

Construction d'ontologie à partir de spécifications de BD

Nathalie Aussenac-Gilles, Mouna Kamel, Marion Laignelet

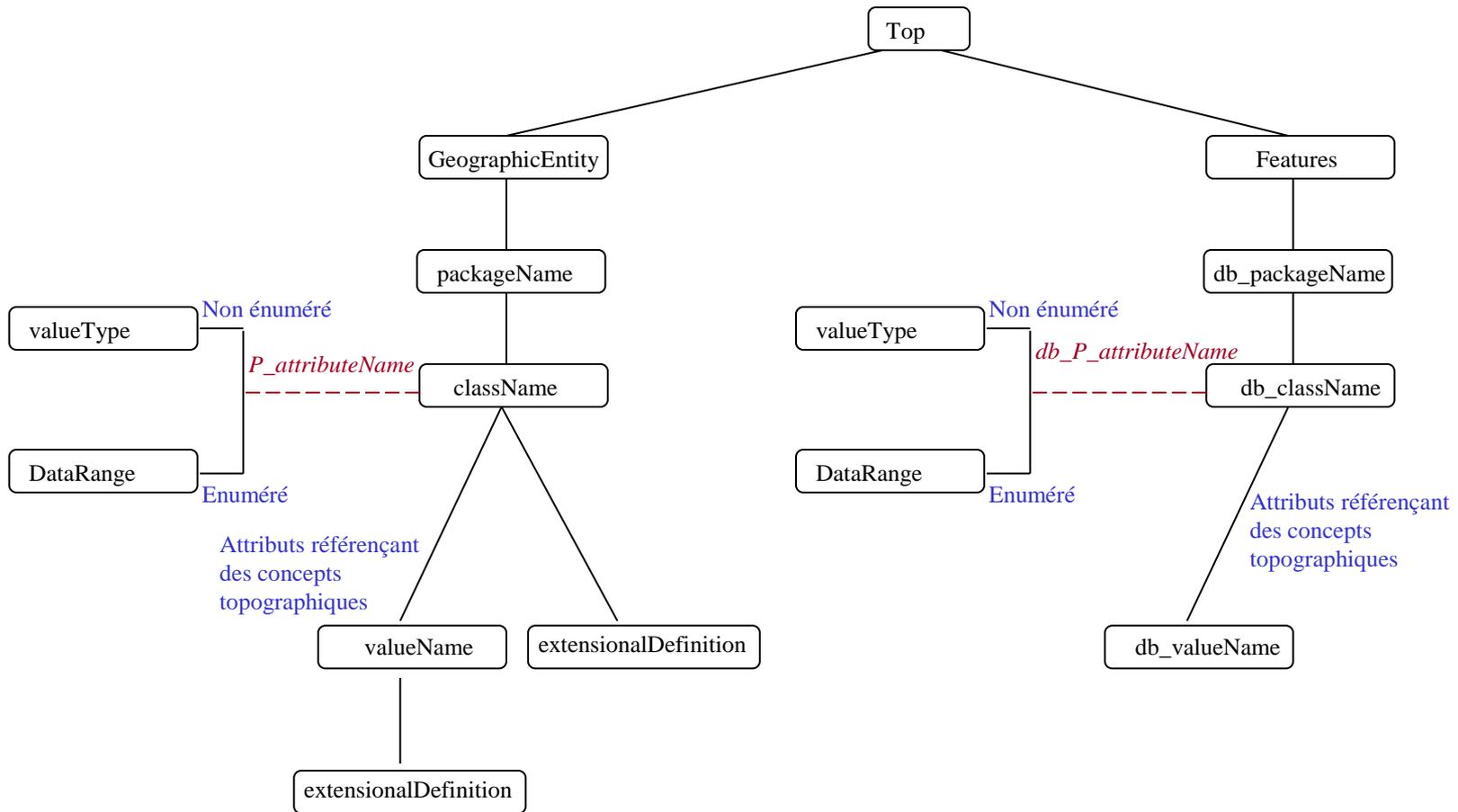
IRIT – IC3

Réunion GEONTO – 21 juin 2010

Nouvelle approche

- Séparation entre les concepts
 - du monde réel
 - du monde de la base de données
- Motivations :
 - liens avec la base de données
 - vocabulaires et domaines qui ne se recouvrent pas complètement

Nouvelle approche



Nouvelle approche

- Nouveaux besoins en terme de formalisme (OWL DL) car utilisation de DAtaRange
- plus possible de développer sous GATE (OWL Light)
- Programme Java qui exploite la structure XML du document (API javax.XML.parsers)
 - En entrée,
 - Le fichier de spécifications
 - La liste des attributs référençant des concepts topographiques
(*statut, classement, nature*, etc. pour BDTopo – *vocation de la liaison, nature, état physique de la route*, etc. pour BDCarto)
 - En sortie,
 - une ontologie au format OWL DL

Extrait de l'ontologie OBDTopo

owl:Thing	
▶	● _Feature
▼	● _GeographicEntity
▼	● http://www.owl-ontologies.com/Ontology1275314039.owl#A - Voies de communication routière
▶	● http://www.owl-ontologies.com/Ontology1275314039.owl#A - Voies de communication routière-Surface de route
▶	● http://www.owl-ontologies.com/Ontology1275314039.owl#A - Voies de communication routière-Tronçon de chemin
▼	● http://www.owl-ontologies.com/Ontology1275314039.owl#A - Voies de communication routière-Tronçon de route
●	http://www.owl-ontologies.com/Ontology1275314039.owl#A - Voies de communication routière-Tronçon de route-Autoroute
●	http://www.owl-ontologies.com/Ontology1275314039.owl#A - Voies de communication routière-Tronçon de route-Autoroutière
▼	● http://www.owl-ontologies.com/Ontology1275314039.owl#A - Voies de communication routière-Tronçon de route-Autre classement
●	http://www.owl-ontologies.com/Ontology1275314039.owl#A - Voies de communication routière-Tronçon de route-Autre classement-Rues
●	http://www.owl-ontologies.com/Ontology1275314039.owl#A - Voies de communication routière-Tronçon de route-Autre classement-Rues piétonnes
●	http://www.owl-ontologies.com/Ontology1275314039.owl#A - Voies de communication routière-Tronçon de route-Autre classement-Voies goudronnées (voies
●	http://www.owl-ontologies.com/Ontology1275314039.owl#A - Voies de communication routière-Tronçon de route-Contre-allée
▶	● http://www.owl-ontologies.com/Ontology1275314039.owl#A - Voies de communication routière-Tronçon de route-Départementale
●	http://www.owl-ontologies.com/Ontology1275314039.owl#A - Voies de communication routière-Tronçon de route-En construction
●	http://www.owl-ontologies.com/Ontology1275314039.owl#A - Voies de communication routière-Tronçon de route-Locale
▶	● http://www.owl-ontologies.com/Ontology1275314039.owl#A - Voies de communication routière-Tronçon de route-Nationale
●	http://www.owl-ontologies.com/Ontology1275314039.owl#A - Voies de communication routière-Tronçon de route-Principale
●	http://www.owl-ontologies.com/Ontology1275314039.owl#A - Voies de communication routière-Tronçon de route-Régionale
▶	● http://www.owl-ontologies.com/Ontology1275314039.owl#A - Voies de communication routière-Tronçon de chemin-Franchissement
▶	● http://www.owl-ontologies.com/Ontology1275314039.owl#A - Voies de communication routière-Tronçon de route-Franchissement
▶	● http://www.owl-ontologies.com/Ontology1275314039.owl#B - Voies ferrées et autres moyens de transport terrestre
▶	● http://www.owl-ontologies.com/Ontology1275314039.owl#B - Voies ferrées et autres moyens de transport terrestre-Tronçon de voie ferrée-Franchissement
▶	● http://www.owl-ontologies.com/Ontology1275314039.owl#C - Transport d'énergie et de fluides
▶	● http://www.owl-ontologies.com/Ontology1275314039.owl#D - Hydrographie terrestre
▶	● http://www.owl-ontologies.com/Ontology1275314039.owl#D - Hydrographie terrestre-Tronçon de cours d'eau-Franchissement

Nombre de concepts

1262 (312 MBD, 950 MR)

Nombre de relations

8 relations (4 MBD, 4 MR)

102 propriétés (62 MBD, 40 MR)

Extrait de l'ontologie OBDCarto

- ▶ ● <http://www.owl-ontologies.com/Ontology1275314039.owl#A2> équipements et sites touristiques
- ▼ ● <http://www.owl-ontologies.com/Ontology1275314039.owl#B>. réseau ROUTIeR
 - ▶ ● <http://www.owl-ontologies.com/Ontology1275314039.owl#B>. réseau ROUTIeR-B-c-1 Carrefour complexe
 - ▶ ● <http://www.owl-ontologies.com/Ontology1275314039.owl#B>. réseau ROUTIeR-B-c-2 Route
 - ▶ ● <http://www.owl-ontologies.com/Ontology1275314039.owl#B>. réseau ROUTIeR-B-c-3 Itinéraire routier
 - <http://www.owl-ontologies.com/Ontology1275314039.owl#B>. réseau ROUTIeR-B-c-4 Début de section
 - <http://www.owl-ontologies.com/Ontology1275314039.owl#B>. réseau ROUTIeR-B-rc-1 A pour successeur
 - <http://www.owl-ontologies.com/Ontology1275314039.owl#B>. réseau ROUTIeR-B-rs-1 Permet d'accéder à
 - <http://www.owl-ontologies.com/Ontology1275314039.owl#B>. réseau ROUTIeR-B-rs-2 Réseau routier
 - <http://www.owl-ontologies.com/Ontology1275314039.owl#B>. réseau ROUTIeR-B-rs-3 Réseau de bac
 - <http://www.owl-ontologies.com/Ontology1275314039.owl#B>. réseau ROUTIeR-B-rs-4 Communication restreinte
 - ▼ ● <http://www.owl-ontologies.com/Ontology1275314039.owl#B>. réseau ROUTIeR-B-s-1 Tronçon de route
 - <http://www.owl-ontologies.com/Ontology1275314039.owl#B>. réseau ROUTIeR-B-s-1 Tronçon de route-0
 - <http://www.owl-ontologies.com/Ontology1275314039.owl#B>. réseau ROUTIeR-B-s-1 Tronçon de route-1
 - <http://www.owl-ontologies.com/Ontology1275314039.owl#B>. réseau ROUTIeR-B-s-1 Tronçon de route-10
 - <http://www.owl-ontologies.com/Ontology1275314039.owl#B>. réseau ROUTIeR-B-s-1 Tronçon de route-2
 - <http://www.owl-ontologies.com/Ontology1275314039.owl#B>. réseau ROUTIeR-B-s-1 Tronçon de route-3
 - <http://www.owl-ontologies.com/Ontology1275314039.owl#B>. réseau ROUTIeR-B-s-1 Tronçon de route-4
 - <http://www.owl-ontologies.com/Ontology1275314039.owl#B>. réseau ROUTIeR-B-s-1 Tronçon de route-6
 - <http://www.owl-ontologies.com/Ontology1275314039.owl#B>. réseau ROUTIeR-B-s-1 Tronçon de route-7
 - <http://www.owl-ontologies.com/Ontology1275314039.owl#B>. réseau ROUTIeR-B-s-1 Tronçon de route-8
 - <http://www.owl-ontologies.com/Ontology1275314039.owl#B>. réseau ROUTIeR-B-s-1 Tronçon de route-9

Nombre de concepts	Nombre de relations
884 (306 MBD, 578 MR)	0 relation 159 propriétés (61 MBD, 98 MR)

Documentation des concepts

The screenshot displays the Protégé interface for editing an ontology. The main window is titled 'CLASS EDITOR for http://www.owl-ontologies.com/Ontology1275314039.owl#B. réseau ROUTIER-B-s-1'. The class being edited is 'http://www.owl-ontologies.com/Ontology1275314039.owl#B. _réseau_ROUTIER-B-s-1 _Tronçon'. The interface is divided into several panes:

- Subclass Explorer:** Shows a hierarchy of classes under 'owl:Thing'. The selected class is 'http://www.owl-ontologies.com/Ontology1275314039.owl#db_A1 bâtiments', which includes subclasses like 'Zone d'habitat', 'Zone d'activité', 'Établissement ac', and 'Bâtiments remar'.
- Class Editor Table:** A table with columns 'Property', 'Value', and 'Lang'. It lists properties such as 'rdfs:comment', 'definition', 'geometryDescription', 'Lemme', 'Origine', 'Reference', and 'Terme'.
- Asserted Condition:** A section for defining logical conditions, currently showing 'http://www.owl-ontologies.com/Ontology1275314039.owl#db_A1 bâtiments' with a 'NECESSARY & SUFFICIENT' condition.
- Disjoin:** A section for defining disjoint classes, currently empty.

Property	Value	Lang
rdfs:comment		
definition	Cette classe comprend toutes les zones d'habitat de la BDCarto ; sont retenus, sur le territoire national : - tous les groupes d'habitations couvrant une superficie supérieure ou égale à 8 hectares ; - les chefs-lieux de commune ; - la plupart des lieux-dits habités de 2 feux et plus, selon la densité de la zone ; - les autres lieux-dits habités situés, soit aux carrefours accessibles par le réseau classé départemental ou national, soit à l'extrémité des culs de sac de longueur supérieure ou égale à 1000 mètres du réseau routier carrossable. A l'étranger, seuls les chefs-lieux de commune (ou équivalent) sont	fr
geometryDescription	Une zone d'habitat est localisée par un sommet géométrique situé en son centre géographique (relation [1,1][0,1]) ; de plus, les zones d'habitat couvrant plus de huit hectares apparaîtront comme zone d'occupation du sol, poste "bâti". Couche géométrique : toponymes	fr
Lemme	A1-s-1 Zone d'habitat	
Origine	Structure	fr
Reference	A1 bâtiments-A1-s-1 Zone d'habitat	
Terme	1 Zone d'habitat	

Discussion BDcarto : concepts étiquetés par des valeurs numériques

The screenshot shows the Protégé 3.4.1 interface. The 'SUBCLASS EXPLORER' on the left lists a hierarchy of classes under 'réseau ROUTIER', including various types of roads and sections. The 'CLASS EDITOR' on the right is set to the class 'Tronçon de route-0'. It displays a table of properties and their values:

Property	Value
rdfs:comment	
definition	liaison principale : Cette valeur représente une densification du maillage routier défini par les tronçons de type autoroutier.
definition	1 chaussée
definition	1 voie
definition	route non revêtue
Lemme	@card@
Origine	Structure
Reference	B. réseau ROUTIER-B-s-1 Tronçon de route- Vocation de la liaison-1
Reference	B. réseau ROUTIER-B-s-1 Tronçon de route- Nombre de chaussées-1
Reference	B. réseau ROUTIER-B-s-1 Tronçon de route- Nombre total de voies-1
Reference	B. réseau ROUTIER-B-s-1 Tronçon de route- Etat physique de la route-1
Reference	B. réseau ROUTIER-B-s-1 Tronçon de route- Nombre de voies chaussée montante
Reference	B. réseau ROUTIER-B-s-1 Tronçon de route- Nombre de voies chaussée descendante-1
Terme	1

At the bottom right, a status bar indicates 'Chargement terminé (100 %)' and 'Plan d'alimentation actuel: Recommandé par HP'.

→ 1 seul concept pour plusieurs concepts labellisés par la même valeur (1 dans l'exemple)

→ On ne sait plus à quelle Référence correspond une Définition

Discussion BDcarto : erreurs dans les spécifications XML

Document Word

B. RESEAU ROUTIER

B-s-1 Tronçon de route

c Attributs

[1] Vocation de la liaison

Cet attribut matérialise une hiérarchisation du réseau routier basée, non pas sur un critère administratif, mais sur l'importance des tronçons de route pour le trafic routier. Ainsi, les quatre valeurs « type autoroutier », « liaison principale », « liaison régionale » et « liaison locale » permettent un maillage de plus en plus dense du territoire.

- 1- **type autoroutier** : autoroute ou voie route à chaussées séparées et carrefours dénivelés (avec entrée ou sortie par une bretelle) sur au moins 5 km.
- 2- **liaison principale** : Cette valeur représente une densification du maillage routier défini par les tronçons de type autoroutier.
Les liaisons principales ont fonction :
 - d'assurer les liaisons à fort trafic à caractère prioritaire entre agglomérations importantes ;
 - d'assurer les liaisons des agglomérations importantes au réseau autoroutier ;
 - d'offrir une alternative à une autoroute si celle-ci est payante ;
 - de proposer des itinéraires de contournement des agglomérations ;
 - d'assurer la continuité, en agglomération, des liaisons interurbaines à fort trafic quand il n'y a pas de contournement possible.
- 6- **liaison régionale** : cette valeur représente une densification du maillage routier défini par les tronçons de type autoroutier et liaison principale.
Les liaisons régionales ont fonction, quand celle-ci n'est pas assurée par des itinéraires de vocation plus élevée :

Document XML

```
<attributes>
  <attribute name=" Vocation de la liaison">
    <attributeName> Vocation de la liaison</attributeName>
    <valueType>Énuméré</valueType>
    <description type="definition"> Cet attribut matérialise une hiérarchisation du réseau routier
    basée, non pas sur un critère administratif, mais sur l'importance des tronçons de route pour le trafic routier.
    Ainsi, les quatre valeurs " type autoroutier ", " liaison principale ", " liaison régionale " et " liaison locale "
    permettent un maillage de plus en plus dense du territoire.</description>
    <enumeratedValues>
      <value name=" type autoroutier : autoroute ou voie route à chaussées séparées et
      carrefours dénivelés (avec entrée ou sortie par une bretelle) sur au moins 5 km.">
        <valueName> type autoroutier : autoroute ou voie route à chaussées
        séparées et carrefours dénivelés (avec entrée ou sortie par une bretelle) sur au moins 5 km.</valueName>
        <description type="definition"></description>
      </value>
      <value name="1">
        <valueName>1</valueName>
        <description type="definition"> liaison principale : Cette valeur représente
        une densification du maillage routier défini par les tronçons de type autoroutier.</description>
      </value>
      <value name="2">
        <valueName>2</valueName>
        <description type="definition"> Les liaisons principales ont fonction
    </description>
  </value>
```

→ Plusieurs cas similaires

Discussion BDcarto : incapacité à acquérir un certain type d'information

```
<attributes>
  <attribute name=" Vocation de la liaison">
    <attributeName> Vocation de la liaison</attributeName>
    <valueType>Enuméré</valueType>
    <description type="definition"> Cet attribut matérialise une hiérarchisation du réseau routier
    basée, non pas sur un critère administratif, mais sur l'importance des tronçons de route pour le trafic routier.
    Ainsi, les quatre valeurs " type autoroutier ", " liaison principale ", " liaison régionale " et " liaison locale "
    permettent un maillage de plus en plus dense du territoire.</description>
    <enumeratedValues>
      <value name=" type autoroutier : autoroute ou voie route à chaussées séparées et
      carrefours dénivelés (avec entrée ou sortie par une bretelle) sur au moins 5 km.">
        <valueName> type autoroutier : autoroute ou voie route à chaussées
        séparées et carrefours dénivelés (avec entrée ou sortie par une bretelle) sur au moins 5 km.</valueName>
        <description type="definition"></description>
      </value>
      <value name="1">
        <valueName>1</valueName>
        <description type="definition"> liaison principale : Cette valeur représente
        une densification du maillage routier défini par les tronçons de type autoroutier.</description>
      </value>
      <value name="2">
        <valueName>2</valueName>
        <description type="definition"> Les liaisons principales ont fonction
      </description>
    </enumeratedValues>
  </attribute>
</attributes>
```

Tronçon de route

String ∈ { Type autoroutier : ..., 1, 2, 6, 7, 8, 10 }

a-pour-Vocation-de-la-liaison

→ Impossibilité d'associer à une des valeurs énumérées sa définition

Discussion BDcarto : paragraphe comme valeur d'énumération

```
<attribute name=" Date de mise en service">
  <attributeName> Date de mise en service</attributeName>
  <valueType>Énuméré</valueType>
  <description type="definition"></description>
  <enumeratedValues>
    <value name="Cet attribut n'est rempli que pour les tronçons en construction (valeur 3 pour
l'attribut Etat physique) ; mois et année de prévision de mise en service pour les tronçons en travaux. Il est sans
objet dans tous les autres cas.">
      <valueName>Cet attribut n'est rempli que pour les tronçons en construction (valeur 3
pour l'attribut Etat physique) ; mois et année de prévision de mise en service pour les tronçons en travaux. Il est
sans objet dans tous les autres cas.</valueName>
      <description type="definition"></description>
    </value>
  </enumeratedValues>
</attribute>
```

→ La valeur dans le DataRange (ou le label d'un concept) est un paragraphe

Discussion BDcarto : attribut sans type

```
<attribute name=" Toponyme">
  <attributeName> Toponyme</attributeName>
  <description type="definition"> Un nœud du réseau routier peut porter un toponyme, si l'un au moins
des tronçons connectés appartient au réseau classé, et si le nœud appartient à l'un des types suivants : Un nœud
composant un carrefour complexe ne porte généralement pas de toponyme. Le toponyme est composé de trois
parties pouvant ne porter aucune valeur (n'existe pas dans les cas ci-dessus et sans objet sinon) : - 11 :
carrefour simple, - 18 : rond-point (diamètre &lt; 30 mètres), - 15 : carrefour aménagé avec passage inférieur ou
toboggan. - un terme générique ou une désignation, texte d'au plus 40 caractères ; - un article, texte d'au plus
cinq caractères ; - un élément spécifique, texte d'au plus 80 caractères.</description>
</attribute>
```

→ Analyse des spécifications essentiellement basée sur le type des attributs

→ Attributs non traités

Bilan - Conclusion

- 1ère approche qui avait été proposée :
 - Génération de règles d'extraction qui offrait un certain degré de généralité
 - Les définitions étaient acquises et associées aux concepts correspondant
 - 2 ontologies OBDDTopo et OBDDCarto exploitables
 - Problème de rattachement des concepts issus des attributs (sous TOP ?)

Bilan - Conclusion

- Nouvelle approche :
 - Le problème de rattachement est résolu à l'aide des DataRange properties
 - Approche valide pour BDTopo (BDCarto ?)
 - Perte de la généricité
 - Les attributs représentant des concepts topographiques varient d'une BD à l'autre
 - Traitement spécifique des valeurs numériques d'attribut présentes dans BDCarto : écrasement des concepts