

## Fiches techniques

---

### **Fiche technique N° 1 : Numérisation des objets géographiques**

La digitalisation (ou numérisation de points) consiste, en utilisant un logiciel approprié (logiciel SIG), à dessiner et à localiser les **objets géographiques** (points, lignes, polygones) ainsi qu'à renseigner les données qui leur sont associées (**identifiants** et **attributs**).

De plus, un objet géographique possède des **propriétés** qui sont des méthodes de calcul appliquées automatiquement par le logiciel SIG et identiques pour tous les objets de même type :

- Pour un point : coordonnées ;
- Pour une ligne : liste ordonnée de points, longueur ;
- Pour un polygone : liste ordonnée de points (non tous alignés), périmètre, superficie.

(Les points qui définissent les lignes et les polygones sont aussi appelés « nœuds »).

On appelle **classe d'objets** l'ensemble des objets géographiques ayant la même définition géométrique et sémantique (donc les mêmes attributs). Tous les objets d'une même classe doivent impérativement être du même type (même si certains logiciels permettent de transgresser cette règle). Par exemple, si le projet nécessite de représenter des tronçons de cours d'eau, ou des mares, tantôt en linéaire, tantôt en polygonal, en fonction de leur largeur, il faudra utiliser 2 classes d'objets distinctes. En polygonal, la superficie du tronçon sera une propriété de l'objet, il n'y aura donc pas d'attribut « superficie », alors qu'en linéaire, un attribut sera nécessaire pour avoir cette information (attribut « largeur moyenne », par exemple).

Pour chaque classe d'objets, il peut être défini d'autres propriétés qui seront calculées automatiquement par le logiciel grâce à des requêtes (intersection, inclusion, distance par rapport à un objet de référence, ...).

Pour chaque projet il doit être défini une méthode de numérisation appropriée (appelée aussi digitalisation pour la partie géométrique) avec des **règles précises qui se déduisent de la qualité des résultats attendus**. Pour les règles de construction de l'identifiant et des données attributaires, cf. *fiche technique n°4 § 4.4*.

#### **Digitalisation : Le calage des points**

On appelle calage la superposition parfaite d'un point d'un objet avec un point d'un autre objet.

La digitalisation des objets géographiques peut être réalisée sur un fond de plan « raster » (IGN BD Ortho®, IGN Scan25®, ...) ou sur un référentiel « vecteur » (IGN BD Topo®, BD Carthage®, PCI Vecteur, ...).

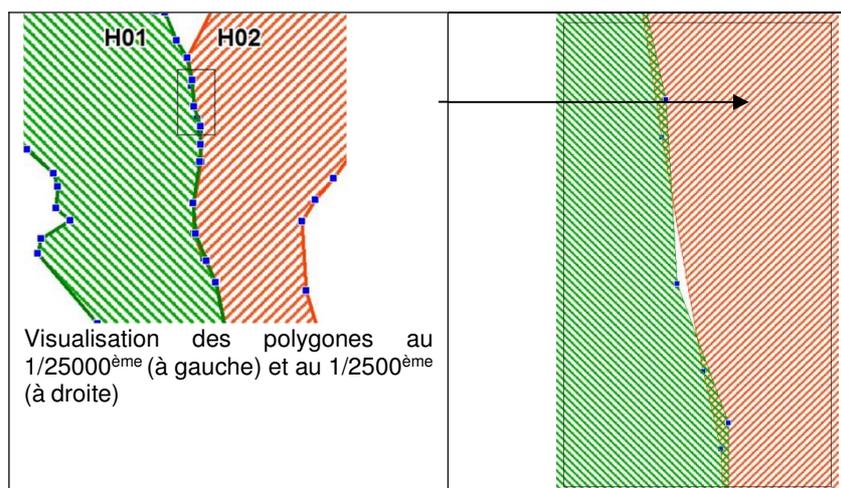
Dans le premier cas, on ne peut pas parler de calage, le fond raster n'étant qu'une photo géo référencée, on ne dispose pas de points de référence.

Dans le second cas, par contre, le référentiel est composé d'objets géographiques délimités par des points sur lesquels on peut positionner les points des objets à digitaliser. Cette méthode est à privilégier chaque fois que cela est possible car les données créées seront cohérentes avec les référentiels existants, et reconnus, sur le territoire d'étude.

En fonction de la nature des objets à créer dans le cadre de l'étude, il est souvent nécessaire de caler des objets d'une même classe entre eux ou avec ceux d'une autre classe.

Par exemple, dans un projet d'inventaire des zones humides, l'entité élémentaire est un objet géographique polygonal qui délimite généralement un habitat humide représenté par un polygone (c'est une liste ordonnée de points appelés nœuds, non tous alignés). Après avoir été identifié sur le terrain et repéré sur un fond de carte approprié, le polygone est dessiné et enregistré dans la classe d'objets « Habitat humide ». Dans ce cas, les relations topologiques à respecter seront l'inclusion de ce polygone « Habitat humide » dans un polygone de la classe « Zone humide effective » et le calage parfait avec les polygones « Habitat humide » voisins ou les limites de la zone humide effective (ou encore les limites d'un autre polygone de type « plan d'eau » ou « cours d'eau »).

### Mise en évidence des imprécisions de la numérisation



Les logiciels SIG sont dotés de fonctions permettant de maximiser la précision de la saisie soit par des constructions automatiques de polygones les uns par rapport aux autres, soit par un système de capture des sommets qui permet la superposition parfaite de chaque nœud. Il est indispensable d'utiliser ces fonctions afin de s'assurer de la fiabilité de la construction car même si les décalages ne sont pas visibles à l'échelle de visualisation (par l'œil de l'utilisateur), elles vont considérablement réduire les possibilités d'exploitation par le logiciel.

Pour plus de détail, consulter les guides pour MapInfo, QGIS ou Arc Gis réalisés par le Forum des Marais Atlantiques en partenariat avec le Conseil Général du Finistère dans le cadre de l'inventaire permanent des zones humides (IPZH29) et téléchargeable à l'adresse : <http://www.zoneshumides29.fr/outils.html>.

### **Digitalisation : établir une méthode de travail**

Les règles de digitalisation doivent être suffisamment précises pour que les objets géographiques créés par différents opérateurs ne présentent pas de grande différence de construction, notamment dans le nombre de points délimitant une ligne ou un polygone. De plus, afin d'optimiser le « poids » des données, il convient de tenir compte de la précision attendue. En effet, quelle que soit l'échelle de digitalisation, la distance entre 2 points consécutifs doit être supérieure ou égale à la valeur de la précision estimée. Ceci implique aussi que certains polygones peuvent être trop petits pour être représentés.

Par exemple, avec une précision attendue de 5 m (pour une échelle d'utilisation inférieure ou égale au 1/5 000<sup>ème</sup>) la distance minimale entre 2 points consécutifs sera de 5 m et le plus petit polygone admissible sera un triangle de 5 m de côté (superficie d'environ 11 m<sup>2</sup>).

Avant un travail important de digitalisation, l'opérateur (ou l'équipe projet) doit donc définir les règles en fonction de la qualité demandée : échelle de digitalisation, espace minimal entre 2 points, choix des points de calage, ou du placement des points par rapport à la représentation de l'objet sur une image raster (par exemple, il y a une infinité de possibilités pour digitaliser une rivière (ligne) sur une photo où l'objet à une largeur souvent plus importante que la précision demandée).

Il est fortement conseillé à l'opérateur de se familiariser avec ces règles de digitalisation sur un secteur test, ce qui permettra de réajuster ou de compléter les règles à suivre et permettra d'estimer le temps total de travail nécessaire à la réalisation du projet de digitalisation. Cela permet aussi de vérifier que la précision obtenue correspond bien à celle qui est attendue.

Le temps passé dans ces réflexions et ces tests n'est donc pas du temps perdu :

- Il permet d'estimer le temps de travail nécessaire,
- Il permet de maîtriser le niveau de qualité
- Il permet d'obtenir des résultats de meilleure qualité (la qualité étant entendue comme la satisfaction d'un besoin, ni moins, ni plus)
- Enfin il permet de gagner du temps : l'erreur la plus fréquente étant de réaliser un travail plus précis que ce qui est nécessaire : ce qui entraîne une perte de temps à la création, mais plus encore à la maintenance et une augmentation significative de la taille des fichiers.

## **Fiche technique N° 2 : Echelle et précision**

L'**échelle** est le rapport qui existe entre une distance sur la carte et celle qu'elle représente dans la réalité. La notion d'échelle n'a de sens que pour un document figé (carte papier principalement).

« *Grande échelle* » signifie que la fraction a un dénominateur petit, par ex. 1/1000 (1 mm → 1 m).

« *Petite échelle* » signifie que la fraction a un dénominateur grand, par ex. 1/100000 (1 mm → 100 m).

Avec les outils informatiques modernes, cette notion a perdu de l'importance : l'échelle peut être modifiée à volonté par l'utilisateur et la distance réelle peut être affichée directement à la demande.

La **précision**, par contre, est une information capitale bien qu'il ne soit pas toujours facile de l'estimer. Sur les cartes papier, la précision est souvent déduite de l'échelle (si l'on estime, par exemple, que la précision du tracé est de l'ordre de 1 mm sur une carte au 1/25000<sup>ème</sup>, on estime la précision à 25 m). Ce nombre correspond à la fourchette d'incertitude du tracé, mesurée sur le terrain : plus le nombre est petit, plus la précision est grande.

Sur un écran d'ordinateur, pour augmenter la lisibilité, il suffit d'augmenter l'échelle, ce qui ne modifie en rien la précision des données affichées. Il faut donc tenir compte de la précision des données acquises sur le terrain ou de celles qui ont servi de référence pour la digitalisation.

Par exemple, un opérateur qui digitalise sur un fond IGN SCAN25<sup>®</sup>, va choisir le 1/15000<sup>ème</sup> pour limiter les risques d'erreur et augmenter son confort de travail : ce qui est vivement recommandé. Dans ce cas, la précision des données produites sera conditionnée par la qualité de son travail, certes, mais surtout par la précision du document source qui est prévu pour une utilisation au 1/25000<sup>ème</sup>, la précision ne pourra en aucun cas être meilleure que 25 m (dans le cas d'une imprécision de l'ordre du millimètre, comme précédemment).

Au contraire, un opérateur qui travaille sur un fond IGN BD ORTHO<sup>®</sup>, devra calculer la précision en fonction des règles et de l'échelle définies pour digitaliser. Si, par exemple il estime à 1 mm l'incertitude du positionnement de ses points à l'écran (ce qui correspond à un travail très soigneux) et qu'il travaille au 1/2000<sup>ème</sup>, il pourra estimer la précision à 2 m. Ce nombre reste à valider par un contrôle sur un secteur test ou un échantillonnage d'objets « faciles » à contrôler.

La précision doit être estimée (et contrôlée) par l'opérateur ou son responsable. **On en déduit une échelle d'utilisation pour les données produites, qui sera mentionnée dans les conditions d'utilisation (métadonnées).** La mention de l'échelle de digitalisation n'est qu'indicative car ce n'est pas le seul critère qui conditionne la précision des données.

Voir aussi fiche technique 4.5 « Note sur les contraintes sémiologiques ».

### Précision absolue / précision relative : cas des acquisitions par GPS :

La précision définie ci-avant peut être qualifiée de relative. En effet, elle définit l'incertitude entre le tracé créé et le fond de plan qui sert de référence à ce tracé. Cependant, elle ne tient pas compte de l'erreur entre le fond de plan ou le référentiel vecteur qui guide le tracé et la réalité du terrain. Cette erreur peut parfois être de plusieurs mètres, ce qui pose un problème dans le cas où l'on veut insérer des acquisitions de points par GPS.

Les mesures GPS sont entachées d'erreurs liées à de nombreux paramètres physiques mais indépendantes d'un fond de plan de référence car elles sont prises directement sur le terrain. Cela peut poser un problème lorsque l'on souhaite intégrer ces relevés à des données digitalisées sur une image ou un référentiel vecteur.

Si, par exemple, des mesures sont effectuées avec un GPS de qualité avec une précision absolue (c'est-à-dire par rapport au terrain) estimée de 1 mètre et qu'elles sont reportées sur une ortho photo dont le décalage avec le terrain varie de - 2 m à + 2 m (précision = 2 m), il en résulte des décalages de 3 m avec le fond de plan pris comme référence.

De plus, la précision d'un GPS est très difficile à estimer car elle ne dépend pas uniquement du récepteur, mais aussi de nombreux facteurs environnementaux (nombre de satellites visibles, obstacles, éruptions solaires ou autre perturbations magnétiques, ...). Elle revêt toujours un aspect statistique qui ne peut être compensé que par un grand nombre de mesures dont on prend la moyenne (éventuellement en 2 temps après élimination des valeurs les plus écartées).

L'utilisation de GPS pour l'acquisition de données nécessite donc un minimum d'expérience et des contraintes fortes pour l'utilisateur. Le choix de cet instrument pour l'acquisition de données répond en général à des besoins spécifiques parfaitement définis.

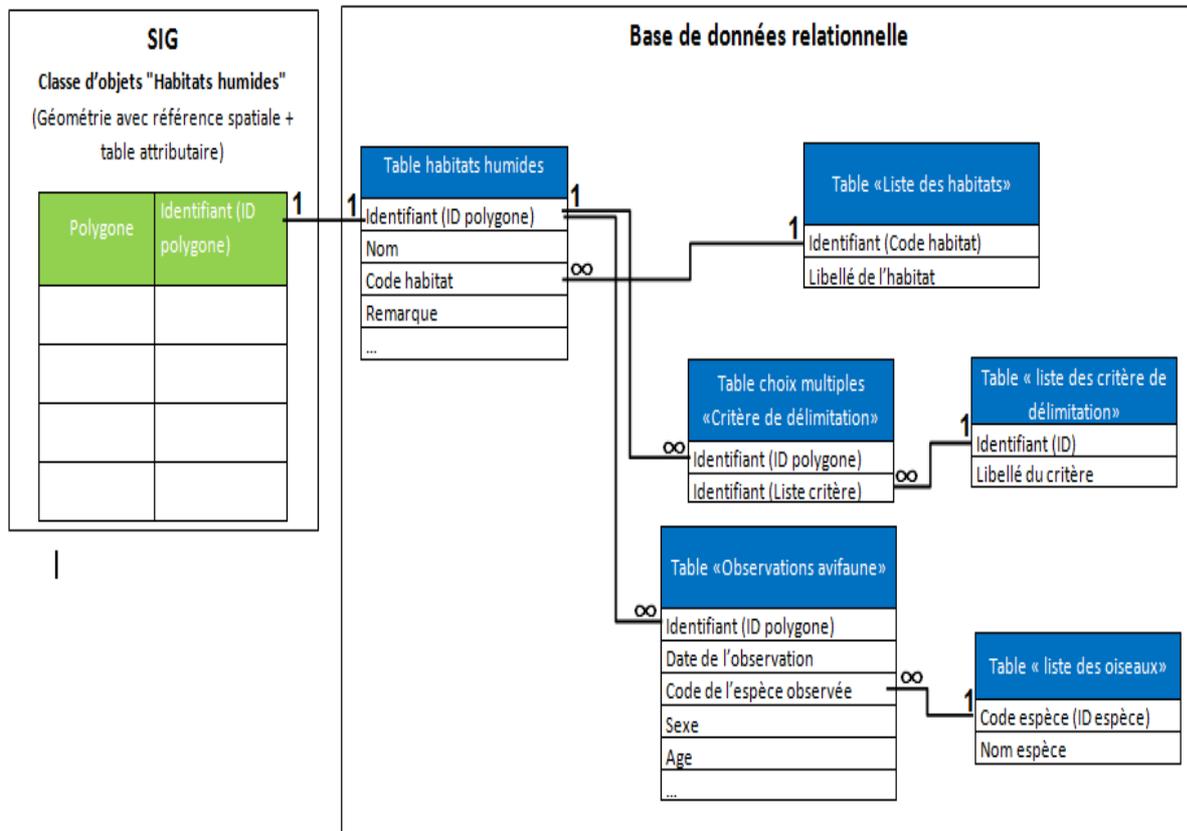
### Fiche technique N° 3 : Données associées aux objets géographiques

Les données attributaires ou attributs (données textuelles ou numériques associées aux objets géographiques) peuvent être stockées dans la base de données géographique elle-même ou dans une base de données « classique » associée de type « tableur » (Excel, ...), ou SGBD (bases de données relationnelles comme Access, SQL Serveur, MySQL, Postgres, ...). En règle générale, on évitera de stocker un grand nombre d'attributs dans la base géographique, et on n'y stockera pas de données de gestion (qui varient rapidement dans le temps ou qui sont liées à un évènement). Un seul attribut est obligatoire dans la base géographique : l'identifiant. C'est l'identifiant qui est un code unique, deux objets distincts ne doivent pas avoir le même identifiant.

#### Le modèle conceptuel (schéma simple de relations entre les données)

Ci-dessous, exemple d'organisation des données « Habitats humides » :

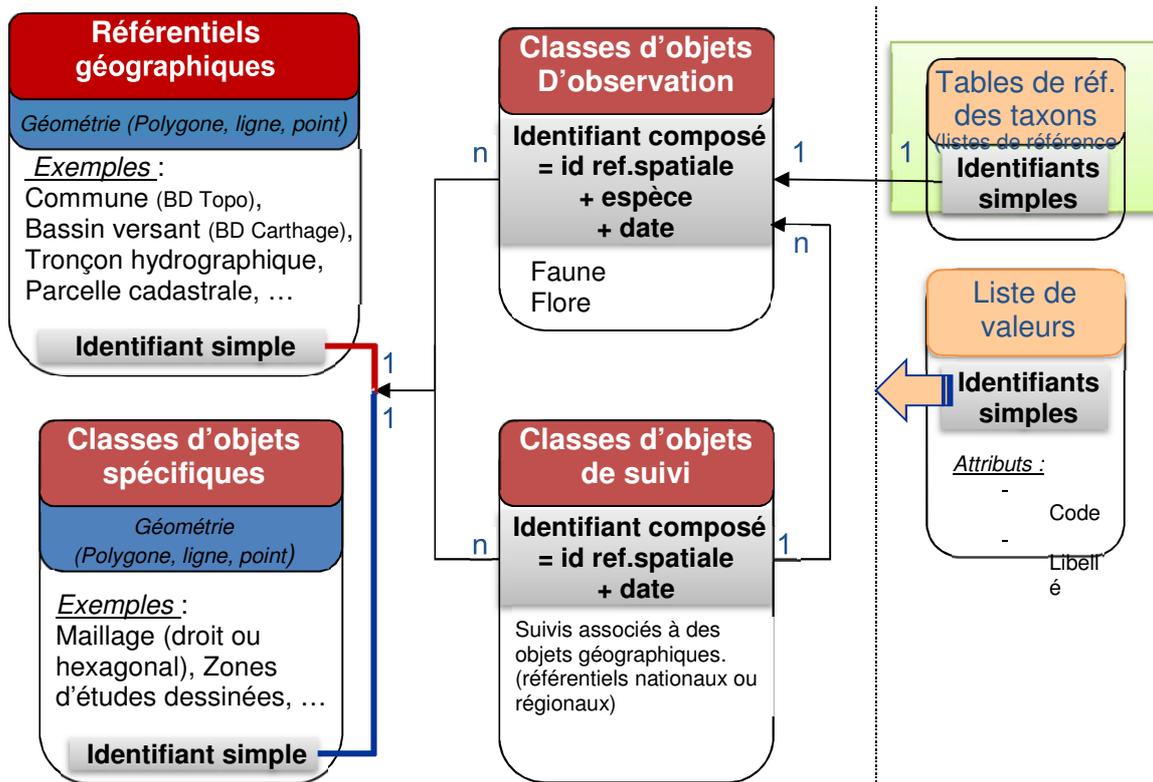
Tous les objets « habitats », de type polygone, sont regroupés dans une même classe d'objets (ou couche d'informations) avec pour unique attribut leur identifiant (ID). Les autres attributs sont stockés dans une base de données (SGBD) structurée en plusieurs tables (ou fichiers) conformément aux règles classiques de modélisation des données.



Comme dans l'exemple ci-dessus, le modèle de données ne doit pas se limiter aux données du SGBD mais doit présenter aussi les relations géométriques des classes d'objets entre elles. De plus, quand elles existent, les relations entre les différentes classes d'objets doivent être schématisées. Cependant, la schématisation décrite ci-dessus n'est pas suffisante pour décrire les caractéristiques des relations géométriques entre les différentes classes d'objets.

**Diagramme des classes pour les données de suivi**

(Exemple du suivi faune et flore utilisé pour la plateforme SIG des Pays de la Loire : GéoPal)



*Note* : les flèches symbolisent des relations logiques (correspondances entre identifiants)

Ce diagramme présente différents types de classes d'objets :

▪ Classes d'objets spécifiques « biodiversité » :

- ✓ **Classes d'objets géographiques** de référence (zone humide, plan d'eau, cours d'eau, ouvrage hydraulique, habitat, ...)

Ces classes d'objets contiennent au minimum les attributs suivants :

- **Identifiant de l'objet** (code unique)
- *Toponyme* (qui peut avoir une valeur nulle)
- *Code typologie* (facultatif mais non nul si utilisé)
- *Dates de création et de mise à jour de la donnée*
- *Identifiant de la fiche de métadonnées associée*
- *Compléments et remarques* (qui peut avoir une valeur nulle)

- ✓ **Classes d'objets de suivi des objets géographiques de référence**, introduisant la dimension temporelle (date d'intervention ou d'état des lieux / diagnostic).

Ces classes d'objets contiennent au minimum les attributs suivants :

- **Identifiant de l'objet géographique de référence** (zone humide, plan d'eau, cours d'eau, ouvrage hydraulique, habitat, ...)
- **Date de référence**
- *Type Intervention* (observation, diagnostic, entretien, restauration, ...)
- *Description* (de l'état, du diagnostic ou de l'action réalisée)
- *Dates de création et de mise à jour de la donnée*
- *Identifiant de la fiche de métadonnées associée*
- *Compléments et remarques* (qui peut avoir une valeur nulle)

- ✓ **Classes d'objets d'observation**, associées à des objets géographiques de référence ou non (objets géographiques de type « point », « ligne » ou « polygone » créés spécifiquement comme support d'observation). Elles concernent les observations de faune ou de flore (référence espèce) et un positionnement temporel (date de l'observation). Les classes d'objets d'observation peuvent être issues d'opérations de suivi lorsque celles-ci traitent aussi le suivi de la faune ou de la flore.
- Classes d'objets généralistes :
  - ✓ **Référentiels géographiques nationaux** (bases de données de type « vecteur » : IGN BD Topo, IGN BD Carthage, IGN BD Parcellaire, PCI Vecteur, ...).
- Classes d'objets non géographiques complémentaires :
  - ✓ **Listes de références nationales ou régionales**
  - ✓ **Listes des valeurs codifiées** utilisées dans les attributs des classes spécifiques « biodiversité » (nommées fichiers ou tables « paramètre », cf. attributs de type « liste de valeurs » ci-après).

### **L'identifiant :**

Dans une classe d'objets (ou couche géographique), chaque objet (ou chaque enregistrement, ou chaque entité) est reconnu par son identifiant, ou clé primaire, qui doit donc être unique. Il n'y a pas de règle absolue pour la codification de cet attribut spécial, indispensable à la cohérence du modèle de données.

On trouve aussi bien des codifications dites « significatives », où la valeur de l'identifiant donne une information sur la nature de l'objet (code géographique, typologie, ...), alors que d'autres ne le sont pas (numérotation chronologique, par exemple).

Les outils modernes de gestion des bases de données permettant d'effectuer des requêtes sur n'importe quel attribut d'une table, il est plus simple et plus facile, aussi bien pour l'opérateur de saisie que pour l'administrateur de la base de données, d'utiliser une codification « non significative » (excepté dans certains cas où une codification plus sophistiquée peut se justifier).

Le principal avantage d'une codification de type « numérotation chronologique » est qu'elle peut se faire de manière automatique avec les outils logiciels standards, ce qui permet un gain de temps non négligeable lors de la saisie des données.

Cependant, il n'est pas souhaitable que le format de l'identifiant soit numérique (possibilité de conflits de formats dans certains cas d'import / export au format texte), il est donc conseillé que le premier caractère soit alphabétique (lettre significative de la classe d'objets par exemple) et que le nombre de caractères soit toujours le même (remplissage par des zéros à gauche du nombre).

### **Exemples simples :**

- Identifiant « Habitat Humide » : H0000101, H0000201, ..., H0000482, ..., H0063015, ...
- Identifiant « Zone Humide » : Z00001, Z00002, ..., Z00004, ..., Z00630, ...

Dans l'exemple ci-dessus, les 5 premiers chiffres de l'identifiant « Habitat Humide » correspondent à la partie numérique de l'identifiant « zone humide » (ou site fonctionnel) à laquelle il appartient (inclusion).

Des formules simples permettent de renseigner automatiquement un attribut de ce type sur toute la table aussi bien avec un logiciel SIG qu'un tableur ou un SGBD.

*Exemple pour les zones humides (ou sites fonctionnels) :*

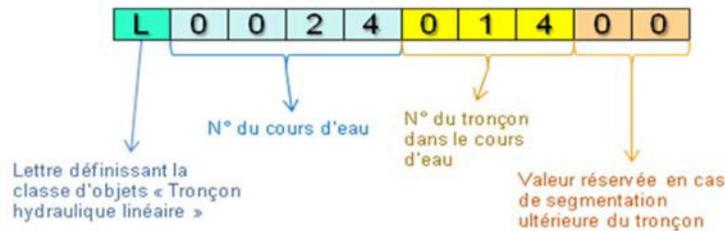
- *Sous MapInfo* : Identifiant = « Z » + Right\$ (« 00000 » + RowID, 5),
- *Sous QGIS* : Identifiant = 'Z' || right('00000' || @row\_number ,5)

*Exemple pour les habitats humides de la zone humide Z00004*

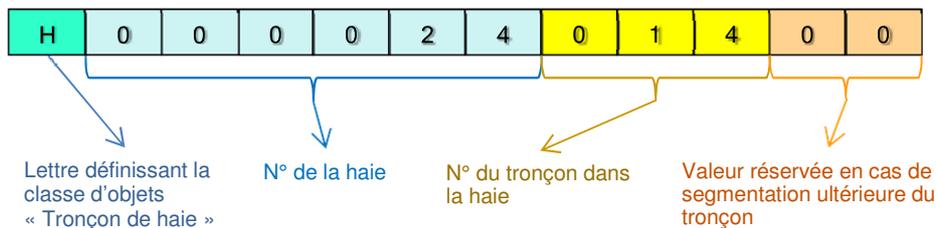
- *Sous MapInfo* : Identifiant = « H00004 » + Right\$ (« 00 » + RowID, 2),
- *Sous QGIS* : Identifiant = 'H00004' || right('00' || @row\_number ,2)

**Exemple de codifications plus complexes :**Pour les cours d'eau :

Dans le cas d'une classe d'objets « tronçon hydraulique linéaire », il peut être souhaité d'avoir une chronologie croissante des tronçons, de l'amont vers l'aval (ou l'inverse), pour chaque cours d'eau.

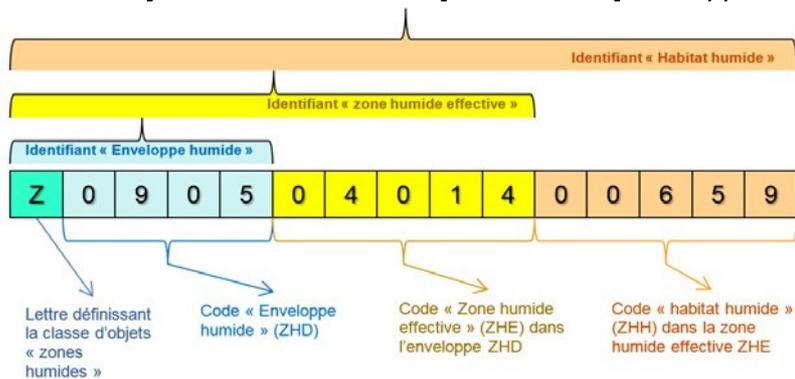
Pour les Haies :

De la même manière, pour la représentation linéaire des haies, il est intéressant d'avoir une codification imbriquée « haie » / « tronçon de haie » :

Pour les milieux et habitats humides :

La codification de l'identifiant peut permettre de rendre compte de l'emboîtement des classes d'objets et surtout de garantir l'unicité du code sur un grand territoire (région, bassin hydrographique, ...) :

[Habitat humide] inclus dans [zone humide effective] inclus dans [enveloppe humide]

Recherche de doublons :

Lorsque les valeurs des identifiants sont saisies manuellement, l'opérateur n'est pas à l'abri d'une erreur qui génère des valeurs identiques pour plusieurs objets différents. Les logiciels SIG ne gèrent pas automatiquement l'unicité des identifiants. La recherche de doublons peut parfois être une manipulation un peu complexe. Une méthode simple consiste à exporter la table attributaire dans un format reconnu par Microsoft Excel (dbf, xls, txt, csv, ...), d'ouvrir le fichier exporté sous Excel, de trier les enregistrements sur la colonne « identifiant » et de comparer automatiquement chaque valeur avec la valeur de la ligne précédente. Cf. copie d'écran ci-après.

	A	B	C	D	E	F	G	H	I	J
1	Nouvelle Colonne	Identifiant	Attribut_1	Attribut_2	Attribut_3	Attribut_4	Attribut_5	Attribut_6	Attribut_7	Attribut_8
2	ok	448000602	Permanent	entre 15 et 50 mètres	Canal, chenal	Non déterminé	Non déterminé	A ciel ouvert		Z5925100
3	ok	448000603	Permanent	entre 15 et 50 mètres	Canal, chenal	Non déterminé	Non déterminé	A ciel ouvert		Z5925101
4	ok	448000604	Permanent	Sans objet	Estuaire	Non déterminé	Non déterminé	A ciel ouvert	Chenal des Gatines	Z5925110
5	ok	448000605	<b>1) - trier les enregistrements par valeurs croissantes de l'identifiant :</b> <i>menu "Données" &gt; "Trier"</i>  <b>2) - insérer une nouvelle colonne :</b> <i>clic droit au dessus de l'entête (colonne de l'identifiant) et "Insersion"</i>  <b>3) - saisir la formule dans la 1<sup>ère</sup> cellule de la nouvelle colonne :</b> $=SI(B2=B1;"DOUBLON";"ok")$  <b>1) - Recopier cette formule dans toutes les cellules de la nouvelle colonne :</b> <i>Selectionner toutes les cellules à partir de la formule, puis touches "Ctrl" et "B"</i>						Chenal des Gatines	Z5925110
6	ok	448000606							Chenal des Gatines	Z5925110
7	ok	448000607							Chenal des Gatines	Z5925110
8	ok	448000608							Chenal des Gatines	Z5925110
9	ok	448000609							Chenal des Gatines	Z5925110
10	ok	448000610							Chenal des Gatines	Z5925110
11	ok	448000611							Chenal des Gatines	Z5925111
12	DOUBLON	448000611								
13	DOUBLON	448000611								
14	DOUBLON	448000611								
15	ok	448000612			Doue du Rouet	Z5925120				
16	ok	448000613			Doue du Rouet	Z5925120				
17	ok	448000614			Chenal de la Prée	Z5925120				
18	ok	448000615			Chenal de la Prée	Z5925120				
19	ok	448000616			Doue du Rouet	Z5925121				
20	ok	448000617			Chenal des Rondées	Z5925121				
21	ok	448000618			Chenal du Martray	Z5925230				
22	ok	448000619			Chenal du Martray	Z5925230				
23	ok	448000620			Chenal des Enfers	Z5925240				
24	ok	448000621	Permanent	Sans objet	Estuaire	Non déterminé	Non déterminé	A ciel ouvert	Chenal des Enfers	Z5925240
25	ok	448000622	Permanent	entre 15 et 50 mètres	Canal, chenal	Non déterminé	Non déterminé	A ciel ouvert	Chenal des Enfers	Z5925240
26	ok	448000623	Permanent	entre 0 à 15 mètres	Canal, chenal	Non déterminé	Non déterminé	A ciel ouvert		Z5925240

### **Les données attributaires (ou attributs) :**

Les attributs peuvent contenir des données de différents types : numérique, alphanumérique (texte libre ou codifié), binaire (vrai/faux, oui/non, 0/1). La saisie de ces données nécessite en général un temps de travail important, il en est de même pour l'indispensable contrôle qui la suit.

**Il est donc impératif, pour les classes d'objets de référence**, c'est-à-dire qui ne sont pas des données « résultat » (issues d'un traitement, pour l'export ou l'édition de documents) **de n'utiliser que des attributs**

- **Véhiculant de l'information utile au projet**, qui pourra être exploitée par interrogation des bases de données (cf. *fiche technique n°5, § 4.5*) ou par des traitements appropriés,
- **Non redondants**, c'est-à-dire qui ne peuvent pas être le résultat d'un traitement effectué sur d'autres attributs ou sur les propriétés géométriques (comme coordonnées géographiques, longueur d'un segment, périmètre ou superficie d'un polygone, relations géométriques avec des objets de cette classe ou d'une autre classe d'objets de référence).

Il est donc important de ne pas sous-estimer le travail de conception des bases de données (cf. *fiches techniques n°2 et n°4*) afin de ne pas oublier d'attribut, mais aussi de ne pas en créer qui ne serviront pas, car ils augmentent inutilement le volume de la base de données et surtout, ils pénalisent le travail de saisie (création et mise à jour).

La règle de non redondance est très importante mais elle n'est pas toujours appliquée, à tort. En effet, non seulement, là aussi, on augmente inutilement le volume de la base de données, mais surtout on introduit l'obligation d'un contrôle supplémentaire (ou d'un traitement de mise à jour).

Exemple : on trouve souvent un attribut superficie dans les classes d'objets de type « polygone » alors que c'est déjà une propriété de l'objet : Si l'on modifie la position d'un point du polygone, la superficie change mais la valeur de l'attribut n'est pas modifiée automatiquement, et donc devient fausse.

**Remarque importante** : il s'agit là, bien sûr, de superficie calculée directement à partir de la géométrie correspondante. Dans certains cas la valeur « superficie » qui est utile provient d'une autre source, elle ne correspond donc pas exactement à la superficie calculée par la propriété de l'objet, il est donc nécessaire d'avoir un attribut « superficie de référence » qui contient cette valeur « officielle ». On rencontre ce cas pour le Plan Cadastral Informatisé (PCI vecteur) : la superficie de référence pour le calcul de l'impôt provient d'une mesure antérieure et ne correspond pas exactement à la valeur donnée par la propriété du polygone « Parcelle cadastrale ».

### Typologie des attributs :

#### **Numérique :**

Ne s'applique qu'aux attributs qui peuvent être utilisés par des formules de calcul arithmétique ou sont le résultat d'une mesure, comme l'altitude d'un point, la largeur moyenne d'un chemin, la profondeur moyenne d'un tronçon de cours d'eau, le nombre d'exploitants d'une unité hydraulique, la section d'ouverture d'un ouvrage hydraulique, la superficie inondable d'une unité hydraulique, le poids maximal que peut supporter un pont, etc.

A certains de ces attributs sont associés des unités de longueur, de surface, de poids qui doivent être précisées dans le dictionnaire de données. Il est conseillé, autant que faire se peut, de choisir une unité adaptée à la précision des objets et qui permette de stocker les valeurs au format « numérique entier ». En effet, l'utilisation du format « numérique flottant » génère des décimales non significatives (nombre de chiffres après la virgule trop élevé) et le format « numérique à nombre de décimales fixe » n'est pas toujours bien interprété lors des imports / exports de données.

Exemples d'attributs de type « numérique entier » dans le cas de classes d'objets ayant une précision plane de 2 à 5 mètres et une précision altimétrique de 20 à 50 centimètres :

- Superficie inondable en mètres carrés,
- Largeur moyenne en mètres,
- Profondeur moyenne en centimètres,
- Section d'un ouvrage en centimètres carrés,
- Poids supporté par un pont en kilogrammes,
- Superficie de l'emprise d'un chemin en mètres carrés,
- Etc.

Pour un attribut de type « numérique », la valeur 0 est ambiguë car c'est souvent la valeur par défaut que va contenir l'attribut lorsqu'il n'est pas renseigné. Il est donc souhaitable que cette valeur ne soit utilisée que pour signaler que l'information n'est pas disponible ou qu'il s'agit d'une erreur de saisie. Il convient donc de préciser dans le dictionnaire de données, en plus de l'unité de mesure, les limites maxi et mini entre lesquelles doit être comprise la valeur saisie pour être considérée comme viable.

*Exemples :*

- Largeur moyenne en mètres, comprise entre 1 et 500
- Profondeur moyenne en centimètres, comprise entre 10 et 2 500
- Etc.

#### **Alphanumérique / texte libre :**

Ce format est utilisé pour des remarques et compléments d'information, la saisie est libre et non obligatoire. Ce type d'attribut n'est pas destiné à une exploitation informatique, il a une valeur purement informative accessible uniquement par la lecture des données associées à un objet géographique.

#### **Alphanumérique / texte codifié ou listes de valeurs prédéfinies :**

Ce format est utilisé dans deux cas :

- La valeur de l'attribut est un identifiant d'une autre classe d'objets de référence (*cf. modèle conceptuel*). L'opérateur de saisie devra s'assurer qu'il existe bien un objet ayant cet identifiant dans la classe d'objets cible.
- La valeur de l'attribut est à prendre dans une liste de valeurs prédéfinie

Dans le second cas, il est vivement conseillé de stocker les différentes valeurs de la liste dans un tableau dit « paramètres », c'est alors le même principe que dans le premier cas.

**L'usage de listes de valeurs prédéfinies est à utiliser autant que possible.** C'est sur ces attributs que pourront être appliquées la plupart des requêtes (interrogation de la base de données) donc plus il y aura d'attributs de ce type, plus la capacité d'exploitation de la base de données sera grande.

Le fait de stocker ces listes de valeurs dans les « paramètres » facilite grandement la saisie (contrôle automatique), l'administration et la maintenance du système. Le contenu de ces « paramètres » est

accessible à l'administrateur du système qui peut créer, modifier, supprimer une valeur de « paramètres » avec effet immédiat sur l'interface de saisie.

### **Binaire :**

Il n'existe que deux valeurs possibles et opposées (vrai/faux, oui/non, 0/1, blanc/noir). Suivant le système de gestion de données, on utilisera pour cela un type « binaire », lorsqu'il existe, ou dans le cas contraire un type « numérique entier court » ou encore un type « alphanumérique » (texte codifié ou liste de valeurs).

Bien que l'information attendue soit d'ordre binaire, on constate que

- il est pertinent d'utiliser une valeur pour dire que l'information n'est pas disponible,
- la valeur 0 est ambiguë car c'est souvent la valeur par défaut que va contenir l'attribut lorsqu'il n'est pas renseigné.

A un format purement binaire, il est donc préférable de choisir une liste de valeurs alphanumériques, qui peuvent être des chiffres ou des lettres qui renvoient à un tableau de « paramètres ».

*Exemple :* classe d'objets des polygones « bassin » (ou surface en eau) dans un marais salé, l'attribut « doux (oui/non) » pourra être de type « texte » sur 1 caractère et prendre les valeurs de la liste suivante :

- « D » correspond à « doux »,
- « S » correspond à « salé » (ou « saumâtre »),
- « Z » correspond à « information non disponible ».

Ou encore

- « 1 » correspond à « doux »,
- « 2 » correspond à « salé » ou « saumâtre »,
- « 9 » correspond à « information non disponible ».

Les termes « salé », « saumâtre » et « doux » devront être définis dans le dictionnaire de données : par rapport au taux en sels dissous, ou par la présence d'espèces associées, ou par la possibilité d'abreuver du bétail, ou ...

Un autre avantage d'utiliser une telle codification est que s'il s'avère nécessaire de faire évoluer l'attribut vers une classification non binaire, l'administrateur pourra le faire sans modifier la structure de la classe d'objets de référence « bassin » ; il suffira juste de modifier les valeurs de la liste dans les « paramètres », par exemple :

- « 1 » correspond à « doux » (moins d'un gramme de sels dissous par litre),
- « 2 » correspond à « légèrement saumâtre » (entre 1 et 10 grammes de sels dissous par litre),
- « 3 » correspond à « saumâtre » (entre 10 et 30 grammes de sels dissous par litre),
- « 4 » correspond à « salé » (au-delà de 30 grammes de sels dissous par litre),
- « 9 » correspond à « information non disponible ».

La codification par des chiffres (bien qu'il s'agisse d'un attribut de type « texte ») est intéressante pour conserver l'ordre de croissance (ici la quantité en sels dissous) lors de l'exploitation de cet attribut (par exemple par une analyse thématique qui va attribuer aux polygones « bassin » une couleur d'autant plus foncée que la quantité de sels dissous sera grande).

### **Les métadonnées :**

De même qu'on utilise un fichier de métadonnées pour décrire et qualifier globalement les données d'une classe d'objets (cf. fiche technique n°6, § 4.6), il est vivement conseillé d'utiliser des attributs pour décrire et qualifier chaque enregistrement. La mise à jour de ces attributs demande un travail supplémentaire à l'opérateur de saisie ou à l'administrateur du système d'information, il est donc important de se limiter strictement aux informations utiles.

## Exemples de métadonnées utiles :

Attribut	Format / Contenu / Valeurs	Pertinence
Origine de la donnée	Liste de valeurs* : Structures à l'origine de la donnée	Uniquement si les données constituant cette table sont de sources multiples
Date de création	Date	Intéressant surtout si les données constituant cette table sont de sources multiples, ou si la saisie a été effectuée sur une longue durée (en plusieurs étapes par exemple)
Date de mise à jour	Date	Toujours utile
Responsable de mise à jour	Liste de valeurs* : initiales des personnes concernées	Toujours utile
Confidentialité	Liste de valeurs* : code confidentialité (par exemple : « 0 » = diffusable, « 1 » = diffusable sous conditions, « 2 » = non diffusable)	Uniquement si la confidentialité n'est pas identique pour tous les enregistrements
Fiabilité**	Liste de valeurs* : code fiabilité (par exemple : « 0 » = incomplet, « 1 » = complet non contrôlé, « 2 » = validation terrain, « 3 » = validé par les acteurs locaux)	Toujours utile
Remarque	Texte libre, codifié ou non codifié, permettant de préciser davantage la qualité des données	Toujours utile, mais remplissage facultatif
Autres attributs de métadonnées	A définir en fonction des résultats attendus	Utilité et faisabilité à argumenter

(\*) - Les listes de valeurs doivent être évolutives : elles sont stockées dans des tables « paramètre » modifiables par l'administrateur.

(\*\*) – Il peut être nécessaire d'avoir plusieurs attributs « fiabilité » : fiabilité géométrique, fiabilité des valeurs attributaires, mais aussi fiabilité en fonction de la source d'information (exemple pour les inventaires des zones humides : « 1 » = délimitation statistique par méthode IBK, « 2 » = délimitation par photo-interprétation, « 3 » = délimitation sur le terrain antérieure à ..., ce qui signifie que cette délimitation a pu évoluer ou que les critères de l'époque n'étaient pas ceux utilisés actuellement, « 4 » = délimitation sur le terrain dans le cadre du projet actuel, etc.).

## **Fiche technique N° 4 : Exploitation des données géographiques.**

### **Interrogation des bases de données (requêtes) :**

Comme indiqué dans la fiche technique n°3, la modélisation des données en amont du projet est fondamentale pour une exploitation pertinente. Les données doivent être structurées pour répondre aux besoins de connaissance et d'aide à la décision (information utile). L'exploitation consiste à utiliser les nombreux outils de requête et d'analyse offerts par les logiciels SIG dans le but de réaliser un document cartographique spécifique ou simplement d'accéder à une partie ciblée de l'information disponible (interrogation des bases de données) ou encore de réaliser des synthèses sous différentes formes.

On oublie souvent que les logiciels SIG sont de puissants outils de calculs, et pas seulement des outils de mise en page pour réaliser des documents cartographiques.

### **Exploitation cartographique des données :**

Par la réalisation d'un document cartographique, il s'agit de délivrer un message clair et non équivoque, adapté à la culture et aux connaissances du public ciblé : éviter à tout prix le risque d'une interprétation erronée par l'utilisateur final (cas très fréquent depuis la démocratisation des SIG, le réalisateur de la carte doit systématiquement se poser la question de ce que comprendra l'utilisateur final, et ce, aux différentes échelles où la donnée est affichable).

*Exemples de défauts courants entraînant une « mauvaise lecture » de la carte :*

- Le domaine d'étude (lieux géographiques où ont été acquises les données) n'est pas, ou pas suffisamment, identifiable sur la carte : l'utilisateur ne va pas différencier les lieux où il n'y a rien (sujet étudié absent) des lieux qui n'ont pas été étudiés (erreur très grave et très fréquente).
- La densité de données exploitées est très variable d'un secteur géographique à un autre mais rien ne l'indique sur la carte (cf. exemple ci-après).
- Le choix, la taille ou la couleur des symboles n'est pas adapté. Certains thèmes ou groupes de données sont trop, ou pas assez, mis en avant par rapport au message que l'auteur veut communiquer. Exemples courants : symbole de type « point » trop gros ou trop petit, symbole de type « ligne » trop épais ou pas assez, bordures de polygone trop épaisses par rapport à la taille du polygone, symboles et couleurs pas assez tranchés, ...
- L'échelle de visualisation est trop grande pour la précision des données : les objets sont décalés par rapport au fond de carte ou à d'autres couches d'information.
- L'échelle de visualisation est trop petite par rapport à la densité des données présentées ou par rapport à la symbologie choisie : l'utilisateur a l'impression que le sujet étudié couvre la quasi-totalité du territoire.

### **Exemple :**

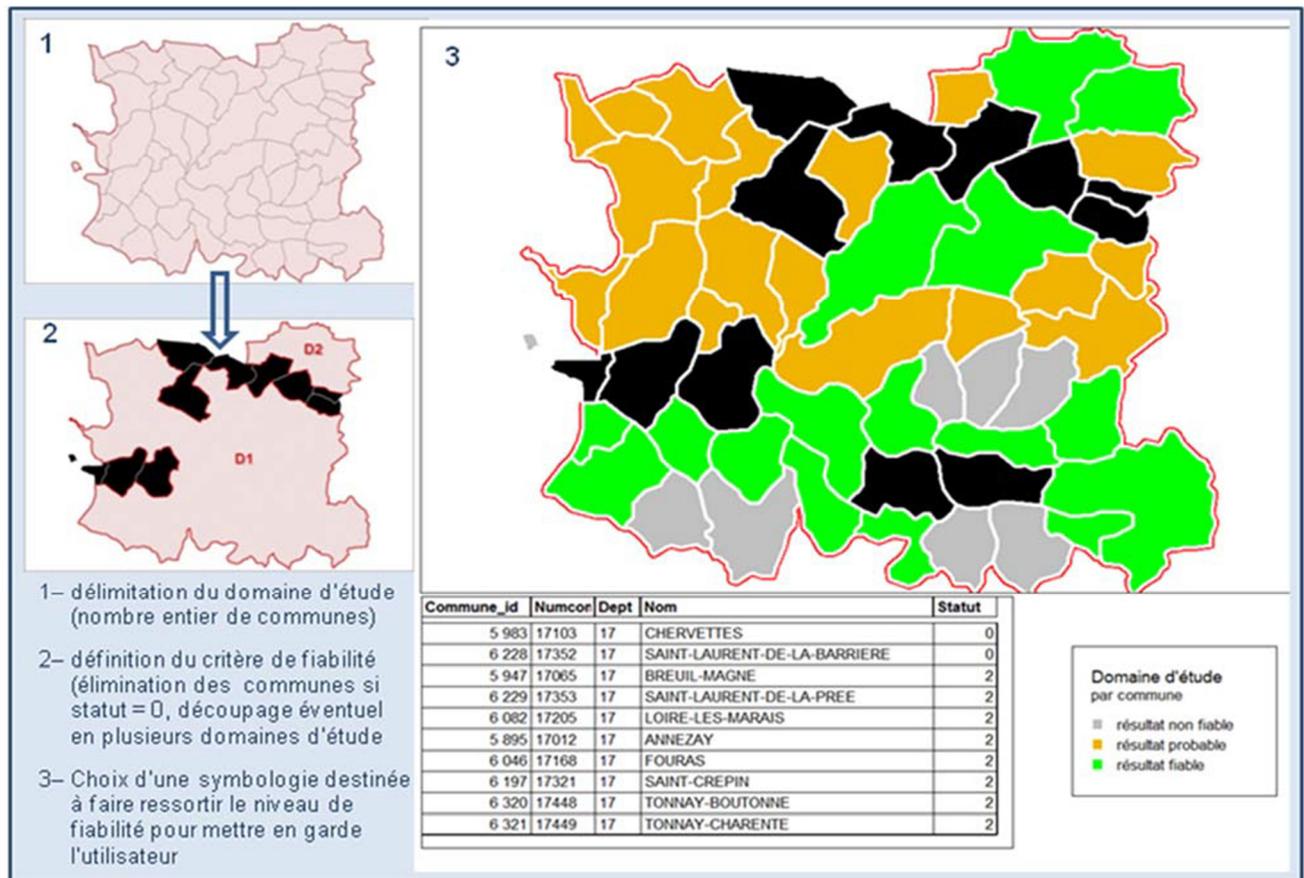
Pour des informations synthétisées par commune (observation d'espèces) et dont la quantité et la qualité peuvent varier sensiblement d'une commune à une autre.

- **Utilisation de messages d'avertissement accompagnant l'affichage des résultats :**

« Attention, le document cartographique ne présente que les espèces qui ont été observées.  
**LE FAIT Q'UNE ESPECE N'AI PAS ETE OBSERVEE  
NE SIGNIFIE PAS QU'ELLE EST ABSENTE SUR LA COMMUNE »**

« Les données mises à disposition reflètent l'état d'avancement des connaissances  
**En aucun cas elles ne sauraient être considérées comme exhaustives. »**

- Utilisation d'une symbologie associée au niveau de fiabilité (unité de synthèse = commune)



### Note sur les contraintes sémiologiques :

Il existe de nombreuses contraintes cartographiques afin de respecter les impératifs de lisibilité de la carte, notamment la surcharge en symboles<sup>1</sup>. Ceci est lié aux limites de notre perception visuelle des éléments cartographiques et à notre capacité d'interprétation des informations fournies par la carte (cf. tableau ci-dessous).

Il existe donc certaines contraintes à la réalisation cartographique, les premières sont propres à l'œil humain. C'est en effet l'outil essentiel à la lecture de documents graphiques, mais il possède des limites dans l'analyse de ces différents éléments<sup>2</sup>:

- Acuité visuelle de discrimination et d'alignement ;
- Seuil de perception ;
- Seuil de séparation ;
- Seuil de différenciation ;
- Seuil dans l'appréciation des couleurs.

Ainsi, sur une carte, les différents éléments doivent donc être :

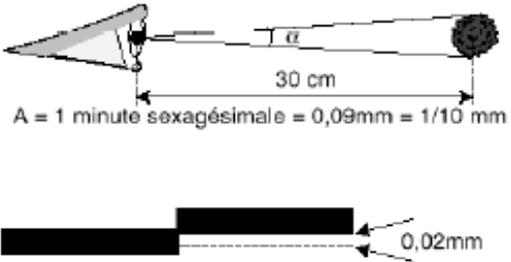
- Perceptibles ;
- Identifiables par leur forme ;
- Suffisamment éloignés des autres éléments ;
- Différenciables les uns des autres.

Lorsque l'échelle est trop petite, la lisibilité se trouve altérée, certaines zones ne sont plus perceptibles ou distinguables. En effet, sur une représentation au 1 / 1 000 000<sup>ème</sup>, Un polygone de 500 m<sup>2</sup> ne sera pas perceptible ni une zone de 1000 mètres de large. A une telle échelle, le fait d'attribuer ou non un contour aux polygones change considérablement le rendu cartographique.

<sup>1</sup> J. Denègre, 2005, Sémiologie et conception cartographique, Hermes science, Paris, page 37.

<sup>2</sup> Sylvain BARD, 2004, Méthode d'évaluation de la qualité de données géographiques généralisées Application aux données urbaines, THESE DE DOCTORAT DE L'UNIVERSITE DE PARIS 6.

En effet, l'épaisseur des contours ne varie pas selon l'échelle, elle reste d'environ 1/2 millimètre, au mieux. Mais à une échelle de 1/1 000 000<sup>ème</sup>, une ligne de 0.5 millimètre de large couvre une bande de 500 mètres. La lecture de la carte s'en retrouve donc complètement faussée car elle laisse penser que les polygones occupent une surface beaucoup plus importante qu'à une échelle plus grande. Cela peut créer des situations d'incompréhension, voire de conflits. A l'inverse, sans contours et toujours pour une échelle trop petite par rapport à la taille des objets, les polygones sont difficilement visibles, voire pas du tout.

CONTRAINTES VISUELLES	REPRÉSENTATION GRAPHIQUE
<p><b>Acuité visuelle de discrémiation</b></p> <p><b>Acuité visuelle d'alignement</b></p>	
<p><b>Seuil de perception</b></p>	<p>Ponctuel    ● 0,2 mm    ■ 0,4 mm    ○ 0,3 mm    □ 0,5 mm                        △ 1 mm    ~ 0,5 mm    ^ 0,6 mm</p> <p>Linéaire    _____ 0,1 mm</p>
<p><b>Seuil de séparation</b></p>	<p>Linéaire    &lt;math&gt;2/10&lt;/math&gt;                        &lt;math&gt;3/10&lt;/math&gt;</p> <p>Ponctuel    &lt;math&gt;2/10&lt;/math&gt;</p>
<p><b>Seuil de différenciation</b></p>	<p>Ponctuel    ●    ●    entre 2 paliers le rapport des surfaces doit être au moins de 2</p> <p>Linéaire    &lt;math&gt;2/10&lt;/math&gt;    Traits rapprochés ecart d'épaisseur 0,1mm</p> <p>              _____    Traits éloignés 0,3 mm minimum</p>

Récapitulatif des dimensions graphiques minimales

Source : Jean Denègre, 2005