

Concepts pour urbanisation des ressources en données par la sémantique et articulation avec les IDG

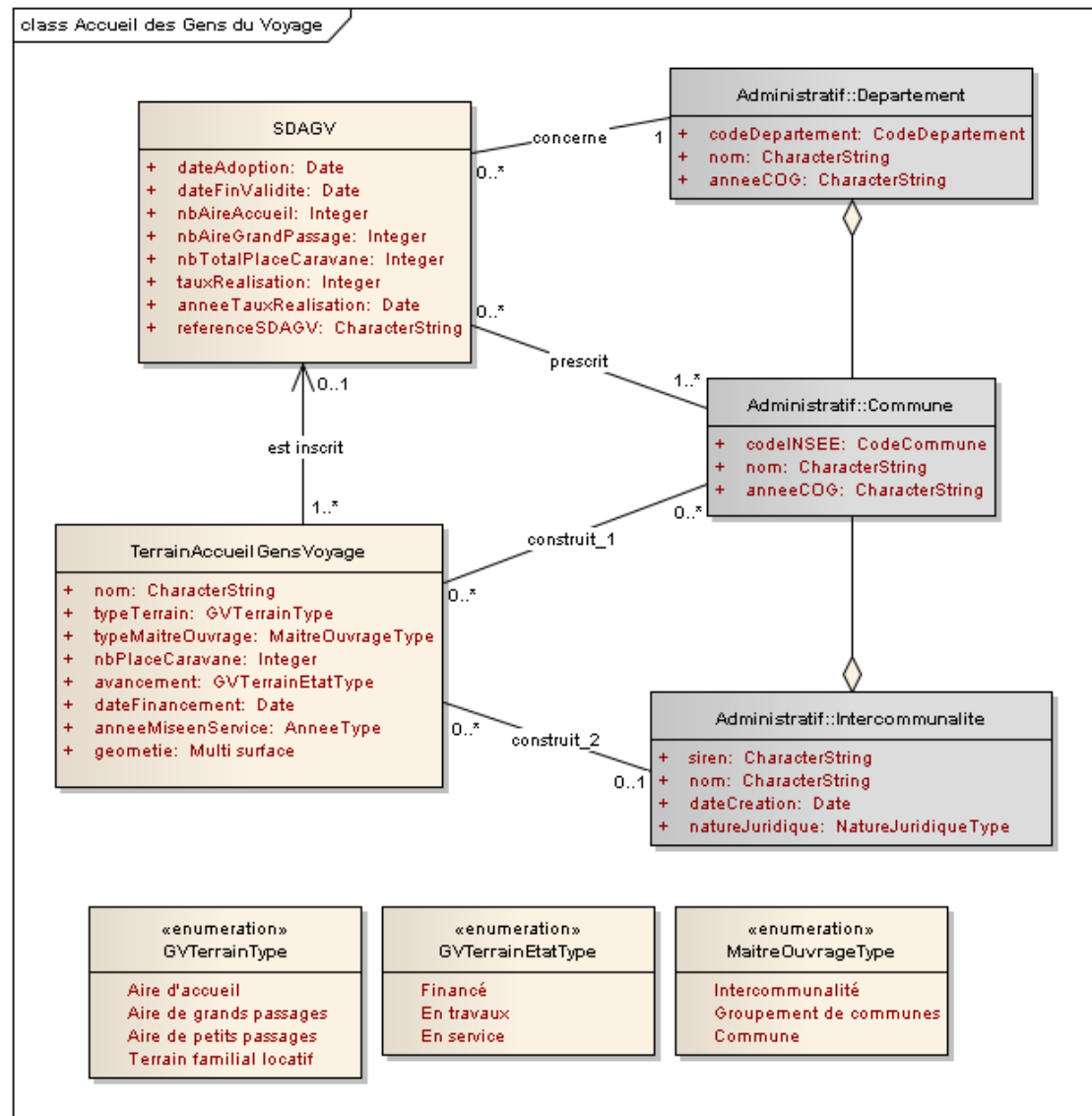


Clément Jaquemet – MTES/CGDD/DRI/SDI/MIG
Février 2018

Définition en modélisation UML des standards

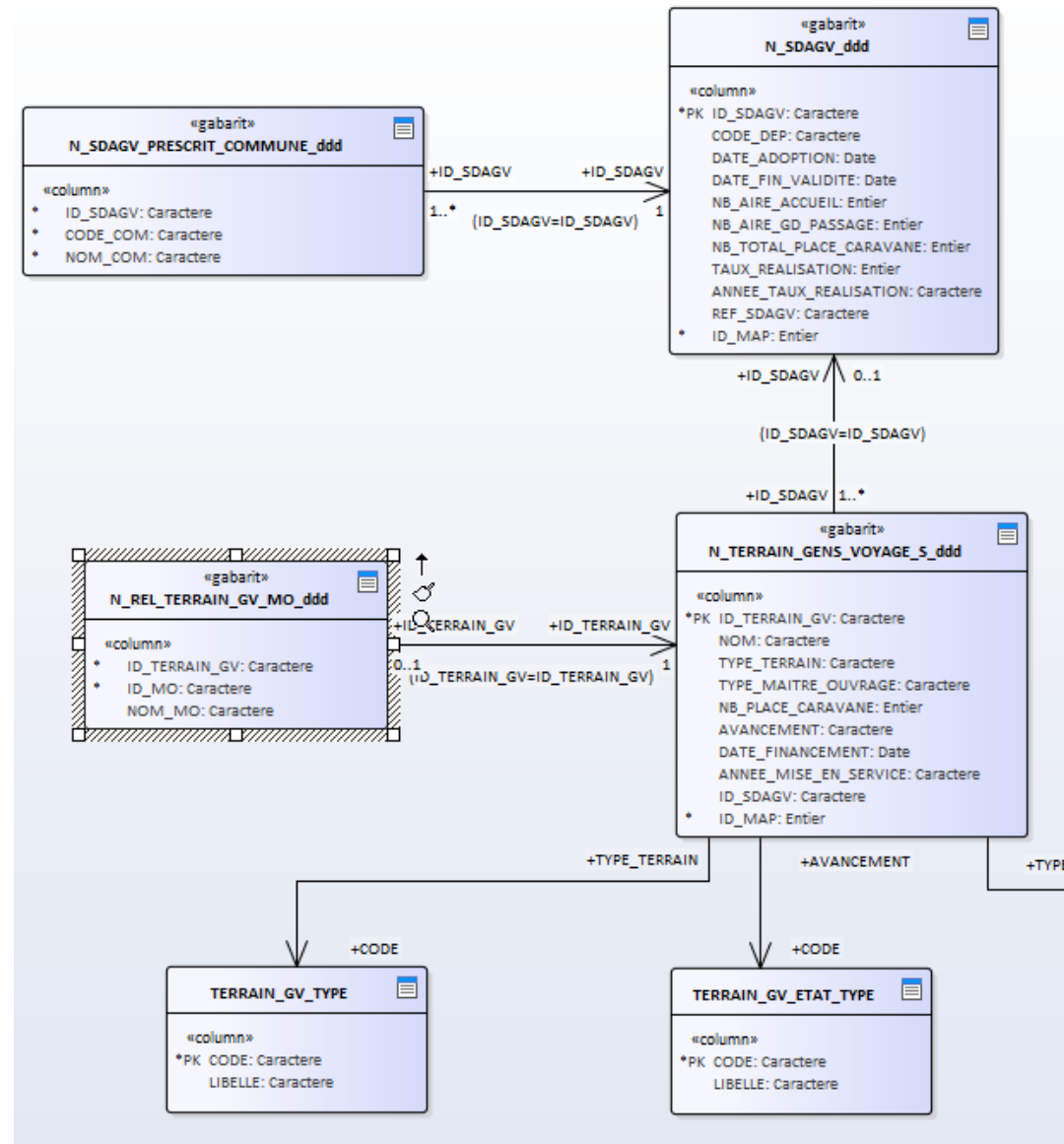
Ex: AGV (Accueil des gens du voyage)

- La classe « commune » est décrite.
- La sémantique de la relation « SDAGV-prescrit-Commune » est présente



Traduction du standard en MDL puis en ressource Prodige / base Postgres

- La classe « Commune » à disparue, elle est traduite par deux propriétés qui la fusionnent avec la relation « prescrit ». (rem : l'année du COG est perdue)
- Dans l'IDG une autre ressource « Communes » peut exister indépendamment. On devra faire une requête SQL pour trouver les infos de la communes



Production des métadonnées (MD) des standards

- A partir du standard on produit des fiches de MDD modèles, ainsi que les fiches catalogues de type de données.

- http://geostandards.developpement-durable.gouv.fr/afficherPageStandard.do?ieii=N_SDAGV

Jeu de données

Fiche de métadonnées Catalogue d'attributs Tables de gabarit

GéoStandard(s) référençant le jeu : [Accueil des gens du voyage](#)

[Télécharger la fiche](#) [Télécharger le fichier de règles](#)

Métadonnées génériques du jeu N_SDAGV	
Date d'actualisation	14/03/2014
Identifiant de la fiche	N_SDAGV_ddd
Intitulé de la ressource	Schéma départemental d'accueil pour les gens du voyage
Résumé de la ressource	Le schéma départemental pour l'accueil des gens du voyage est rendu obligatoire par la loi du 5 juillet 2000 relative à l'accueil des gens du voyage. Le SDAGV précise en outre « la destination des aires permanentes d'accueil, permanentes d'accueil et les communes où celles-ci doivent être réalisées ». Le SDAGV précise en outre « la destination des aires permanentes d'accueil, permanentes d'accueil et les communes où celles-ci doivent être réalisées ». Le SDAGV fixe le nombre d'aires d'accueil à réaliser et le nombre de places de caravane à réaliser. Il est révisable tous les 6 ans à compter de la date de signature du SDAGV. Les communes de plus de 5000 habitants figurent obligatoirement dans le SDAGV si elles ont donné au préalable leur accord ou si elles en ont fait la demande. La gestion d'une aire réalisée en commun par plusieurs communes se fait par le biais d'une convention intercommunale.
Catégories thématiques	Société Planification/Cadastre
Thème INSPIRE	
Mots-clés	HABITAT_POLITIQUE_DE_LA_VILLE/N_POLITIQUE
Géométrie	composé
Type de représentation spatiale	vecteur
Résolution spatiale	1/2000
Zone d'application	France métropolitaine
Généalogie de la ressource	Ces données sont issues du SDAGV adopté par le préfet de département et le président du conseil général. Il est révisable tous les 6 ans à compter de la date de signature du SDAGV. Les communes de plus de 5000 habitants figurent obligatoirement dans le SDAGV si elles ont donné au préalable leur accord ou si elles en ont fait la demande. La gestion d'une aire réalisée en commun par plusieurs communes se fait par le biais d'une convention intercommunale.
Sources des données	sans objet - Standard de données COVADIS Accueil des gens du voyage, version 1 - : Source thématique : schéma départemental approuvé ...
Restrictions sur l'accès	Pas de restriction d'accès public selon INSPIRE

Jeu de données

Fiche de métadonnées Catalogue d'attributs Tables de gabarit

Propriétés du jeu N_SDAGV

Identifiant de la fiche de métadonnées	N_SDAGV_ddd
Définition	Table contenant la liste des schémas départementaux d'accueil pour les gens du voyage

Catalogue d'attributs du jeu N_SDAGV

Nom	Type/Longueur	Définition
ID_SDAGV	chaîne (5)	Identifiant unique du SDAGV correspondant à la concaténation de CODE_DEP et DATE_ADOPTION
CODE_DEP	chaîne (3)	Code INSEE du département sur lequel s'applique le SDAGV
DATE_ADOPTION	date	Date de signature du schéma départemental
DATE_FIN_VALIDITE	date	Date de fin de validité du schéma départemental
NB_AIRE_ACCUEIL	entier	Nombre fixé par le SDAGV d'aires d'accueil à réaliser pour le département
NB_AIRE_GD_PASSAGE	entier	Nombre fixé par le SDAGV d'aires de grands passages à réaliser
NB_TOTAL_PLACE_CARAVANE	entier	Nombre total de places de caravane à réaliser en aire d'accueil
TAUX_REALISATION	entier	Taux (en pour-cent) de réalisation du schéma départemental égal au rapport entre le nombre total de places de caravane prescrites par le schéma et le nombre total de places de caravane réalisées
ANNEE_TAUX_REALISATION	chaîne (4)	Année de calcul du taux de réalisation du SDAGV
REF_SDAGV	chaîne (254)	Nom du fichier contenant le schéma départemental approuvé
ID_MAP	entier	Identifiant technique à rajouter pour un stockage de la table en base de données

Compléter la méthode de standardisation de la Covadis par :

- **Une méthode de structuration permettant une « conception structurée »,**
- **Une méthode de validation sémantique visant à la validation formelle des modèles,**
- **Des référentiels sémantiques aidant à l'analyse et à la conception des relations et des propriétés dans les modèles,**
- **Une méthode d'intégration de modèles sur la base d'une analyse sémantique,**
- **Une méthode de structuration de la connaissance permettant un meilleur passage vers le niveau logique.**

Comment traduit le standard UML en ontologie ?

- Distinguer ontologies linguistiques (synonyme, hyperonyme, hyponyme, etc.) et ontologies conceptuelles (représenter des concepts généraux, indépendamment de leurs représentations dans un langage particulier)
- Correspondances utilisées entre les métamodèles UML et RDF Schema : permet la production d'une ontologie « locale »

UML ☒	RDFS ☒
Class ☒	<code>rdfs:Class</code> ☒
Generalization ☒	<code>rdfs:subClassOf</code> ☒
Association ☒	<code>rdf:Property</code> ☒
Attribute ☒	<code>rdf:Property</code> ☒
InstanceOf ☒	<code>rdf:type</code> ☒
Attribute type : String ☒	<code>rdfs:Literal</code> ☒
Attribute value ☒	<code>rdf:value</code> ☒

Traduction du modèle UML AGV en ontologie locale

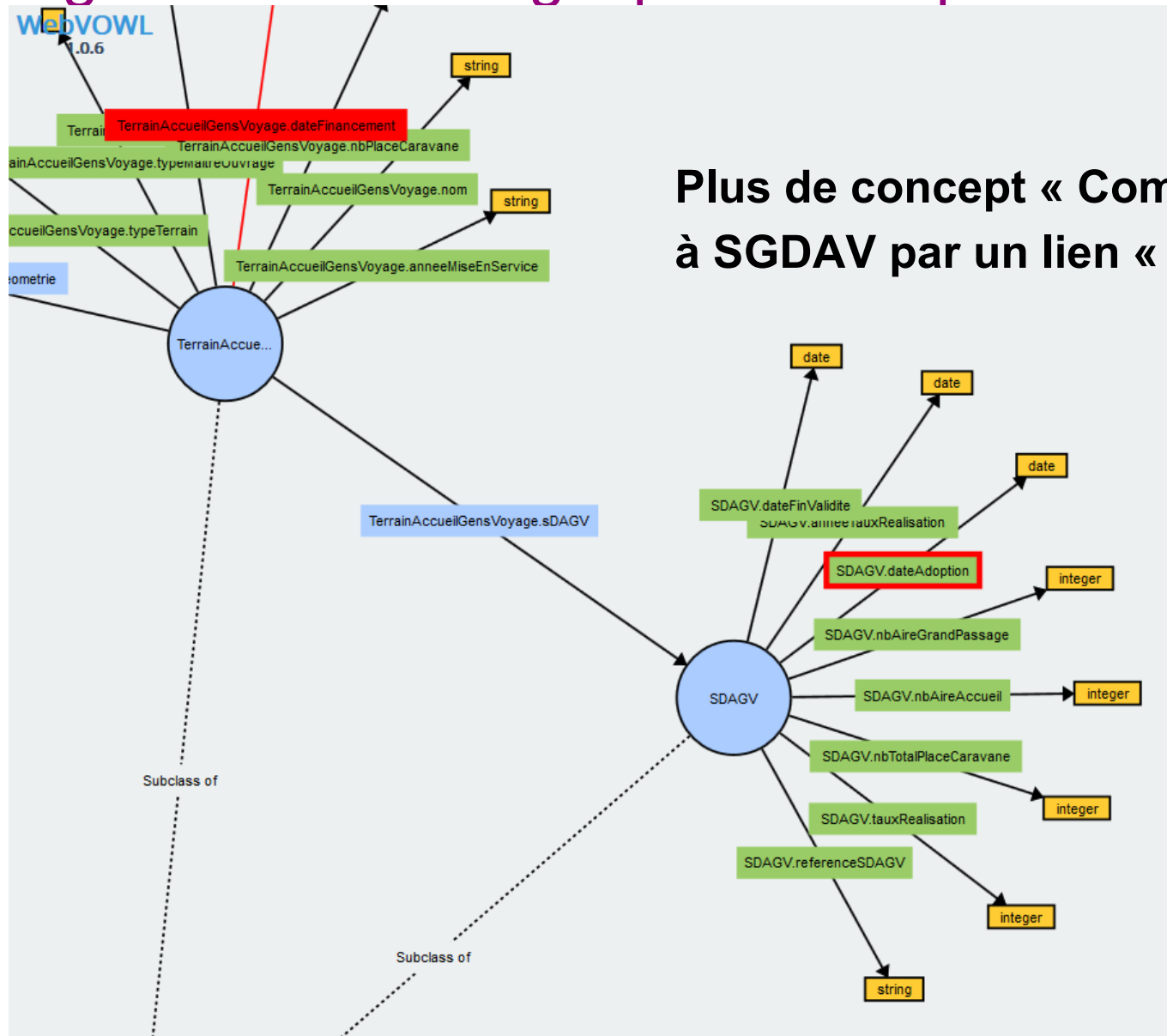
```
<http://data.cerema.fr/covadis/owl/agv>      a      owl:Ontology ;
  owl:imports
    <http://www.opengis.net/ont/gml> ,
    <http://www.opengis.net/ont/geosparql> ;

  owl:versionIRI   "http://data.cerema.fr/covadis/owl/agv/1.0" ;
  owl:versionInfo "1.0" ;
  skos:definition   "-- Name --\r\nSchéma d'application de l'Accueil des
  Gens du Voyage (AGV)\r\n-- Definition --\r\nSchéma d'application de
  l'Accueil des Gens du Voyage (AGV)"@fr
.../...
agv:SDAGV a      owl:Class ;
  rdfs:subClassOf gsp:Feature ;
agv:SDAGV.anneeTauxRealisation
  a      owl:DatatypeProperty ;
  rdfs:domain agv:SDAGV ;
  rdfs:range  xsd:date .
.../...
```

Attention :

il faut assurer la pérennité de l'URI de l'ontologie = standard+version

Perte d'information si passage via le « schéma de règle » sur l'ontologie produite à partir du MDL



Plus de concept « Commune » relié à SGDAV par un lien « prescrit-par »

Réutilisation d'ontologies externes à intégrer dans la modélisation UML et sa traduction en ontologie

- Pour les communes, par exemple celle de l'INSEE ? :

```
<owl:Ontology rdf:about="http://data.lirmm.fr/ontologies/geo-fr">
  <rdf:type rdf:resource="http://www.w3.org/ns/adms#SemanticAsset"/>
  <dcterms:title>Ontologie géographique de l'INSEE</dcterms:title>
  <dcterms:description>Ontologie des objets géographiques, en particulier des territoires administrat
  <dcterms:modified rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2011-05-04</dcterms:modified
  <vann:preferredNamespaceUri>http://rdf.insee.fr/geo/</vann:preferredNamespaceUri>
  <vann:preferredNamespacePrefix>geo-fr</vann:preferredNamespacePrefix>
  <foaf:homepage rdf:resource="http://data.lirmm.fr/ontologies/geo-fr.html"/>
  <dcterms:created rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2011-05-04</dcterms:created>
  <dcterms:publisher>http://data.lirmm.fr/ontologies/geo-fr#LIRMM</dcterms:publisher>
  <dcterms:partOf rdf:resource="http://data.lirmm.fr/ontologies"/>
  <dcterms:type rdf:resource="http://purl.org/adms/assettype/Ontology"/>
  <dcterms:status rdf:resource="http://purl.org/adms/status/UnderDevelopment"/>
  <dc:creator rdf:resource="http://data.lirmm.fr/ontologies/geo-fr#fran%C3%A7ois"/>
  <dc:creator rdf:resource="http://data.lirmm.fr/ontologies/geo-fr#Franck%20cotton"/>
  <dc:creator rdf:resource="http://data.lirmm.fr/ontologies/geo-fr#Bernard%20Vatant"/>
</owl:Ontology>
```

- En particulier, urbaniser les standards avec les modèles INSPIRE ou européen (ISA2, etc.)
- Linked Open Vocabularies (LOV) ? (630 vocabulaires, dont 50 en français) et Schema.org ?
- Distinguer : ontologies locales (domaine spécifique), et ontologies génériques (états, actions, espace...)

INSPIRE : Guidelines for the RDF encoding of spatial data

▪ Exigences (exemples)

- Le nom de l'ontologie doit être construit comme suit :
 - ontology-name = "http://inspire.ec.europa.eu/ont/" + app-schema-code
- L'espace de noms pour les composants d'ontologie doit être :
 - ontology-namespace = ontology-name + "#"
- La *version* doit également être utilisée pour affecter une *version IRI* à l'ontologie, qui est définie comme suit :
 - versionIRI = ontology-name + "/" + version
- Si une classe RDFS / OWL d'un vocabulaire existant est connue pour représenter de manière appropriée la sémantique d'un type UML, alors le type doit être mappé à cette classe.

▪ Des conseils d'encodage sur des propriétés classiques

- EX : Encodage des éléments de documentation courants

Table 1. Encoding of common documentation items

Documentation item	Encoding with RDF/OWL property
natural / human readable name	rdfs:label
definition	skos:definition
Description	rdfs:comment

Mappage des types de l'ISO 19103 pour INSPIRE

Table 2. Mapping of types from ISO 19103

ISO 19103 Type	RDFS/OWL Class or Datatype
Boolean	xsd:boolean
CharacterString	xsd:string
Date	xsd:date
DateTime	xsd:dateTime
Decimal	xsd:decimal
Integer	xsd:integer
Number	xsd:double
Real	xsd:double
Any	rdfs:Class
Angle	base:Measure
Area	base:Measure
Distance	base:Measure
Length	base:Measure
Measure	base:Measure
Probability	base:Measure
Velocity	base:Measure
Volume	base:Measure

Mappage des types géométriques pour INSPIRE

Table 3. Mapping of types from ISO 19107 to their implementations as GeoSPARQL Geometry subclasses		
ISO 19107 Types	Simple Feature Ontology Class	GML Ontology Class
GM_Aggregate	sfowl:GeometryCollection	gmlowl:MultiGeometry
GM_Curve	sfowl:LineString	gmlowl:Curve or gmlowl:LineString
GM_MultiCurve	sfowl:MultiLineString	gmlowl:MultiCurve
GM_MultiPoint	sfowl:MultiPoint	gmlowl:MultiPoint
GM_MultiPrimitive	sfowl:GeometryCollection	gmlowl:MultiGeometry
GM_MultiSurface	sfowl:MultiPolygon	gmlowl:MultiSurface
GM_Object	sfowl:Geometry	gmlowl:AbstractGeometry
GM_Point	sfowl:Point	gmlowl:Point
GM_PolyhedralSurface	sfowl:PolyhedralSurface	gmlowl:PolyhedralSurface
GM_Primitive	sfowl:Geometry	gmlowl:AbstractGeometricPrimitive
GM_Ring	sfowl:LinearRing	gmlowl:Ring or gmlowl:LinearRing
GM_Solid	-	gmlowl:Solid
GM_Surface	sfowl:Polygon	gmlowl:Surface or gmlowl:Polygon
GM_Tin	sfowl:TIN	gmlowl:Tin
GM_Triangle	sfowl:Triangle	gmlowl:Triangle

Articulation avec ISA² et avec les registres

▪ **Articulation avec ISA² : Core Vocabularies**

- Ex. : L'implémentation RDF / OWL de 'Address' du schéma 'INSPIRE Annex I - Adresses' doit être une classe rdfs: subClass de la classe "Address« de [ISA Programme Location Core Vocabulary](#)

▪ **Si la signification des codes d'une énumération est évidente, ou si les codes sont auto-descriptifs, alors l'énumération devrait être représentée par un OWL *DataOneOf* ,**

- c'est-à-dire un type de données RDFS qui spécifie la liste restreinte de littéraux utilisant une déclaration owl: oneOf. ,

▪ **Sinon, l'énumération doit être encodée comme une liste de codes :**

- Les listes de codes INSPIRE - et les extensions - doivent être représentées comme des schémas conceptuels SKOS, leurs codes étant des concepts SKOS. Ils doivent être publiés et gérés dans le registre INSPIRE.

Enrichissement des standards en sémantique

- **Pour la gestion des standards :**
 - Gérer les liens entre MD de données (MDD) et MD catalogue d'attributs (feature catalogue : MDFC)
 - Gérer des URI pour toutes ces fiches de MD
 - Déjà nécessaire pour la gestion des standards dans Prodigé V4.1 avec [ajout d'un lien de la MD vers les scripts SQL pour créer les standards en base PostGIS](#)
- **Pour la gestion des modèles**
 - Enrichir la modélisation UML des informations sémantiques (ajout des concepts des ontologies)
 - Enrichir/utiliser les outils de modélisation (Enterprise Architect) pour produire ou intégrer les ontologies et produire des fiches de MD enrichies
 - Utiliser les capacités d'extension d'Enterprise Architect « Model Driven Generation (MDG) » pour personnaliser les modèles d'EA
 - Utilisation des « tagged values » et de stéréotypes
 - Ajouter les transformations personnalisées sur ces modèles enrichis

Exemple d'enrichissement des propriétés

Ajouter les IRI des concepts correspondant, par exemple :

L'attribut « CodeINSEE » de la classe « Administratif::Commune » est enrichi par dans le modèle UML par l'IRI du concept correspondant (exemple avec l'ontologie INSEE <http://rdf.insee.fr/def/geo/insee-geo-onto.ttl> , des équivalences sont établies avec les ressources RDF publiées par l'IGN, voir <http://data.ign.fr>).

donc la propriété « CODE_COM » de N_SDAGV_PRESCRIT_COMMUNE_ddd hérite de cet enrichissement, ce qui permet de garder la complétude sémantique et de faire du contrôle sur les types de données.

- **code insee de la commune**
- **IRI : <http://rdf.insee.fr/def/geo#codeCommune>**
 - **igeo:codeCommune a owl:DatatypeProperty ;**
 - **rdfs:domain igeo:CommuneOuAssimilee ;**
 - **rdfs:isDefinedBy <<http://rdf.insee.fr/def/geo>> ;**
 - **rdfs:label "code commune"@fr ;**
 - **rdfs:range xsd:token ;**
 - **rdfs:subPropertyOf igeo:codeINSEE .**
- **Rem : les éléments de l'ontologie INSEE manquent de propriétés d'annotation lisibles pour les humains (rdfs:comment ou skos:definition), mais on les trouvent dans l'ontologie de l'IGN.**

Catalogue d'attributs du jeu N_SDAGV_PRESCRIT_COMMUNE			
Nom	Type/Longueur	Définition	Obligatoire
ID_SDAGV	chaîne (5)	Identifiant du SDAGV	OUI
CODE_COM	chaîne (5)	Code INSEE de la commune concernée par une prescription du SDAGV	OUI
NOM_COM	chaîne (50)	Nom de la commune	OUI

Définir le mappage sémantique entre modèles

- **Utiliser les correspondances sémantiques entre propriétés et classes pour urbaniser et déclarer/générer les règles de transformations entre des classes de plusieurs standards**
 - En particulier, proposer des vues INSPIRE des standards
 - Possibilité d'ajouter ces scripts de traduction /ou création de vues via des liens dans les fiches de MD modèle
- **Le niveau sémantique ajouté en chapeau du niveau conceptuel des standards permettra de communiquer, capitaliser et mettre en cohérence les standards :**
 - mise en évidence des redondances, des concepts sous-entendus, des choix d'agrégation et de représentation implicites, etc.
- **L'urbanisation peut se faire de manière itérative, en comparant la description de deux standards pour identifier les concepts similaires ou sémantiquement très proches en s'appuyant sur la mise en place de référentiels sémantiques :**
 - catalogues de types de relation et de types de données, vocabulaire, etc.

Principe méthodologique

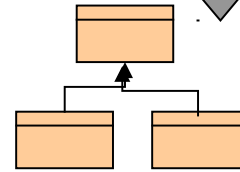
Niveau sémantique

Ontologies linguistiques (vocabulaire standard et métiers)

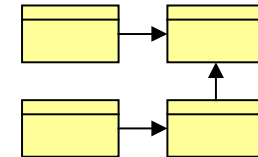
Ontologies locales (par standard+Base Covadis)

Ontologies externes (INSPIRE+autres)

Niveau Conceptuel UML




EOLIEN




AGV

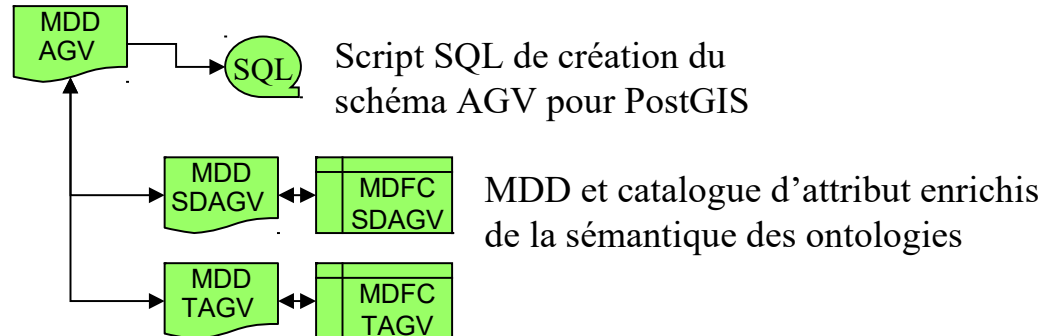
Niveau Ressource IDG

Légende :

 Métadonnées de données


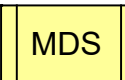
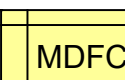
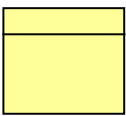
 Métadonnées Catalogue d'Attributs

URI : <http://geostandards/agv/1.0>



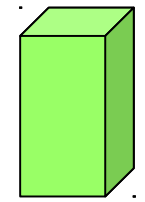
Utilisation de la MDD du standard pour créer les ressources


Légende :

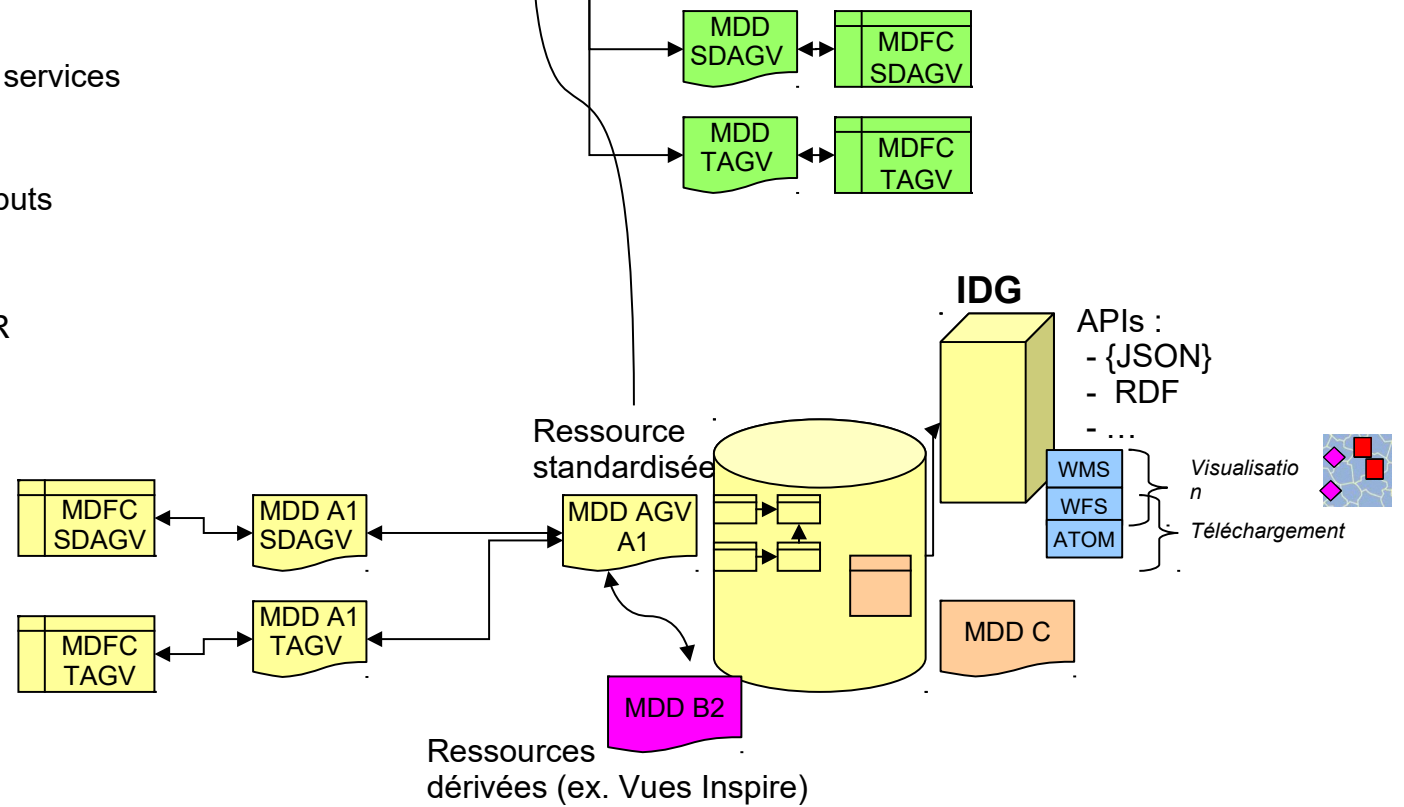
-  Métadonnées de données
-  Métadonnées de services
-  Métadonnées Catalogue d'Attributs
-  Tables de SGBDR

Serveur de standards

URI : <http://geostandards/agv/1.0>



 Script SQL de création du schéma AGV pour PostGIS

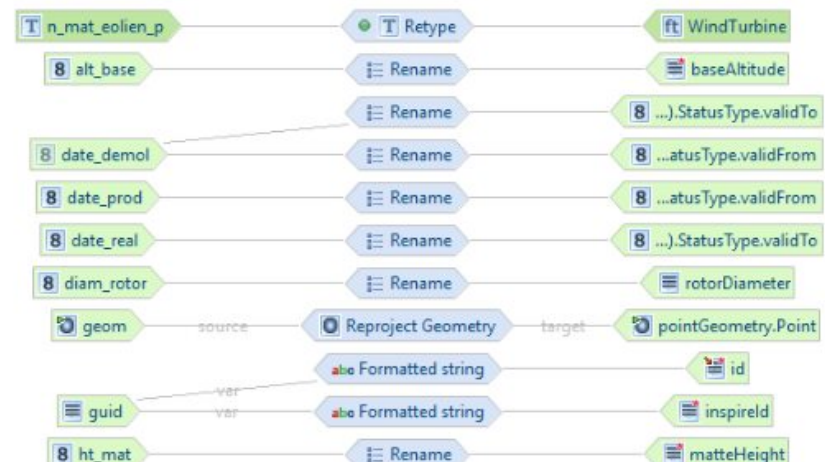


Actuel mapping INSPIRE d'un standard

- Exemple: processus actuel de l'extension INSPIRE du standard « Éolien terrestre version 2.0 – 29 mai 2017 »**
 - Correspondance entre deux schémas conceptuels
 - Mise en correspondance avec les attributs
 - Import des données en PostGIS
 - Production du schéma .xsd des données GML Inspire (à partir du projet EA Inspire par ShapeChange-2.0.0)
 - Rem : Ils sont téléchargeables sous le « schema repository » maintenu par le JRC <http://inspire.ec.europa.eu/schemas/>
 - Transformation par HALE (HUMBOLDT Alignment Editor)
 - Entrée : PostGIS
 - Sortie .xsd GML
 - Règles suivant correspondances

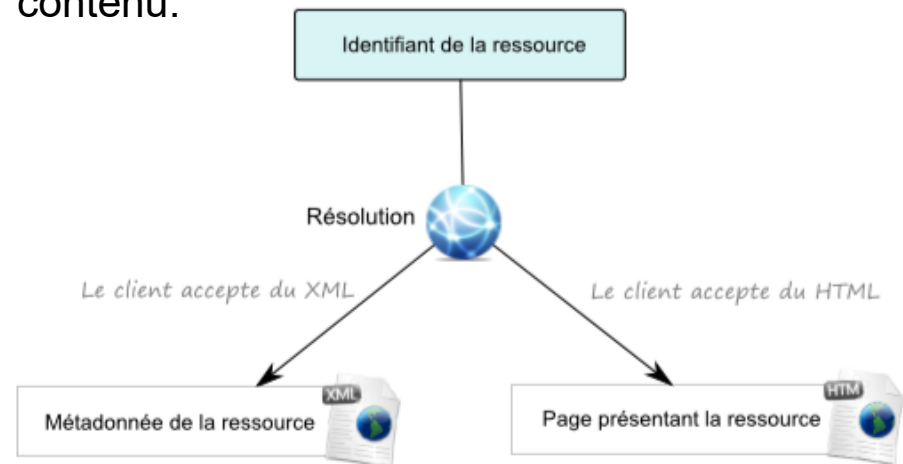
source et le jeu des exemples données :

Schéma d'extension INSPIRE		Règle	Gabarit COVADIS	
Type	Attribut / rôle d'association / contrainte		Table	Champ / Relation
<WindTurbine>		retype	N_MAT_EOLIEN_P	
<WindTurbine>	baseAltitude	rename	N_MAT_EOLIEN_P	ALT_BASE
<WindTurbine>	bladeHeight	rename	N_MAT_EOLIEN_P	HT_MAX
<WindTurbine>	id	formatted string (guid)	-	guid
<WindTurbine>	identifier	rename	N_MAT_EOLIEN_P	ID_MAT
<WindTurbine>	inspireId	formatted string (guid)	-	guid
<WindTurbine>	matteHeight	rename	N_MAT_EOLIEN_P	HT_MAT
<WindTurbine>	nacelleHeight	rename	N_MAT_EOLIEN_P	HT_NACELLE
<WindTurbine>	nationalCadastralReference	rename	N_MAT_EOLIEN_P	N_PARCEL
<WindTurbine>	nominalPower	rename	N_MAT_EOLIEN_P	PII_NOMINAL



Dans une IDG on a donc une MD par ressource

- cf. le guide du CNIG sur les URI (= <MD_Identifier> de la MD)
- **RECOMMANDATIONS NATIONALES :**
 - L'identifiant de la ressource (jeu de données ou service) devra pointer a minima vers la fiche de métadonnées en XML.
 - La ressource vers laquelle pointe l'identifiant peut également être choisie via un mécanisme de négociation de contenu.



- **une ressource peut avoir une ou plusieurs IRU, par ex. :**
 - <MD_Identifier> = MonIDG/{fileidentifieur} (URI résolvable par l'IDG)
 - <MD_Identifier> = http://geocatalogue.fr/id/{fileidentifieur} (résolvable par le géocatalogue)

Une API - Open API Initiative (OAI) pour l'IDG ?

Exemple d'une API globale à l'IDG pour télécharger ses ressources en GeoJSON.

- Sur « MonIDG/ » :
 - `{dsid}` étant l'ID interne du dataset (« fileIdentifiant » par ex.)
- Une ressource peut regrouper plusieurs classes d'objet `{sot}`

- Descendre à l'URI de l'instance ?
 - Données figées
 - Données vivantes

- Offrir aussi du JSON-LD ?
- Offrir un point SPARQL sur l'IDG ?

The screenshot displays an API documentation interface with the following endpoints:

- GET /** Returns a short description of the API
- GET /doc** Forward to complete documentation site
- GET /spec** Forward to API specification in Swagger
- GET /terms** Returns the terms of use

dataset URI ▾

- GET /{dsid}** the dataset URI returns the dataset metadata

download service ▾

- GET /download/{dsid}** Get Download Service Metadata
- GET /download/{dsid}/describe** Describe Spatial Data Set
- GET /download/{dsid}/get** Get Spatial Data Set
- GET /download/{dsid}/describe/{sot}** Describe Spatial Object Type
- GET /download/{dsid}/get/{sot}** Get Spatial Object

Evolution du WFS 3.0 avec OAI et schema.org

- Cf. https://rawgit.com/opengeospatial/WFS_FES/master/docs/17-069.html#_evolution_from_previous_versions_of_wfs
- Le " Core " du WFS n'impose aucun encodage ou format pour les représenter les ressources.
- spécifie les représentations de ces ressources dans les encodages couramment utilisés pour les données spatiales sur le Web: HTML, GeoJSON, Geography Markup Language (GML).
- Recommandation pour HTML : ajouter les annotations [Schema.org](https://schema.org/)
- recommande de prendre en charge GeoJSON pour l'encodage des données (*réserve le GML pour les cas complexes*)
- utilise des fragments d'OpenAPI 3.0 comme exemples et pour formaliser les exigences. Cependant, il n'est pas nécessaire d'utiliser OpenAPI 3.0.
 - ***Le service DOIT fournir une définition de l'API sous « BaseURL/api »***
 - ***Si le document de définition de l'API utilise OpenAPI 3.0, le document DOIT être conforme à la classe d'exigences OpenAPI Specification 3.0.***

Exemple de mise en œuvre de l'OpenAPI d'un WFS

- **Sur la « BaseURL » du service :**
- **/:**
 - décrit les collections de caractéristiques dans l'ensemble de données
- **/api/conformance:**
 - liste toutes les classes d'exigences spécifiées dans une norme (p. ex., WFS 3.0) à laquelle le serveur est conforme
- **/ {dsid} :**
 - Récupérer l'ensemble des données de la ressource « dataset » (en HTML ou GeoJSON)
- **/ {dsid}/{fid} :**
 - Récupérer l'objet identifié « fid » de la ressource
 - Comme une ressource peut-être multi-classes, le « fid » peut lui-même être structuré en « classe » / « id de l'instance », ou non.