



BONNES PRATIQUES CONCERNANT LA PRODUCTION ET L'ACTUALISATION D'UN JEU DE DONNEES GEOGRAPHIQUES PRODUIT AU BENEFICE DE LA DGRNE

à l'intention de producteurs internes de la DGRNE, des personnes qui suivent et élaborent des marchés de production externes, et des producteurs externes

Version du 14/11/2007

1. INTRODUCTION

1.1. Contexte général, pourquoi avoir ces bonnes pratiques ?

La Direction générale des Ressources naturelles et de l'Environnement (DGRNE), au travers de sa Cellule SIG, s'inscrit et s'intègre dans une **politique d'échange de données géographiques** dont l'objectif est de permettre le partage des informations géographiques entre utilisateurs, applications, systèmes et lieux différents. La DGRNE a pour objectif de produire et fournir des données de « qualité »⁽¹⁾.

Actuellement la base de données de la DGRNE (BD_REF SIG DGRNE) contient plus de cent données qui sont consultées/utilisées par de nombreux utilisateurs, dont :

1. des **utilisateurs internes** au Ministère de la Région Wallonne (MRW) qui peuvent se connecter aux données de la BD_REF SIG DGRNE :
 - **directement** au moyen d'un SIG-logiciel (STAR-APIC et ESRI) ;
 - indirectement, au moyen d'un SIG-logiciel (STAR-APIC et ESRI), grâce à des **services** utilisant la BD_REF SIG DGRNE ;
 - indirectement, sans SIG-logiciel, grâce à d'une des dix **applications** Intranet (comme OGEAD ou la cartographie du Patrimoine Naturel) se basant sur la BD_REF SIG DGRNE.
2. des **utilisateurs externes** au MRW, dont les communes, bureaux d'étude, universités et simples citoyens qui peuvent accéder aux données de la BD_REF SIG DGRNE :
 - indirectement au moyen de leur SIG-logiciel, sur une copie des données (de la BD_REF SIG DGRNE) pour laquelle ils ont signés une convention d'utilisation ;
 - indirectement, au moyen de leur SIG-logiciel, par des services utilisant la BD_REF SIG DGRNE ;

¹ La qualité est bien entendu une notion relative : un même produit peut donner lieu à des évaluations de qualité différentes selon des besoins distincts. L'ISO (norme - 8402) définit la qualité comme : « l'ensemble des propriétés ou caractéristiques d'un produit ou service qui lui confère l'aptitude à satisfaire des besoins exprimés ou implicites des utilisateurs ».

- indirectement, sans SIG-logiciel, grâce à l'une des dix applications Internet (comme OGEAD ou la cartographie du Patrimoine Naturel) se basant sur la BD_REF SIG DGRNE.

On remarque donc aisément qu'il est nécessaire d'avoir une **base de donnée** qui **centralise** l'ensemble des données à jour et que chaque client lourd (application) ou léger (SIG-logiciel) doit **se connecter sur la même source unique** de données.

On se rend également compte que la **procédure de mise à jour de données doit être simplifiée au maximum** pour que la Cellule SIG puisse mettre rapidement les données produites en interne ou externe. Pour cela il est nécessaire de respecter quelques règles lors de la création / mise à jour des données. En effet, vu tout ce qui a déjà été énoncé on comprend aisément que **la Cellule SIG ne peut plus se permettre de restructurer et nettoyer chaque jeu de données lors des mises à jour, et cela d'autant plus dans l'urgence.**

1.2. Quelles bonnes pratiques ?

Pour permettre le partage entre différents utilisateurs des jeux de données géographiques de la DGRNE, il est donc nécessaire, même primordial **de définir et décrire ces informations, de structurer et d'encoder les données suivant une méthode préalablement arrêtée.** Cela permet, d'une part, de suivre correctement le cycle de vie d'une donnée et, d'autre part, que celle-ci soit utilisée de manière pertinente.

Pour améliorer et établir la qualité de données géographiques, il est nécessaire de **respecter quelques règles préalablement définies.** Cela suppose avoir défini a priori une référence à laquelle seront confrontées les données utilisables et le contrôle de la qualité consistera à déterminer si le produit répond bien aux spécifications établies.

Il est par exemple tout à fait important de respecter le Modèle Conceptuel de Données (MCD) qui, en toute logique, a dû être défini avant le début de la production de la donnée. Il est également nécessaire de décrire les données et les informations concernant la généalogie (sources de données utilisées et traitements effectués). La précision et l'exactitude de position ou attributaire, l'exhaustivité, la cohérence logique et l'actualité sont autant des composantes qui permettent de décrire, d'appréhender et d'évaluer plus en détail la qualité des données géographiques.

Pour assurer et rencontrer des données de « meilleures qualités » sans trop complexifier et augmenter considérablement la charge de travail, la Cellule SIG préconise de **se focaliser sur la concordance entre le MCD qui aura été défini préalablement** (conjointement entre le producteur et la Cellule SIG), **la validation des données reçues et la rédaction de leurs métadonnées respectives.** La validation des données est effectuée aussi bien par le producteur (tout au long de la production) que par la Cellule SIG (au cours de l'étape de validation et de réception finale des données). Si les données ne correspondent pas au MCD, la donnée pourra être refusée et il sera demandé au producteur (interne ou externe) d'y apporter les corrections nécessaires. C'est donc bien à lui de veiller à l'adéquation de ses données par rapport au modèle défini en collaboration avec la Cellule SIG. Il en est évidemment de même pour les métadonnées.

1.3. Structure du document

Les bonnes pratiques présentées dans ce document sont regroupées en trois parties. La première décrit **le niveau logique** du jeu de données. Elle traite de la définition du jeu de données et du Modèle Conceptuel de Données. Le **niveau physique** est abordé dans

la seconde partie en trois points : les données attributaires, les données spatiales et les critères de qualité qui s'y rapportent directement. La troisième partie reprend quelques considérations sur les **métadonnées** dont il aura été fait mention à plusieurs endroits dans ce document.

2. NIVEAU LOGIQUE DU JEU DE DONNÉES

Cette partie reprend les attentes envers le ou les jeux de données géographiques à réaliser ou à actualiser. Outre une définition précise du jeu de données, la description logique reprend un MCD. Tout producteur d'un jeu de données géographiques doit respecter les exigences décrites et définies en ce point.

2.1. DÉFINITION DU JEU DE DONNÉES

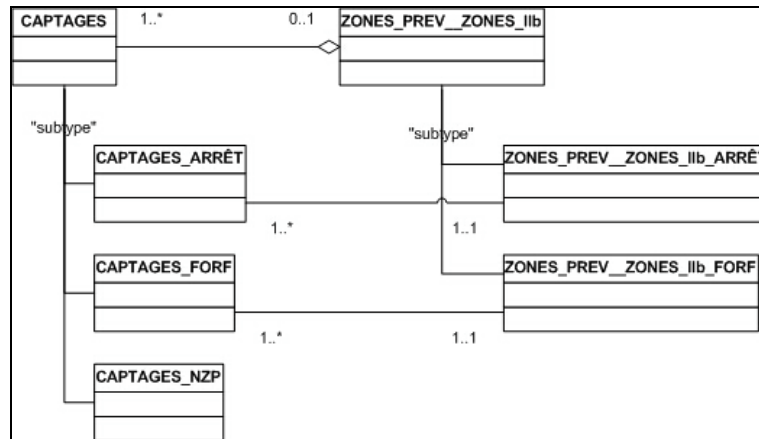
Définir un jeu de données, c'est donner les informations de base le concernant. Ces informations doivent au minimum comprendre :

- le nom explicite du jeu de données et son identifiant (abreg_data) ;
- l'objectif, le résumé associé, les mots-clés et la (ou les) thématique(s) associée(s) ;
- l'information sur la mise à jour (fréquence, personne de contact et information complémentaire - cette information peut concerner le jeu de données en entier ou un attribut ou objet, c'est à préciser), la version de ces données, l'état des données (finalisée, en cours,...), les dates (de création, de mise à jour et éventuellement de publication) ;
- les personnes de contacts (propriétaire, créateur, gestionnaire,...) ;
- la langue utilisée ;
- le type de représentation spatiale, l'étendue spatio-temporelle, le système de référence spatiale utilisé, l'échelle ou la résolution au sol, la gamme d'échelles pour laquelle la donnée peut être utilisée ;
- les copyrights, licences, droits de propriété intellectuelle et autre doivent être mentionnés, cela tant au niveau des données qui ont servi à la création qu'au niveau de la donnée en cours de production ;
- la généalogie doit également être expliquée (pour le jeu de donnée en entier, un attribut ou un objet). Ce point est repris dans la partie « qualité » du niveau de description physique des données.

Lors de la création du jeu de données, il est nécessaire de décrire ces informations au niveau des métadonnées (cf. point 4 relatif aux métadonnées). Au cours de la mise à jour, il s'agit de vérifier l'actualité des métadonnées.

2.2. MODÈLE CONCEPTUEL DE DONNÉES

Une des toutes premières étapes lors de la création de données est de mettre au point le modèle conceptuel de données ou MCD, qui est un ensemble de concepts et de règles permettant de définir la structuration des données de manière précise en vue de son intégration et de sa production à l'aide d'outils informatiques spécifiques (SIG-logiciels notamment). Il permet donc de définir les différentes entités spatiales (classes d'objets), leur type d'implantation (zonale, linéaire, ponctuelle), leurs attributs et les différents liens entre eux. Il doit être complété d'une légende associée, adéquate. Le MCD de chacun des jeux de données se définit en concertation entre le producteur et la Cellule SIG.



Exemple : Diagramme de classe simplifié - Classe d'objets géographiques « Captages » et ses relations directes avec les « zones de prévention éloignée » (arrêtées et forfaitaires)

Une fois le MCD élaboré et approuvé, la production de la donnée peut débuter⁽²⁾. Le MCD doit être scrupuleusement respecté par le producteur au cours de la phase de production. Lors de la réception finale des données par la Cellule SIG (en fin de production), l'adéquation entre le modèle défini et celui délivré sera vérifié. Une donnée non conforme pourra être refusée et des corrections, modifications devront être apportées. Il est donc important de définir au mieux le MCD préalablement et que le producteur vérifie tout au long du processus de création l'adéquation entre la donnée et son modèle. Les appellations des entités (classes d'objets) et champs attributaires doivent autant être respectées que les domaines spécifiques définis (cf. point 3 relatif au niveau physique des données). Il est évident que les informations décrivant le MCD devront se retrouver au niveau des métadonnées (cf. point 4 relatif aux métadonnées).

² Si, pour une donnée déjà existante dans la Base de Données géographiques de Référence de la DGRNE (BD_REF), aucun MCD n'a encore été élaboré, c'est la structure de la donnée telle qu'elle se trouve dans la BD_REF qui fait office de référence. A terme et de manière progressive, les structures de ces données seront revues, complétées et décrites par le producteur et la Cellule SIG et ce, sous l'initiative d'une des parties.

3. NIVEAU PHYSIQUE DU JEU DE DONNÉES

3.1. DONNÉES ATTRIBUTAIRES

Les données attributaires doivent être en adéquation avec ce qui a été spécifié lors de la création du MCD. Les noms, alias, taille, type, nombre de décimales et domaine d'utilisation de chaque attribut est décidé de manière conjointe avant la phase de production. Les données doivent respecter ces exigences tant qu'une modification n'est pas proposée par une des parties et acceptée par toutes.

Si vous produisez une donnée qui existe dans la BD_REF SIG DGRNE mais qui n'a pas encore de MCD, en attendant la formalisation de ce dernier c'est la donnée présente actuellement au niveau de la BD_REF SIG DGRNE qui est considérée comme référence. De plus, si aucun domaine pour les attributs n'a été défini, il est primordial de vérifier leur cohérence et d'attribuer des noms et valeurs explicites.

Lors de la création, de la production ou de la mise à jour d'une donnée, tout producteur doit à la fois vérifier l'adéquation entre le jeu de données à livrer et ces spécifications au niveau attributaire, et élaborer la partie des métadonnées correspondante (cf. point 4 relatif aux métadonnées).

Si de trop nombreuses inadéquations sont identifiées par la Cellule SIG, une donnée pourra être jugée non conforme et des corrections, modifications devront être apportées par le producteur.

3.2. REPRÉSENTATION SPATIALE DU JEU DE DONNÉES - DONNÉES SPATIALES

Les exigences concernant la représentation spatiale ont dû être définies préalablement avec la Cellule SIG⁽³⁾. Il s'agit de vérifier que le géoréférencement a bien été fait correctement et que celui-ci est mentionné, décrit au niveau des métadonnées.

Pour un jeu de données au format vectoriel, les exigences concernent le degré de complexité des relations spatiales entre les objets (définition des contraintes topologiques) et le type de géométries qui définiront les classes d'objets (géométries simples OGC / ISO⁽⁴⁾).

Pour les jeux de données au format matriciel, les exigences portent sur le nombre d'axes spatio-temporels, le type de données (point ou cellule) et les éventuels paramètres de transformation entre les coordonnées images et les coordonnées géographiques ou cartographiques. Il faut également mentionner l'état du géoréférencement (réalisé ou non, voire non-précisé). Si les données sont à géoréférencer et qu'il existe des points de contrôle et des paramètres d'orientation des capteurs, ceux-ci sont à préciser également. De plus, il sera nécessaire de fournir le nombre de pixels sur chacun des axes et la résolution spatiale déterminée.

³ Le géoréférencement des données géographiques se fera toujours selon le système de projection en vigueur pour les cartes topographiques belges (actuellement la projection Lambert belge 72 dont l'ellipsoïde de référence est celui de Hayford - international 1924).

⁴ Un modèle de géométrie simple est composé de points, lignes, polygones avec segments.

3.3. INFORMATIONS SUR LES CRITÈRES DE LA « QUALITÉ »

Traditionnellement, il est demandé une série d'informations sur la généalogie et sur les tests effectués pour contrôler la précision et l'exactitude (de position et sémantique), la cohérence et la complétude.

Il est primordial de connaître la **généalogie** de la donnée, pour déterminer comment l'exploiter et quelles en sont les limites. Cette généalogie reprend :

- les informations sur la source des données⁽⁵⁾ et les systèmes d'acquisition ;
- la description de l'ensemble des processus et traitements⁽⁶⁾ ayant permis de réaliser le jeu de données géographiques avec les personnes de contact correspondantes (créateur) et les dates de traitements ;
- les référentiels (systèmes de coordonnées, systèmes de projection, datum) ;
- les informations réglementaires (informations juridiques ou réglementaires à l'origine de la création de la donnée).

Ce point détaillé est obligatoire, il doit se retrouver dans les métadonnées.

Si le **géoréférencement** est fait et qu'il existe des points de contrôle, il faut les décrire dans les métadonnées, donner les coordonnées des quatre coins du jeu de données matricielles en relation avec les coordonnées correspondantes du système de référence spatial associé et la position des coordonnées au sein du pixel correspondant à sa localisation terrestre.

Ensuite, si certains **tests de qualité** ont été effectués (ou que des informations concernant la qualité existe), c'est ici qu'ils seront décrits, tout comme leurs résultats. Ces tests peuvent concerner la précision et l'exactitude de position⁽⁷⁾, la précision et l'exactitude sémantique; la complétude (ou l'exhaustivité), la cohérence logique et la précision temporelle. Quelques uns de ces concepts sont explicités ci-dessous.

- La **précision de position** concerne :
 - les précisions absolues des objets (concordance des coordonnées mesurées avec les valeurs vraies ou acceptées comme telles) ;
 - la précision de la matrice (concordance des valeurs de position dans la grille avec les valeurs vraies ou acceptées comme telles) ;
 - la précision relative (concordance des positions relatives des objets spécifiés avec les positions relatives respectives vraies ou acceptées comme telles).
- Par **complétude**, nous entendons la description de la présence ou de l'absence d'objets, de leurs attributs et de leurs relations. L'exhaustivité est un paramètre de qualité qui indique le degré de complétude de la donnée par rapport à ce qu'elle est sensé représenter. Le contrôle peut consister à voir si la couverture de la zone est complète, si le nombre d'entités (classes d'objets) modélisées est égal au nombre de classes d'objets / concepts identifiés par la thématique, si les objets modélisés ont le bon nombre d'attributs, si tout ce que contient le MCD est bien porté dans la donnée. L'exhaustivité contrôle aussi bien le manque que le surplus d'information par rapport à ce que la donnée devrait contenir.

⁵ Veillez à utiliser des données sources cohérentes et fiables. Cela concerne également les limites administratives et les fonds de plan. La Cellule SIG peut toujours apporter une aide technique sur ce point.

⁶ Parmi les traitements que vous effectuez avant de donner une version de votre donnée, n'oubliez pas de supprimer tous les doublons dans les enregistrements et ceux qui ne sont pas prévus par le MCD dans les attributs, de même que supprimer les enregistrements et attributs non nécessaires (par exemple ceux de travail). Mettez la donnée en adéquation avec son MCD.

⁷ Même si vous ne faites aucun test de mesure de qualité de positionnement, vous devez vérifier au minimum qu'aucune donnée ne se trouve en dehors de la zone d'étude.

- La **cohérence logique** reprend le degré d'adhésion (de conformité) aux contraintes définies préalablement. Elle décrit la cohérence au niveau des
 - relations topologiques ;
 - domaines de valeurs ;
 - représentations cartographiques des différents objets.

Pour chacune de ces classes d'informations, il faut donner :

- la méthode du test (directe ou indirecte, interne ou externe), sa description claire, son identification ;
- la date pendant laquelle la mesure de qualité a été appliquée ;
- une conformité à ces tests, valeurs obtenues et signification de la valeur des résultats.

4. LES MÉTADONNÉES, QU'EN EST-IL CONCRÈTEMENT?

A de nombreuses reprises ce document fait référence à la complétion de métadonnées. Il est clair que des métadonnées associées aux données doivent être livrées en même temps que chacune des versions de la donnée et que des métadonnées à jour devront toujours accompagner les données.

Tout d'abord rappelons qu'une métadonnée est une donnée qui renseigne sur la nature de certaines autres données et qui permet ainsi leur utilisation pertinente. Une métadonnée géographique décrit le contenu, le type, la nature, la référence spatiale, la qualité, la disponibilité et d'autres caractéristiques des données, services et applications géographiques. On distingue les métadonnées de découverte qui permettent à un utilisateur externe d'avoir un minimum d'information sur la donnée, des métadonnées d'usage qui sont plus complètes.

Etant donné que l'éditeur de métadonnées selon le Profil wallon (normes ISO 19115 & 19139 adoptées par la région wallonne) n'est pas accessible à tous et que ce format donne un document difficilement utilisable, nous demandons simplement l'information brute associée à chaque rubrique du Profil wallon. La Cellule SIG se chargera ensuite de rédiger et de compléter l'information pour qu'elle soit en conformité avec le format « xml » du Profil wallon.

Un document, reprenant les principales rubriques des métadonnées selon le Profil wallon, peut être donné au format "html"⁽⁸⁾. Ce document peut être imprimé et annoté à la main, complété en "html" ou mieux associé à un document Word qui comprend les différentes informations. Il est également possible d'utiliser un éditeur de métadonnée d'ESRI et de donner les informations complémentaires sur un document annexe. Le passage vers le format "xml" du Profil wallon se fera via la Cellule SIG, mais c'est bien le producteur de la donnée qui doit fournir toute l'information nécessaire!

Pour plus de renseignements, n'hésitez pas à nous contacter sur sig.dgrne@mrw.wallonie.be

⁸ Il sera vierge ou contiendra les métadonnées de la précédente version des données.

Annexe : glossaire

Nous reprenons ci-dessous quelques termes courants que nous avons explicités dans la Foire Aux Questions de notre Portail SIG de la DGRNE.

- Qu'est-ce qu'un attribut ?

Un attribut est une caractéristique d'une entité cartographique. Les attributs d'une rivière, par exemple, peuvent comprendre son nom, sa longueur, sa profondeur moyenne, etc. Un SIG bureautique stocke les attributs dans des tables et les lie aux entités cartographiques qu'ils décrivent

- Qu'est-ce qu'une métadonnée ?

Une métadonnée est une donnée qui renseigne sur la nature de certaines autres données et qui permet ainsi leur utilisation pertinente. Une métadonnée géographique décrit le contenu, le type, la nature, la référence spatiale, la qualité, la disponibilité et d'autres caractéristiques des données, services et applications géographiques.

On distingue les métadonnées de découverte qui permettent à un utilisateur externe d'avoir un minimum d'information sur la donnée, des métadonnées d'usage qui sont plus complètes.

- Qu'est-ce que la geodatabase ?

Une geodatabase est un ensemble de jeux de données géographiques à utiliser dans ArcGIS. Il existe différents types de jeux de données géographiques, notamment des classe d'objets, des tables attributaires, des jeux de données raster, des jeux de données réseau, des topologies, etc

- Qu'est-ce que la Base de Données géographiques de Référence de la DGRNE ?

La Base de Données géographiques de Référence de la DGRNE (BD_REF) organise et structure les données géographiques de la DGRNE de façon cohérente, et met ces données à disposition des agents du MET / MRW de manière conviviale (continuité spatiale sur l'ensemble du territoire wallon, rapidité d'accès,...). Vous trouverez plus d'information sur la BD_REF en suivant ce lien.

- Qu'est-ce qu'une donnée géographique, un jeu de données géographiques ?

Une donnée géographique est une représentation des objets ou des phénomènes réels observés à la surface terrestre à un moment donné. La représentation reprend leur position géographique, leur forme et leur description.

Parmi les types de représentation des données géographiques, on distingue généralement les données images (ou raster) et les données vectorielles (points, lignes, polygones).

Un jeu de données géographiques regroupe un ensemble de données géographiques présentant un certain nombre de caractéristiques communes. On regroupe généralement dans un jeu de données l'ensemble des informations se rapportant à une thématique.

- Qu'est-ce qu'une thématique ?

Une thématique est un ensemble d'informations relatives à un même thème (par exemple "localisation", "sol", "nature et paysage", "eau", etc.).

- Qu'est-ce qu'une donnée thématique ?

Une donnée thématique est une donnée géographique qui répond à un thème précis (Réseau hydrographique correspond au thème "eau").