



RÉPUBLIQUE  
FRANÇAISE

*Liberté  
Égalité  
Fraternité*



Géosciences pour une Terre durable

**brgm**

# EVOLUTIONS/DYNAMIQUES EN COURS DANS LE WEBSEMANTIQUE/LINKEDDATA

22/01/2021





# Des bonnes pratiques

- Data on the Web Best Practices

- W3C Recommendation

31 January 2017

- <https://www.w3.org/TR/dwbp/>

[Best Practice 1](#): Provide metadata

[Best Practice 2](#): Provide descriptive metadata

[Best Practice 3](#): Provide structural metadata

[Best Practice 4](#): Provide data license information

[Best Practice 5](#): Provide data provenance information

[Best Practice 6](#): Provide data quality information

[Best Practice 7](#): Provide a version indicator

[Best Practice 8](#): Provide version history

[Best Practice 9](#): Use persistent URIs as identifiers of datasets

[Best Practice 10](#): Use persistent URIs as identifiers within datasets

[Best Practice 11](#): Assign URIs to dataset versions and series

[Best Practice 12](#): Use machine-readable standardized data formats

[Best Practice 13](#): Use locale-neutral data representations

[Best Practice 14](#): Provide data in multiple formats

[Best Practice 15](#): Reuse vocabularies, preferably standardized ones

[Best Practice 16](#): Choose the right formalization level

[Best Practice 17](#): Provide bulk download

[Best Practice 18](#): Provide Subsets for Large Datasets

[Best Practice 19](#): Use content negotiation for serving data available in multiple formats

[Best Practice 20](#): Provide real-time access

[Best Practice 21](#): Provide data up to date

[Best Practice 22](#): Provide an explanation for data that is not available

[Best Practice 23](#): Make data available through an API

[Best Practice 24](#): Use Web Standards as the foundation of APIs

[Best Practice 25](#): Provide complete documentation for your API

[Best Practice 26](#): Avoid Breaking Changes to Your API

[Best Practice 27](#): Preserve identifiers

[Best Practice 28](#): Assess dataset coverage

[Best Practice 29](#): Gather feedback from data consumers

[Best Practice 30](#): Make feedback available

[Best Practice 31](#): Enrich data by generating new data

[Best Practice 32](#): Provide Complementary Presentations

[Best Practice 33](#): Provide Feedback to the Original Publisher

[Best Practice 34](#): Follow Licensing Terms

[Best Practice 35](#): Cite the Original Publication



# Des bonnes pratiques

- Spatial Data on the Web Best Practices

- Travail W3C/OGC
- W3C Working Group Note

28 September 2017

- <https://www.w3.org/TR/sdw-bp/>

Best Practice 1: Use globally unique persistent HTTP URIs for Spatial Things

Best Practice 2: Make your spatial data indexable by search engines

Best Practice 3: Link resources together to create the Web of data

Best Practice 4: Use spatial data encodings that match your target audience

Best Practice 5: Provide geometries on the Web in a usable way

Best Practice 6: Provide geometries at the right level of accuracy, precision, and size

Best Practice 7: Choose coordinate reference systems to suit your user's applications

Best Practice 8: State how coordinate values are encoded

Best Practice 9: Describe relative positioning

Best Practice 10: Use appropriate relation types to link Spatial Things

Best Practice 11: Provide information on the changing nature of spatial things

Best Practice 12: Expose spatial data through 'convenience APIs'

Best Practice 13: Include spatial metadata in dataset metadata

Best Practice 14: Describe the positional accuracy of spatial data



# Dynamique du web des données de R&D -> production

Une « appropriation » / reprise de ces logiques par la sphère commerciale

- 2011 : schema.org
  - Bing, Google, Yahoo, Yandex
  - ⇒ Amélioration indexation pages web par les crawler (sémantisation du contenu des pages)
  - Depuis, support puis recommandation usage JSON-LD appelant entre autre des vocabulaires issus de Schema.org
  - ⇒ Nombreux test tools / validators fournis
- => De nombreuses fournisseurs de données / agences environnementales s'y mettent
- => côté sphère INSPIRE
  - 2018 INSPIRE conference Google Dataset Search workshop
  - 2019 “Workshop on making spatial data discoverable through mainstream search engines”



# Un changement de paradigme en cours

Une URI vers plusieurs représentations de la même donnée

- Négociation de contenu par profil
  - W3C Working Draft 26 November 2019
  - <https://www.w3.org/TR/dx-prof-conneg/>
  - <https://github.com/w3c/dx-conneg>

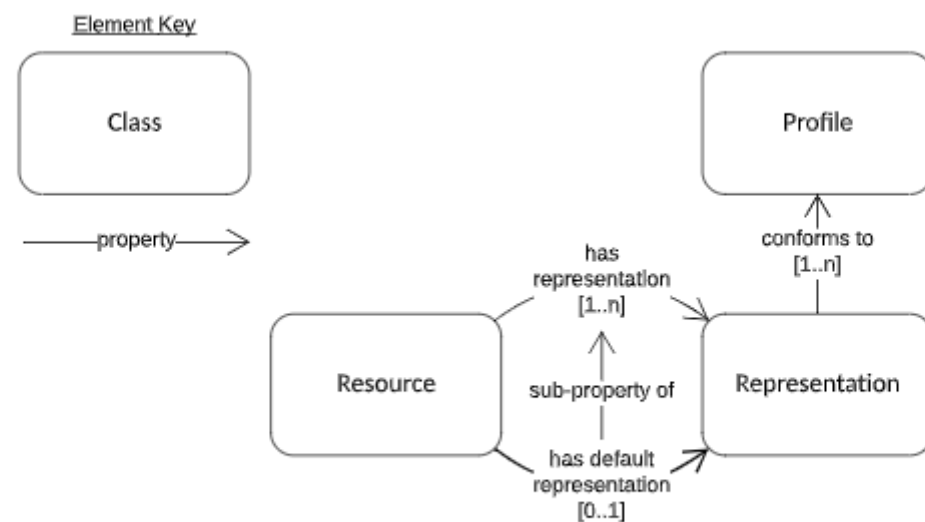


Figure 1 A diagram of the Alternate Profiles data model implemented in OWL [OWL2-OVERVIEW].

# Un changement de paradigme en cours

## Une URI vers plusieurs représentations de la même donnée

- Négociation de contenu par profil

**EXAMPLE 8:** Server returns a Link header to indicate available profiles

```
HEAD /resource/a HTTP/1.1
Accept: text/turtle
[more request headers]

---

HTTP/1.1 200 OK
Content-Type: text/turtle
Content-Profile: <urn:example:profile:x>
Link:
  <http://example.org/resource/a>;
    rel="canonical";
    type="text/turtle";
    profile="urn:example:profile:x",
  <http://example.org/resource/a>;
    rel="alternate";
    type="text/turtle";
    profile="urn:example:profile:y",
  <http://example.org/resource/a>;
    rel="alternate";
    type="application/xml";
    profile="urn:example:profile:x",
  <http://example.org/resource/a>;
    rel="alternate";
    type="application/xml";
    profile="urn:example:profile:y",
  <http://example.org/resource/a>;
    rel="alternate";
    type="text/html"
[more response headers]
```

Ajout de champs dans  
l'en-tête http



In [Example 8](#), for each of the different Media Type / Profile combinations, the URI of the resource remains unchanged and it is the **Accept** and **Accept-Profile** headers for Media Type and Profile respectively that alter the returned response.

# Un changement de paradigme en cours

Une URI vers plusieurs représentations de la même donnée

- Test d'interopérabilité OGC SELFIE (Second Environmental Linked Features Interoperability Experiment (SELFIE))
  - <https://github.com/opengeospatial/SELFIE>
  - <https://docs.ogc.org/per/20-067.html>



# De plus en plus de sémantiques sur les données

## Les communautés métiers s'approprient le WebSémantique

- Data Catalog Vocabulary (DCAT) - Version 2
  - W3C Recommendation 04 February 2020  
<https://www.w3.org/TR/vocab-dcat-2/>
  - Description données, service, catalogue -> cf logique 'à la ISO 19115/19139, à la Geonetwork'
- Observations & Measurements (O&M)
  - W3C SSN/SOSA : travail conjoint W3C/OGC sur O&M V2  
<https://www.w3.org/TR/vocab-ssn/>
  - Sera mis à jour suite évolution O&M en V3



# De plus en plus de sémantiques sur les données

## Les communautés métiers s'approprient le WebSémantique

- Ex d'ontologies en cours côté OGC

Basées sur le travail historique de maïeutique des groupes mais pas une conversion UML -> OWL

(cf rapport test interopérabilité OGC ELFIE Annexe A : [https://docs.ogc.org/per/18-097.html#ontology\\_from\\_uml](https://docs.ogc.org/per/18-097.html#ontology_from_uml) )

- Hydrographie -> HY\_Features

- OGC / Organisation Météorologique Mondiale (WMO Chy)
- Équivalence vers INSPIRE Hy et, ex : Carthage/Topage en France

- Hydrogéologie -> GroundWaterML 2.0

- OGC / Organisation Météorologique Mondiale (WMO Chy)

- Géologie -> GeoscienceML

- OGC / IUGS CGI



# De plus en plus de sémantiques sur les données

## Les communautés métiers s'approprient le WebSémantique

- Monde de la recherche -> Research Data Alliance (RDA)
  - Vocabulary Services Interest Group : <https://www.rd-alliance.org/groups/vocabulary-services-interest-group.html>
  - Data Discovery Paradigms Interest Group: <https://www.rd-alliance.org/groups/data-discovery-paradigms-ig>
- + lien avec ESIP Federation (NASA, NOAA, USGS)  
=> <https://github.com/ESIPFed/science-on-schema.org>
- BioSchema
  - -> presentation Franck





# Spécifications d'API

Les spécifications considèrent maintenant le WebSémantique et les logiques de données liées

- Un tournant : début collaboration W3C/OGC dans les années 2014
- OGC API Records (revision CSW) : travail en cours
  - “Basic information model: strongly considers using DCAT”
  - “Linked data : Wherever possible, provisions will be made to include "links" sections to accomodate the use of such hypermedia or linking elements.”
- OGC API Feature (revision WFS) - Core
  - Links section pour les aspects hypermedia (et aussi pour lier elements/collection)
  - + logique de LinkedData selon déploiement (cf présentation pôle INSIDE)
- OGC SensorThings API (revision SOS) – Part 1
  - 100% hypermedia driven (HATEOAS), logique de LinkedData selon déploiement (cf présentation pôle INSIDE)



# WebSémantique et pile d'outils open source

Des implémentation OpenSource disponibles non réservées aux 'sachants' du domaine

- Registres de codeList/vocabulaire
  - Epimorphics LDRegistry :
    - Utilisation par de nombreuses "grandes" organisations
    - Lien vers fédérations registre EU
- Cataloguage : Geonetwork
  - Intégration JSON-LD dans réponse
  - Migration DCAT -> DCAT V2 à faire
- Exposition d'objets métiers (WFS 2, OGC API Features) : Geoserver
  - Négociation de contenu (pas par profil encore)
  - Feature templating plugin -> JSON-LD
- Exposition d'observation (SensorThings API) : Fraunhofer - FROST
  - Selon SOSA/SSN prototype dans le cadre de (S)ELFIE.
  - Note : par design SensorThings API est déjà conforme Observations & Measurements (O&M V2)



# Conclusion

De nombreux fondements ont maturés et sont utilisés en production

- La pile/séquence catalogue > services > données a vécu
  - toutes les ressources sont au même niveau,
  - prêtes à être indexées par des moteurs de recherche
  - certains implémentant des vocabulaires plus riches (ex : Google et DCAT)
- Passage en production dans tous les secteurs
  - Autorité publique : De nombreux fournisseurs de données / agences environnementales s'y mettent à l'échelle mondiale
  - Recherche (via RDA et CODATA) mouvement de fond pour exposer les données de la recherche avec de la sémantique ET en les liant
  - Secteur privé
    - De nombreuses sociétés de services se montent sur le sujet (France, EU, International)
    - Point "Godwin" : certains acteurs s'interrogent de savoir pourquoi leurs données ne sont pas dans le Google DataSet search

# Conclusion

De nombreux fondements ont maturés et sont utilisés en production

- 2 grands pans pour le WebSémantique : LinkedData / Reasoning
  - La majorité des implémentations en cours de mise en place dans les organismes publiques vise les aspects LinkedData
  - Le reasoning (les inférences) arrive ensuite
- Par la standardisation internationale également (ex : ISO TC 211 -> série ISO 191xx)
  - ISO/OsGeo : GeoLexica
  - 2020/2021 : Groupe informel en cours : 'considering the impact of non-relational technologies on TC211 standards' (lire : utilisation de plus de méthodes du WebSémantique)



# Merci

[a.feliachi@brgm.fr](mailto:a.feliachi@brgm.fr)

[s.grellet@brgm.fr](mailto:s.grellet@brgm.fr)

