



Les linked data géographiques ... et INSPIRE l'expérience Datalift

Journée AFIGEO GT OGC OpenData

Saint Mandé 6 mars 2013

Nathalie Abadie – IGN France

Faycal Hamdi - CNAM

Plan

Les linked data

Le projet Datalift

Les linked data géographiques

... et INSPIRE

Linked data

- <http://www.numerama.com/magazine/23701-open-data-europeana-ouvre-les-donnees-de-20-millions-d-oeuvres.html>
- Les choix techniques du World Wide Web Consortium (W3C) pour supporter la découverte, le croisement, la réutilisation des données :
 - Utiliser des URI pour identifier des *choses* (...dont on parle)
 - Utiliser des URI HTTP déréréférencables par des personnes et par des agents automatiques
 - Lorsqu'une URI est entrée dans un navigateur fournir une information pertinente à propos de la *chose* identifiée en réponse
 - Publier des liens entre URIs pour faciliter le parcours du Web de description en description.

Linked Open Data (LOD)

5 étoiles

- ★ Publier votre contenu sur le Web sous une licence ouverte
- ★★ Le publier sous forme de données structurées (ex : Excel)
- ★★★ Utiliser un format non propriétaire
- ★★★★ Utiliser des URIs pour identifier les *choses*, dérèfèrenceables
- ★★★★★ Créer des liens entre vos données et d'autres données

Le projet Datalift

- Projet ANR CONTINT (2010-2014) :
 - Une plate-forme pour aider des fournisseurs de contenu à publier leurs données sous forme de linked data.



INRIA

Atos
Origin



IGN

INSTITUT NATIONAL
DE L'INFORMATION
GÉOGRAPHIQUE
ET FORESTIÈRE

EURECOM



MONDECA

Le projet Datalift

- Idées directrices :
 - Des choix de publication (façon de faire les URI, schémas RDF utilisés) auront des conséquences sur l'intégration des contenus publiés:
→ **aider les fournisseurs de contenus à faire les bons choix.**
 - L'intégration (construction des liens) est encore manuelle
→ **proposer des outils pour une construction la plus automatique possible des liens.**
 - Un point clé pour le succès/accélération du Web des données est d'avoir une certaine masse de données, et des données de confiance
→ **IGN, INSEE, etc.**

Le projet Datalift

- Les URIs :
 - Données GEOFLA: L'URI
<http://data.ign.fr/id/geofla/departement/51>

Renvoie la page suivante:



Description du nœud <<http://data.ign.fr/id/geofla/departement/51>>

Rechercher dans

Sujet ↕	Predicat	Objet
http://data.ign.fr/id/geofla/departement/51	http://data.ign.fr/id/geofla/code-dept	"51"
http://data.ign.fr/id/geofla/departement/51	http://data.ign.fr/id/geofla/nom-dept	"MARNE"
http://data.ign.fr/id/geofla/departement/51	http://data.ign.fr/id/geofla/code-chf	"108"
http://data.ign.fr/id/geofla/departement/51	http://data.ign.fr/id/geofla/nom-chf	"CHALONS-EN-CHAMPAGNE"
http://data.ign.fr/id/geofla/departement/51	http://data.ign.fr/id/geofla/x-chf-lieu	"7999"^^< http://www.w3.org/2001/XMLSchema#integer >
http://data.ign.fr/id/geofla/departement/51	http://data.ign.fr/id/geofla/y-chf-lieu	"68738"^^< http://www.w3.org/2001/XMLSchema#integer >
http://data.ign.fr/id/geofla/departement/51	http://data.ign.fr/id/geofla/code-reg	"21"
http://data.ign.fr/id/geofla/departement/51	http://data.ign.fr/id/geofla/nom-region	"CHAMPAGNE-ARDENNE"
http://data.ign.fr/id/geofla/departement/51	http://data.ign.fr/id/geofla/st-astext	"MULTIPOLYGON(((4.67018320729181 48.5318873023232,4.6655623683

Le projet Datalift

- Les URIs :

- Distinction entre objet du monde réel et objet de BD
Préconisé dans la littérature.

Exemple Ordnance Survey:

Cette URI identifie la ville de Southampton

<http://data.ordnancesurvey.co.uk/id/7000000000037256>

Quand on la déréférence, on accède:

- À la donnée RDF qui décrit la ville de Southampton

- (<http://data.ordnancesurvey.co.uk/doc/7000000000037256.rdf>)

- À une page HTML qui décrit la ville de Southampton

- (<http://data.ordnancesurvey.co.uk/doc/7000000000037256.html>)

Donnée RDF qui décrit la ville de Southampton:

```
<rdfs:label>The City of Southampton</rdfs:label>
<spatialrelations:northing rdf:datatype="http://www.w3.org/2001/XMLSchema#decimal">112882.8</spatialrelations:northing>
<admingeo:hasAreaCode>UTA</admingeo:hasAreaCode>
<spatialrelations:touche rdf:resource="http://data.ordnancesurvey.co.uk/id/7000000000017596"/>
<spatialrelations:touche rdf:resource="http://data.ordnancesurvey.co.uk/id/7000000000043511"/>
<spatialrelations:touche rdf:resource="http://data.ordnancesurvey.co.uk/id/7000000000017733"/>
<spatialrelations:touche rdf:resource="http://data.ordnancesurvey.co.uk/id/7000000000017765"/>
<geo:long rdf:datatype="http://www.w3.org/2001/XMLSchema#decimal">-1.399128</geo:long>
<skos:notation rdf:datatype="http://statistics.data.gov.uk/def/administrative-geography/StandardCode">00MS</skos:notation>
<rdf:type rdf:resource="http://data.ordnancesurvey.co.uk/ontology/admingeo/UnitaryAuthority"/>
<rdf:type rdf:resource="http://data.ordnancesurvey.co.uk/ontology/admingeo/Borough"/>
<spatialrelations:contains rdf:resource="http://data.ordnancesurvey.co.uk/id/7000000000017705"/>
<spatialrelations:contains rdf:resource="http://data.ordnancesurvey.co.uk/id/7000000000017537"/>
<spatialrelations:contains rdf:resource="http://data.ordnancesurvey.co.uk/id/7000000000017539"/>
<spatialrelations:contains rdf:resource="http://data.ordnancesurvey.co.uk/id/7000000000017565"/>
<spatialrelations:contains rdf:resource="http://data.ordnancesurvey.co.uk/id/7000000000017711"/>
<spatialrelations:contains rdf:resource="http://data.ordnancesurvey.co.uk/id/7000000000017707"/>
<spatialrelations:contains rdf:resource="http://data.ordnancesurvey.co.uk/id/7000000000017698"/>
<spatialrelations:contains rdf:resource="http://data.ordnancesurvey.co.uk/id/7000000000017720"/>
<spatialrelations:contains rdf:resource="http://data.ordnancesurvey.co.uk/id/7000000000017710"/>
<spatialrelations:contains rdf:resource="http://data.ordnancesurvey.co.uk/id/7000000000017703"/>
<spatialrelations:contains rdf:resource="http://data.ordnancesurvey.co.uk/id/7000000000017566"/>
<spatialrelations:contains rdf:resource="http://data.ordnancesurvey.co.uk/id/7000000000017591"/>
<spatialrelations:contains rdf:resource="http://data.ordnancesurvey.co.uk/id/7000000000017535"/>
<spatialrelations:contains rdf:resource="http://data.ordnancesurvey.co.uk/id/7000000000041760"/>
<spatialrelations:contains rdf:resource="http://data.ordnancesurvey.co.uk/id/7000000000017699"/>
<spatialrelations:contains rdf:resource="http://data.ordnancesurvey.co.uk/id/7000000000017570"/>
<georss:point>50.913855 -1.399128</georss:point>
<skos:prefLabel>The City of Southampton</skos:prefLabel>
```

Page HTML qui décrit la ville de Southampton:



The City of Southampton

A description of the resource identified by <http://data.ordnancesurvey.co.uk/id/7000000000037256>

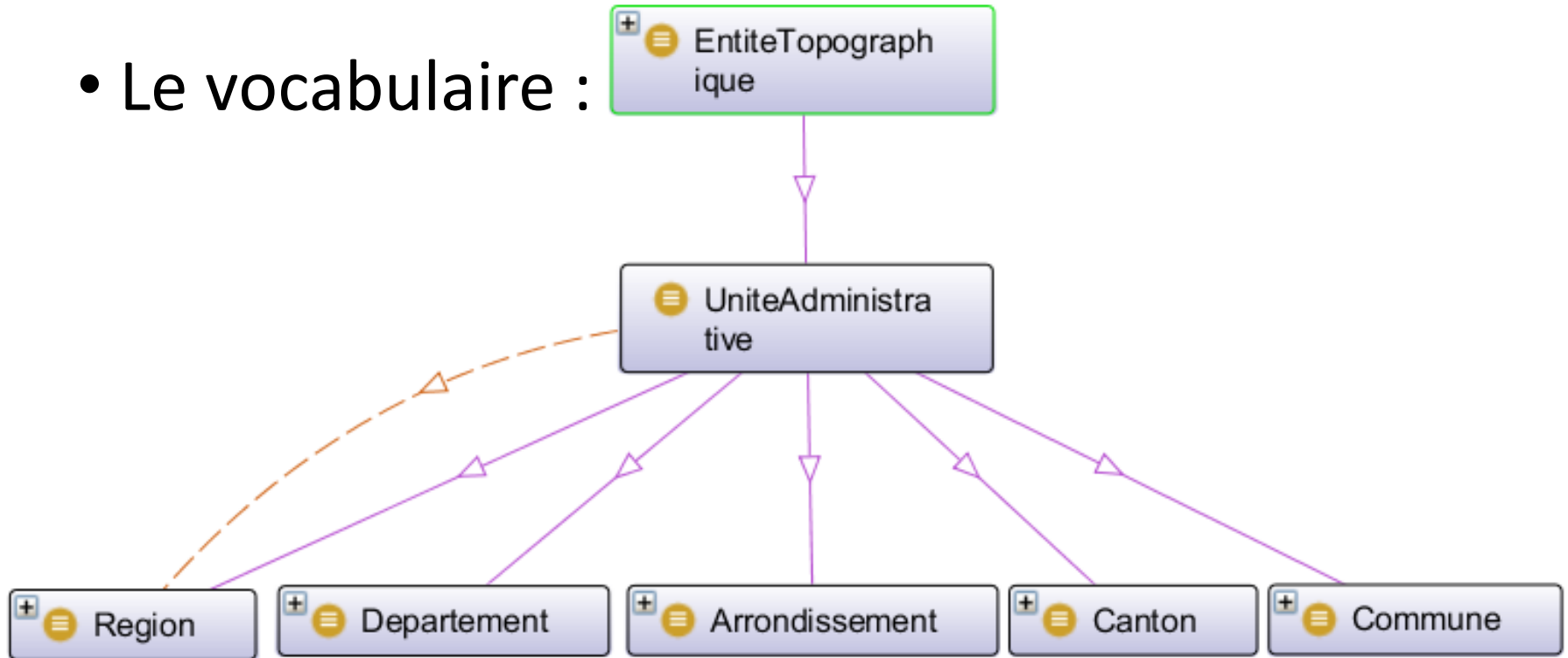
Preferred label	The City of Southampton
Label	The City of Southampton
Alternative label	City of Southampton
Latitude	50.913855
Longitude	-1.399128
Type	Unitary Authority Borough
Gss code	E06000045
Area code	UTA
Census code	00MS
Unit i d	37256
In european region	South East
Ward	Harefield Bevois Redbridge Bitterne Portswood

Le projet Datalift

- Les URIs :
 - Distinction entre objet du monde réel et objet de BD est prévue et en cours de mise en place:
 - URI identifiant le département 51 (monde réel):
<http://data.ign.fr/id/geofla/departement/51>
 - URI identifiant la données RDF correspondante:
<http://data.ign.fr/data/geofla/departement/51>
 - La gestion des versions des jeux de données est faite au travers des URIs:
 - <http://data.ign.fr/id/geofla/version>
 - <http://data.ign.fr/id/geofla/>: c'est la dernière version du jeu de données

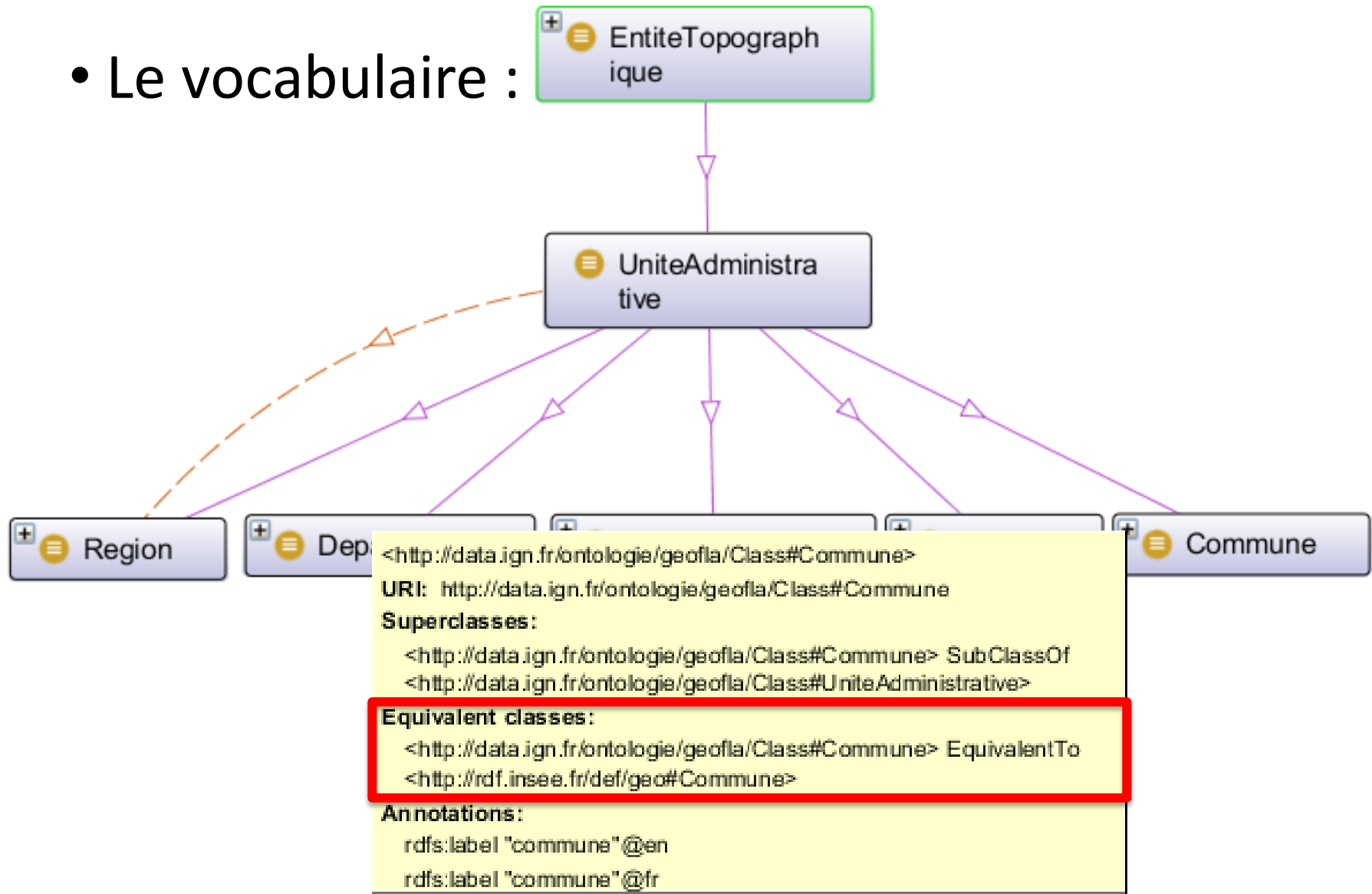
Le projet Datalift

- Le vocabulaire :



Le projet Datalift

- Le vocabulaire :



Le projet Datalift

- Le vocabulaire :
 - Le vocabulaire BDTOPO.owl est en ligne:
<http://data.ign.fr/ontology/topo.owl>
 - Un vocabulaire est en cours d'élaboration pour représenter des géométries structurées:
 - repose sur le vocabulaire GeoSPARQL
 - y ajoute des géométries structurées (pas en WKT)

Le projet Datalift

- L'interconnexion : Comparaison de vocabulaires, interconnexion de données

Les approches mises en œuvre dans le domaine Linked Data s'appuient sur des mesures de distance simples

Des travaux au COGIT ont étudié l'appariement multicritère: géométrie + valeurs d'attributs + sémantique (comparaison de vocabulaires)

→ Intégrer cette approche aux outils d'interconnexion

Le projet Datalift

Perspective: Un vocabulaire intégrant des connaissances pour l'interconnexion...

Objectif: Ajouter du sens à la propriété reliant un objet à sa géométrie

- Quelles sont les règles de saisie de la géométrie?
- Quelle est la nature de la relation entre un objet thématique et l'objet topographique via lequel il est localisé?

INSPIRE

(<http://inspire.jrc.ec.europa.eu/index.cfm/pageid/48>)

« INSPIRE is based on a number of common principles:

- Data should be **collected only once** and kept where it can be maintained most effectively.
- It should be possible to **combine** seamless spatial information from **different sources** across Europe and share it with **many users and applications**.
- It should be possible for information collected at one **level/scale** to be shared with all levels/scales; detailed for thorough investigations, general for strategic purposes.
- Geographic information **needed for good governance at all levels should be readily and transparently available**.
- Easy to **find what** geographic information is available, **how** it can be used to meet a particular need, and under which conditions it can be acquired and used. »

Linked data

- “ We are surrounded by data – data about the performance of our local schools, the fuel efficiency of our cars, a multitude of products from different vendors, or the way our taxes are spent. By helping us make better decisions, this data is playing an increasingly central role in our lives and driving the emergence of a data economy. [..]
- This raises three key questions:
 - How best to provide access to data so it can be most easily reused?
 - How to enable the discovery of relevant data within the multitude of available data sets?
 - How to enable applications to integrate data from large numbers of formerly unknown data sources?
 - Just as the World Wide Web has revolutionized the way we connect and consume documents, so can it revolutionize the way we discover, access, integrate and use data.”

Heath and Bizer, 2010, “Linked Data: Evolving the Web Into a Global Data Space“

Linked data

“ We are surrounded by data – data about the performance of our local schools, the fuel efficiency of our cars, a multitude of products and services, or the way our taxes are spent. The data is playing a role in a number of ways, and a new way of connecting the data is playing a role in a number of ways.”

- This

– I

– H

m

– H

nu

- Just as

consumers are revolutionizing the way we connect and integrate data, we are revolutionizing the way we discover, access, integrate and use data.”

Ressemble beaucoup aux Linked Data en termes d'objectifs mais des points de vue différents :

INSPIRE : point de vue des fournisseurs de données géographiques (notion d'échelle, de termes d'utilisation)

Linked Data : point de vue des consommateurs de données du Web (notion de réutilisabilité, données brutes et licence ouverte)

Heath and Bizer, 2010, “Linked Data: Evolving the Web Into a Global Data Space”

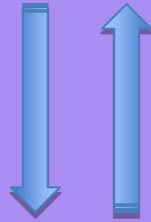
... comparatif

Linked data

INSPIRE

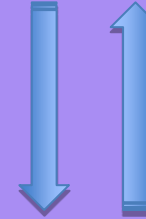
APIs

Déréférencement
d'URI
Requêtes SparQL



- Présentation
- Contenu (RDF)

Requêtes WMS,
WFS, ..



- Visualisation
Cartographique
- Contenu (GML, ...)

INTEGRATION

Mécanismes de classification (OWL)

Vocabulaires partagés

Interconnexions explicites :

<URI, sameAs, URI>

URIs sur les objets du monde réel

Catalogues

Bibliothèques de détection et correction
d'incohérences entre les géométries

Requêtes spatiales, URI

URI sur les *spatial objects* (Feature)

CONTENU

Contenus si possible
structurés et utilisant des
vocabulaires connus

Données dans un schéma
fédéré

... comparatif (suite)

- Linked Open Data :
 - Publier des données brutes sous licence ouverte (pas comme INSPIRE, Google maps, ni 4square, ... plutôt comme OSM)
 - data.gouv initiatives

- INSPIRE :
 - La visualisation est gratuite,
 - « it should be easy to find under which conditions data can be acquired or used »,
 - Recherches sur comment évaluer la performance des SDI et la valeur d'une brique d'information dedans,

Points d'accroche

- Les URIs : fournir des URI pour les points d'accroche géographiques, faisant référence aux objets du monde réel
- L'intégration : mixer le spatial et le sémantique
- Les vocabulaires : promouvoir le schéma fédéré INSPIRE comme vocabulaire du web des données
- Les modèles économiques : apprenons du Web des données les nouveaux modes de création de valeur: subscription model, advertising model, authority model, etc. (voir <http://www.chiefmartec.com/2010/01/7-business-models-for-linked-data.html>)...

Les linked data géographiques ... et INSPIRE l'expérience Datalift



Merci pour votre attention.

Plus d'informations sur

<http://www.datalift.org/>