



Le schéma GML

Gabarit physique du standard de données INPG-SINP v1.0.0.1

Rémy Jomier (MNHN/SPN)

1	Introduction.....	2
2	Structure générale.....	2
2.1	Découpage des fichiers.....	2
2.2	Flexibilité de la structure du GML	3
3	Contenu du fichier	4
3.1	Balises	4
3.2	Vocabulaire contrôlé	4
3.3	Annotation.....	5
4	Le Modèle Logique de Données du standard (MLD).....	5
4.1	Implémentation des classes	6
4.2	Gestion des attributs facultatifs	6
4.3	Gestion des attributs obligatoires conditionnels	6
5	Les modèles du Standard	6
5.1.1	Modèle Logique de Données.....	6
5.1.2	Modèle Physique de Données (MPD).....	8



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 pour les occurrences de taxon (cf CR du 14 octobre 2013). Ce format est retenu pour les autres standards, à des fins de cohérence technique (INPG-SINP dans le cas présent). 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).

Ce travail a fait l'objet d'un sous-groupe rassemblant des structures ayant expérimentées la mise en place de fichiers d'échange dans le cadre de système d'information : le MNHN et l'IGN. Le Sandre (Dimitri Meunier) intervenant sur le Système d'Information sur l'Eau (SIE) a été consulté.

Remarque : en avril 2014, le SIE a décidé de temporiser l'utilisation de format GML pour remplacer le format XML à cause de l'impact que cela a sur les outils existants du SIE, développés depuis de nombreuses années et gérés par différents partenaires. Le SINP n'est pas dans la même configuration car les plateformes sont en cours de création et le standard d'échange est un nouvel élément du SINP.

Le numéro de version de ce document est dérivé du numéro de version du standard dont il est dérivé : v1.0.0.x veut dire qu'il est issu de la version 1.0 du standard, v1.2.1.x de la version 1.2.1 d'un standard. Le numéro en lieu et place du « x » est ensuite incrémenté en fonction des modifications faites au XSD lors des phases de test.

2 Structure générale

2.1 Découpage des fichiers

Un fichier GML échangera zéro à plusieurs sites représentant une partie ou la totalité d'un jeu de données. 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 » et la balise de chaque sujet d'observation est « featureMember ». Elles n'apparaissent pas dans le schéma GML (xsd) mais doivent être ajoutées dans les fichiers GML.

Par ailleurs, une balise boundedBy (correspondant à une boîte englobant la totalité des données géographiques présentes dans le fichier) est nécessaire dans chaque fichier GML, de par la nature du GML. Elle sera remplie de la façon qui suit :

```
<gml:boundedBy>
```

```
<gml:Null>Métadonnée</gml:Null>
</gml:boundedBy>
```

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>xxx 	<a>xxx
 		<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>xxx	yyy
 	yyy	<c>zzz</c>
<c> </c>	<c>zzz</c>	<a>xxx
	yyy	<c>zzz</c>
	<a>xxx	yyy
	<c>zzz</c>	<a>xxx 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 cartes marines, aucune des balises de SiteCarteMarine n'est nécessaire).

Nous proposons de fixer l'ordre des balises. 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 Site est flaggée « Root » (c'est elle qui servira d'élément racine parmi les différents autres éléments du standard).

Remarque :

- Les contrôles de conformité des fichiers au gabarit peuvent être faits en les opposant au schéma GML.

3 Contenu du fichier

3.1 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 4.1.

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 chaque objet.

Cet identifiant n'est pas défini par le standard de données INPG-SINP. Il conviendra aux expéditeurs de la donnée de diffuser le leur s'il existe ou d'en générer un s'il n'existe pas. Il aura toujours la forme « ID_ » suivi d'un nombre, et devra être unique au sein du fichier.

3.2 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 ou hors SINP (Stratigraphie, Communes ou départements par l'INSEE). Elles sont considérées comme des *CodeList* en langage UML.
- d'une liste de valeur interne comme le vocabulaire contrôlé de typologie1 ou phenomeneGeol. 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 les départements sont mis à jour chaque année.

Nous proposons de faire référence aux vocabulaires contrôlés dans le fichier GML, ce qui permet potentiellement de vérifier que la valeur transmise dans le champ est une valeur autorisée par l'énumération ou la CodeList. Elle permet en outre d'accéder aux valeurs sans avoir recours au dictionnaire de données PDF.

Concrètement, cela peut se traduire de deux façons :

- Soit la nomenclature est inscrite directement dans le schéma. La nomenclature est directement consultable mais la mise à jour de la nomenclature implique de créer une nouvelle version du schéma GML pour prendre en compte les modifications.
- Soit la nomenclature est appelée via une URL : le fichier fait le lien avec le référentiel disponible dans une forme interrogeable (en XML par exemple).

Avec cette solution, la prise en compte des mises à jour des listes de vocabulaire se fait automatiquement. Cependant, elle est plus lourde à mettre en place et à utiliser :

- la ressource doit être disponible sur la plateforme nationale sous une forme exploitable pour le parseur, ce qui n'est actuellement pas le cas et peut être lourd à mettre en place pour les référentiels complexes et de grande volumétrie comme TAXREF, Commune, EspaceNaturel, Maille et MasseEau.

- il est alors nécessaire d'utiliser un parseur spécifique pour valider la donnée ou de développer du code pour récupérer la liste de valeurs et la comparer au fichier.

Une solution plus légère possible : l'URL peut aussi simplement diriger vers une page informative sur le référentiel à utiliser.

En prenant en compte les particularités des référentiels, et de l'opérationnalité de la mise en œuvre, il est proposé pour cette première version d'intégrer les référentiels à faible volumétrie et faible mise à jour directement dans le gabarit. En attendant que les référentiels plus complexes soient en format interrogeables par URL, les liens ne seront pas intégrés ou dirigeront vers une page informative sur les référentiels à utiliser (à voir avec la plateforme nationale).

Dans un premier temps, les liens URL ne dirigeront pas vers les référentiels sous une forme lisible machine. A terme, il est prévu de rendre les référentiels disponibles dans un format callable par URL de la plateforme nationale afin de faciliter la vérification de conformité des fichiers. Les nomenclatures plus simples disparaîtront ensuite du schéma pour être mises à disposition dans un format accessible pour les machines. Les énumérations disparaîtront au profit des CodeLists.

Remarques :

- Dans le cadre du Système d'Information sur l'Eau (SIE), le Sandre a fait le choix d'inclure dans les XSD les vocabulaires contrôlés à faible volumétrie, les référentiels plus complexes ne sont pas liés aux XSD.

- Les nomenclatures dans le schéma GML seront de plus, disponibles en CSV pour intégration dans les bases de données.

3.3 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 :

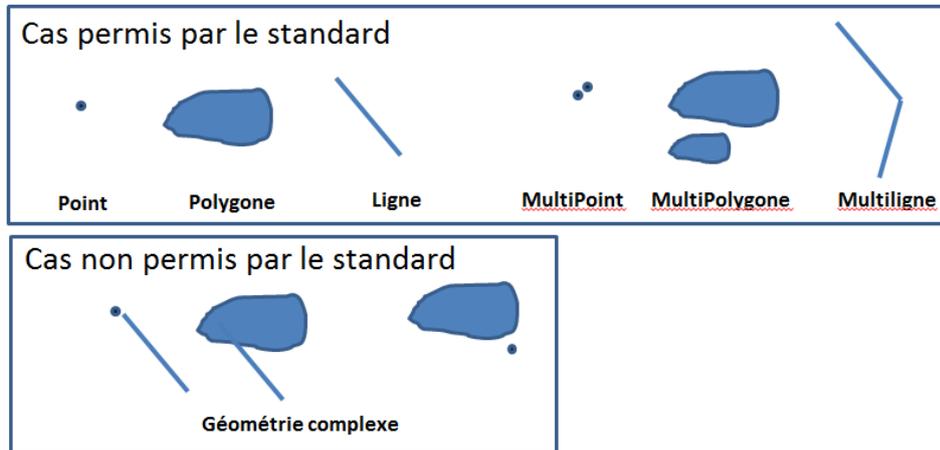
```
<xs:annotation>
<xs:documentation source = « nom »>Standard INPG-SINP</xs:documentation>
<xs:documentation source = « versionDictionnaire »>1</xs:documentation>
<xs:documentation source = «versionSchemaXSD »>1</xs:documentation>
<xs:documentation source = «auteurs»>SINP</xs:documentation>
<xs:documentation source = «statutDoc»>Validé</xs:documentation>
<xs:documentation source = «description »>Ce schéma permet de d'échanger les données de géotopes.</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

Afin de simplifier le format du fichier GML, la géométrie est mise en GM_object (toute géométrie permise), sauf pour les objets points (GM_point). 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 juste nécessaire que ce contrôle soit effectué.



Voir exemples de types autorisés (et comment les remplir) en annexe 1.

4.2 Gestion des attributs facultatifs

Les champs facultatifs sont notés voidable, cf chap 2.2. Les recommandés et obligatoires conditionnels également.

4.3 Gestion des attributs obligatoires conditionnels

Afin de gérer au mieux les attributs obligatoires conditionnels dépendant d'autres attributs, la balise englobante suivante est créée :

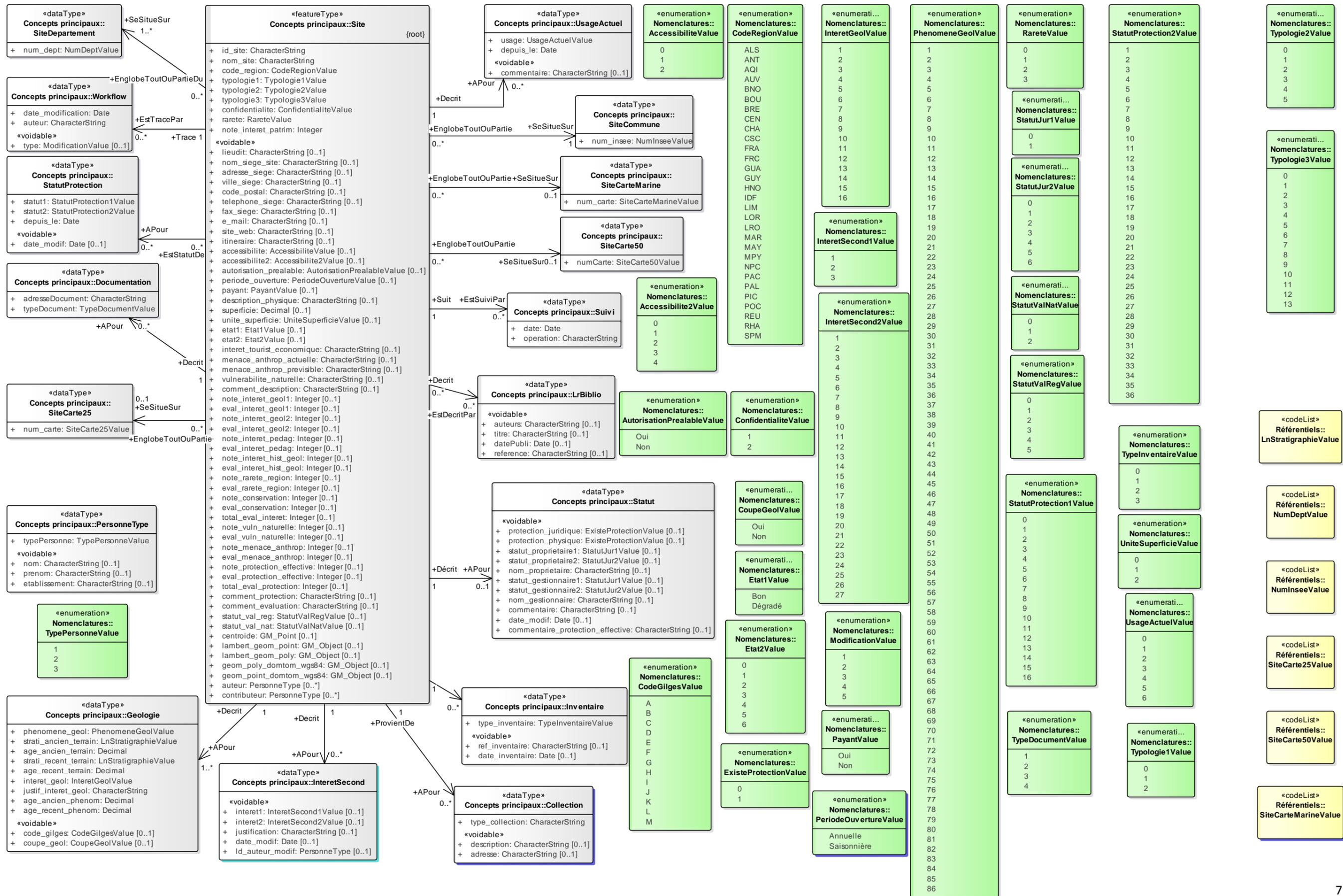
- PersonneType.

Les autres attributs obligatoires conditionnels sont notés « voidable ». Cependant, cela ne veut pas dire qu'ils sont facultatifs, il faut se référer aux règles pour savoir quand ils sont potentiellement non renseignés.

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 (Voir page 7).



5.1.2 Modèle Physique de Données (MPD)

Le MPD est disponible en schéma GML (Standard_INPG_v1.xsd).

ANNEXE 1 : Exemples de remplissage de géométrie

Pour un objet GM_Point :

```
<!-- GM_Point -->
<gml:Point gml:id="ID_123" srsName="EPSG:2154">
  <gml:coordinates cs="," decimal="." ts=" " >376024.0,6707107.0 </gml:coordinates>
</gml:Point>
```

Pour un objet GM_Object, qui peut prendre en compte points, lignes, polygones, multipoints, multilignes, et multipolygones uniquement

```
<!-- GM_Object (accepte point, ligne, multiligne, polygone, multipolygone)-->
<!-- Pour un point-->
<gml:Point gml:id="ID_123" srsName="EPSG:2154">
  <gml:coordinates cs="," decimal="." ts=" " >376024.0,6707107.0 </gml:coordinates>
</gml:Point>

<!-- Pour une ligne-->
<gml:LineString gml:id="ID_321" 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_226" 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 multipoint -->
<gml:MultiPoint gml:id="ID_151" srsName="EPSG:2154">
  <gml:pointMember>
    <gml:Point gml:id="ID_152">
      <gml:coordinates>2.079641,45.001795</gml:coordinates>
    </gml:Point>
  </gml:pointMember>
  <gml:pointMember>
    <gml:Point gml:id="ID_153">
      <gml:coordinates>2.718330,45.541131</gml:coordinates>
    </gml:Point>
  </gml:pointMember>
</gml:MultiPoint>

<!-- Pour une multiligne-->
<gml:MultiCurve gml:id="ID_151" srsName="EPSG:2154">
  <gml:curveMember>
    <gml:LineString gml:id="ID_152" srsName="EPSG:2154">
      <gml:posList>
        45.256 -110.45 46.46 -109.48 43.84 -109.86
      </gml:posList>
    </gml:LineString>
  </gml:curveMember>
  <gml:curveMember>
    <gml:LineString gml:id="ID_153" srsName="EPSG:2154">
      <gml:posList>
        45.256 -110.45 46.46 -109.48 43.84 -109.86
      </gml:posList>
    </gml:LineString>
  </gml:curveMember>
</gml:MultiCurve>
```

```

<!-- Pour un multipolygone -->
<gml:MultiSurface gml:id="ID_127" srsName="EPSG:2154">
  <!--Premier polygone-->
  <gml:surfaceMember>
    <gml:Polygon gml:id="ID128" 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>
  </gml:surfaceMember>
  <!--Deuxième polygone-->
  <gml:surfaceMember>
    <gml:Polygon gml:id="ID129" 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>
  </gml:surfaceMember>
</gml:MultiSurface>

```