de données
  - Approche RESTful et hypermedia (dont JSON-LD)
  - CSW, WMS, WPS, WCS etc…

⇒ Vers un recalage des pratiques OpenData, linked data, interopérabilité ?

Contact

- S.Grellet@brgm.fr

