



# SINP

Système d'information  
sur la Nature et le Paysage



## Le schéma GML

### Gabarit physique GML v1.0.0.0 du standard de données national SINP « Occurrences d'habitats » v1.0

Auteur : Rémy Jomier (UMS Patrimoine Naturel)

Testeurs du GML : N/A

1	Introduction.....	2
2	Structure générale.....	2
2.1	Découpage des fichiers.....	2
2.2	Flexibilité de la structure du GML .....	2
3	Contenu du fichier.....	3
3.1	Encodage des fichiers .....	3
3.2	Balises.....	3
3.3	Contenu des balises.....	4
3.4	Vocabulaire contrôlé .....	4
3.5	Annotation.....	4
4	Le Modèle Logique de Données du standard (MLD).....	4
4.1	Implémentation des classes .....	5
4.2	Gestion des attributs facultatifs.....	5
4.3	Gestion des attributs obligatoires conditionnels .....	5
4.4	Changements par rapport au standard .....	<b>Erreur ! Signet non défini.</b>
5	Les modèles du Standard .....	5
5.1.1	Modèle Logique de Données.....	5
5.1.2	Modèle Physique de Données (MPD).....	6
	ANNEXE 1 : Exemples de remplissage pour des attributs de type GM_Object (coordonnées géographiques).....	7

# 1 Introduction

L'objectif de ce document est de présenter l'implémentation du dictionnaire de données en Geography Markup Language (GML), le format choisi par le GT Standard de données (cf CR du 14 octobre 2013). Les fichiers GML, comme les XML, sont créés à partir d'un schéma de référence en XSD, appelé dans ce document « schéma GML ». Les classes et les attributs sont matérialisés par des balises.

L'utilisation du GML 3.2.1 permet de respecter les normes ISO (norme ISO 19136 publiée en 2007) et INSPIRE (format préconisé par INSPIRE).

## 2 Structure générale

### 2.1 Découpage des fichiers

Un fichier GML échangera zéro à plusieurs observations représentant une partie ou la totalité d'un jeu de données, ayant ou non un regroupement parent. Ce découpage pourra être optimisé selon les performances des plateformes dans l'échange de données.

Dans les fichiers d'échange, la balise englobante du jeu de données est « FeatureCollection » suivie de la balise « FeatureMembers ». Elles n'apparaissent pas dans le schéma GML (xsd) mais doivent être ajoutées dans les fichiers GML. La balise de chaque station est « Station »

### 2.2 Flexibilité de la structure du GML

La structure du gabarit GML peut être plus ou moins verrouillée :

- dans la présence de toutes les balises, même si elles sont vides. En effet, dans un GML/XML, un champ vide peut se concrétiser par une balise vide ou par une absence de balise. Ainsi, si a, b, c sont des balises d'un GML et que la balise b est vide, car l'attribut est facultatif par exemple, alors la présence de toutes les balises n'est pas obligatoire : la structure du fichier est pour autant toujours conforme.

Fichier avec les balises a, b, c :	b est facultatif et non renseigné :	ou
<a> </a>	<a>xxx </a>	<a>xxx</a>
<b> </b>	</b>	<c>zzz</c>
<c> </c>	<c>zzz</c>	

- dans l'ordre des balises. L'ordre des balises n'est pas fixé. Par exemple : les balises a, b, c peuvent se présenter en b, a, c ou c, a, b etc ; pour autant, la structure du fichier est toujours conforme.

Fichier avec les balises a, b, c :	Autres possibilités :	
<a> </a>	<a>xxx</a>	<b>yyy</b>

<b> </b>	<b>yyy</b>	<c>zzz</c>
<c> </c>	<c>zzz</c>	<a>xxx</a>
	<b>yyy</b>	<c>zzz</c>
	<a>xxx</a>	<b>yyy</b>
	<c>zzz</c>	<a>xxx</a> etc

Le choix de verrouiller ces aspects de la structure est impactant pour la validation de la conformité des fichiers et leur utilisation. En effet, plus la structure est fixée, et plus la validation du fichier peut se faire avec des parseurs simples et plus la récupération et l'interrogation des données sont facilitées.

Si un concept n'est pas utilisé, il n'a pas à être présent (si par exemple on n'utilise pas les attributs additionnels, aucune des balises des attributs additionnels n'est nécessaire).

L'ordre des balises est fixé. Dans le schéma GML, cela est représenté par la balise <xs:sequence>. Remarque : conformément au concept du GML, les balises objet correspondant aux classes sont laissées libres. La classe Station est flaggée « Root ».

Remarque :

- Les contrôles de conformité des fichiers au gabarit peuvent être faits par les plateformes.

### 3 Contenu du fichier

#### 3.1 Encodage des fichiers

Les fichiers GML seront encodés en UTF-8 sans BOM de manière à permettre aux outils de les lire de façon correcte. Chaque fichier sera en début de fichier précédé des balises suivantes : **<?xml version="1.0" encoding="UTF-8"?>**

#### 3.2 Balises

Les balises correspondent aux classes et aux attributs du dictionnaire de données.

Des choix d'implémentation ont été faits pour implémenter un Modèle Logique de Données, qui lui-même est traduit en Modèle Physique de données (le schéma GML).

Les choix d'implémentation sont présentés au chapitre 4.

Les définitions de chaque élément (classe, attribut, énumération, CodeList) sont ajoutées dans le schéma GML.

Pour l'attribut géographique, le GML rend obligatoire l'échange d'un identifiant unique de l'objet. Cet identifiant n'est pas défini par le GT Standard de données. Il conviendra aux plateformes R/T de diffuser le leur s'il existe ou d'en générer un s'il n'existe pas.

### 3.3 Contenu des balises

Afin d'éviter des problèmes majeurs dans le traitement des fichiers XML, les données de type chaîne de caractères devront  **systématiquement**  être écrites comme suit :

```
<![CDATA[Chaîne de caractères]]>
```

### 3.4 Vocabulaire contrôlé

Le vocabulaire contrôlé représente les valeurs de référence à utiliser pour renseigner un champ. Il peut s'agir :

- d'un référentiel géré au niveau national dans le cadre du SINP (HABREF) ou hors SINP (Communes ou départements par l'INSEE). Elles sont considérées comme des *CodeList* en langage UML.
- d'une liste de valeurs interne comme le vocabulaire contrôlé de techniqueCollecte. Les listes fermées de valeurs sont des énumérations, représentées par des *Enumeration* en langage UML.

Ces différentes CodeLists ou énumérations peuvent faire l'objet de mises à jour plus ou moins importantes. Les référentiels comme TaxRef et Commune sont mis à jour chaque année alors que le vocabulaire contrôlé de techniqueCollecte : « In situ, Télédétection, Lidar... » n'a pas vocation à beaucoup évoluer.

Pour des raisons d'évolutivité potentielle des nomenclatures, toutes les nomenclatures deviennent des CodeList. Cela implique que le XSD n'aura pas à être modifié quand les nomenclatures évolueront. L'inconvénient est que le XSD seul ne permettra pas de valider le contenu des balises qui suivent une CodeList.

### 3.5 Annotation

Une balise de documentation peut être insérée en annotation pour présenter les caractéristiques du schéma. Il est proposé de rajouter une balise d'annotation présentant des informations sur la création et le sujet du schéma.

Ces informations sont en en-tête du fichier de schéma XSD mais elles n'apparaissent pas dans les fichiers de données XML et GML. Ci-dessous la balise proposée :

```
<xs:annotation>
<xs:documentation source = "nom">Occurrences d'habitats</xs:documentation>
<xs:documentation source = "versionDictionnaire">1.0</xs:documentation>
<xs:documentation source = "versionSchemaXSD">1.0</xs:documentation>
<xs:documentation source = "auteurs">SINP</xs:documentation>
<xs:documentation source = "statutDoc">Validé</xs:documentation>
<xs:documentation source = "description">Le but du standard "Occurrences d'habitats" est de
permettre l'échange d'informations sur la biodiversité entre les acteurs du
SINP.</xs:documentation>
</xs:annotation>
```

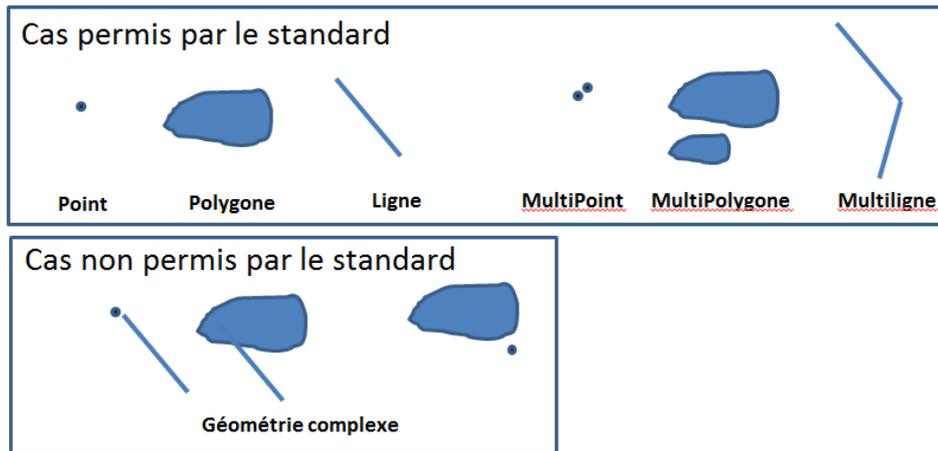
## 4 Le Modèle Logique de Données du standard (MLD)

Le passage du MLD au MPD se fait sans option d'implémentation mais selon les règles des formats techniques, ici selon les règles du standard GML notamment.

## 4.1 Implémentation des classes

Les classes « Station » et « ObjetGeographique » sont liées par une association 1-1 : elles pourraient être fusionnées. Cependant, il est choisi d'implémenter les deux classes en balises afin de séparer les attributs géographiques, des attributs de la station en elle-même.

Remarque : Afin de simplifier le format du fichier GML, la géométrie est mise en GM\_object (toute géométrie permise). Cela fait que le schéma permettra de véhiculer des objets complexes, ce qui normalement n'est pas permis par le standard, mais cela simplifie beaucoup le format. Il sera nécessaire que ce contrôle soit effectué par les plateformes.



## 4.2 Gestion des attributs facultatifs

Les balises facultatives n'ont pas forcément lieu d'être si elles sont vides.

## 4.3 Gestion des attributs obligatoires conditionnels

Afin de gérer au mieux les attributs obligatoires conditionnels dépendant d'autres attributs, les balises englobantes suivantes sont créées :

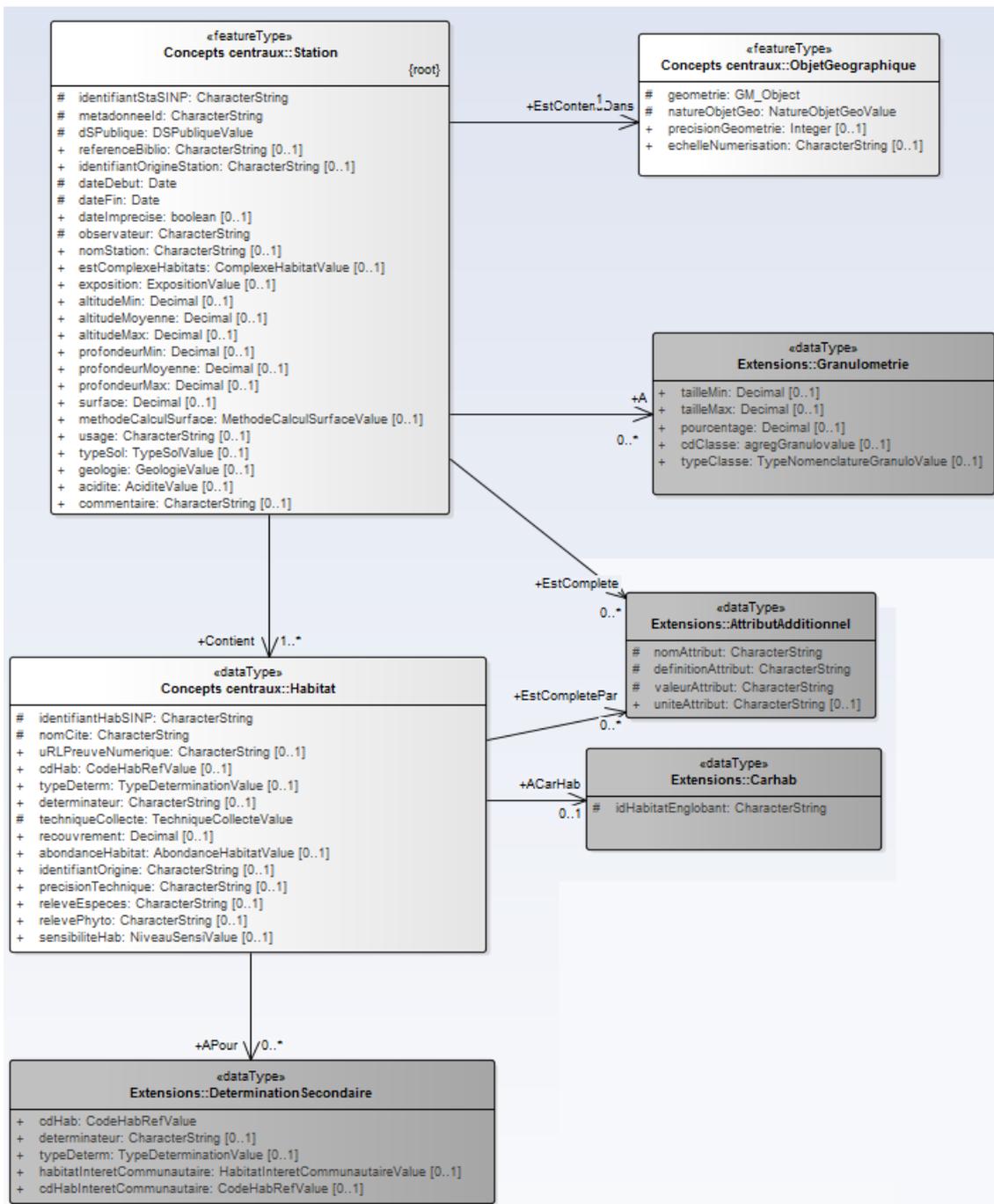
- ObjetGeographiqueType
- Habitat

Il faut se référer aux règles pour savoir quand les attributs sont potentiellement non renseignés. Par exemple, dans le cas où une preuve n'existe pas, il n'est pas nécessaire de fournir la balise. Le schéma permet cette possibilité.

# 5 Les modèles du Standard

## 5.1.1 Modèle Logique de Données

Le MLD produit à partir d'Enterprise Architect est disponible en format propriétaire EA. Notez bien l'absence de trait entre le sujet d'observation et le regroupement, et l'absence de trait récursif sur le concept de regroupement, par rapport au modèle logique du standard. Ces différences sont liées aux choix d'implémentation, mais n'impactent pas les informations présentes ou les relations prévues. Le modèle est présenté dans le document du standard.



### 5.1.2 Modèle Physique de Données (MPD)

Le MPD est disponible en schéma GML (SOH\_1\_0\_extern.xsd).

## ANNEXE 1 : Exemples de remplissage pour des attributs de type GM\_Object (coordonnées géographiques)

Pour un attribut de type GM\_Object (la géométrie de l'objet géographique) avec ses différents types possibles, point, ligne, polygone, multipolygone

```
<!-- Pour un point-->
<gml:Point gml:id="ID_123123123" srsName="EPSG:2154">
  <gml:coordinates cs="," decimal="." ts=" " >376024.0,6707107.0</gml:coordinates>
</gml:Point>

<!-- Pour une ligne-->
<gml:LineString gml:id="ID_3216547" srsName="EPSG:2154">
  <gml:posList>
    45.256 -110.45 46.46 -109.48 43.84 -109.86
  </gml:posList>
</gml:LineString>

<!-- Pour un polygone-->
<gml:Polygon gml:id="ID_22626" srsName="EPSG:2154">
  <gml:exterior>
    <gml:LinearRing>
      <gml:posList>
        45.256 -110.45 46.46 -109.48 43.84 -109.86 45.256 -110.45
      </gml:posList>
    </gml:LinearRing>
  </gml:exterior>
</gml:Polygon>

<!-- Pour un multipolygone -->
<!-- DECLARATION DU MULTI POLYGONE -->
<gml:MultiSurface gml:id="ID_4984312313816" srsName="EPSG:2154">
  <!-- DECLARATION DU PREMIER POLYGONE -->
  <gml:surfaceMember>
    <gml:Polygon gml:id="ID_498431">
      <gml:exterior>
        <gml:LinearRing>
          <gml:posList>45.256 -110.45 46.46 -109.48 43.84 -109.86 45.256 -110.45</gml:posList>
        </gml:LinearRing>
      </gml:exterior>
    </gml:Polygon>
  </gml:surfaceMember>
  <!-- DEUXIEME POLYGONE -->
  <gml:surfaceMember>
    <gml:Polygon gml:id="ID_498431">
      <gml:exterior>
        <gml:LinearRing>
          <gml:posList>45.256 -110.45 46.46 -109.48 43.84 -109.86 45.256 -110.45</gml:posList>
        </gml:LinearRing>
      </gml:exterior>
    </gml:Polygon>
  </gml:surfaceMember>
</gml:MultiSurface>
```