Données nationales / régionales

Document produit par Dominique Laurent et al., IGN, 15/02/19

Annotations de Marc Leobet (LBT), MIG, 21/02/19.

Enjeux des données de référence nationales/régionales

Contexte général : révision du règlement sur le rapportage

« Member States are required to tag selected spatial data sets that are considered to be the **national or regional reference** data set for the given theme or spatial object type(s). »

A priori, la notice explicative demande de taguer ces jeux de données de référence mais ne fournit aucune définition de ce qu’est un jeu de référence.

L’objectif donné dans la note explicative est d’avoir des indicateurs plus pertinents pour le suivi de la mise en œuvre d’INSPIRE (par rapport au nombre de jeux de données utilisé actuellement). La Commission européenne estime que des bases de données coordonnées au niveau national expriment un niveau plus élevé de maturité.

Autres bénéfices possibles pour les utilisateurs :

* Faciliter les recherches sur le GéoPortail européen ; faciliter la constitution de jeux de données pan-Européens par moissonnage des couvertures nationales ou régionales
* Faciliter les recherches sur le GéoCatalogue français.

Impact potentiel pour les producteurs :

* Recevoir le tag « national » donne de la reconnaissance officielle au jeu de données et augmente ses chances d’être découvert et utilisé
* On peut supposer que les contraintes d’interopérabilité seront essentiellement sur ces jeux de données à couverture nationale ou régionale (il n’y a effectivement pas grand intérêt à mettre dans un modèle européen un morcellement de jeux locaux).

Impact potentiel pour la France :

* Bien que cela ne semble pas être le cas dans le règlement révisé sur le suivi d’INSPIRE, le nombre de jeux de données de référence à couverture nationale ou régionale sera probablement utilisé dans le futur comme dénominateur pour vérifier la conformité à certains règlements INSPIRE (ex : interopérabilité) ; le choix judicieux de ces jeux de données peut permettre à la France d’avoir de « bons » indicateurs.

LBT : Le règlement sur l’interopérabilité ne sera pas modifié à court terme en ce sens, et probablement pas avant 2023 La proposition étant d'être mieux au service des usagers, c'est celui-ci qui doit fixer le choix judicieux, charge ensuite à l'Etat de faire en sorte que les indicateurs soient bons.

Proposition : il faut privilégier l’objectif européen, en sélectionnant uniquement les données d’intérêt pour des utilisateurs pan-Européens ou transfrontaliers

LBT : la position ministérielle est qu’une IDG européenne n’a pas d’avenir sans IDG nationale au service, d’abord, des habitants. Il est de ma responsabilité d’en informer les membres de la commission RMO : toute charge supplémentaire doit d’abord apporter un bénéfice aux utilisateurs et entreprises françaises.

Par ailleurs, j’ai appris que la diffusion de données harmonisées via Eurogeographics n’avait pas trouvé d’utilisateur. Il me semble que cela condamne la proposition.

En revanche, que l’IDG française participe à améliorer l’IDG européenne avec un niveau de convergence supplémentaire est une perspective naturelle.

Stratégie de sélection

On peut envisager 3 options :

* 1 : on a pour objectif de faire l’exercice une fois pour toutes courant 2019 ; on essaie d’avoir la liste la plus « exacte » possible ; on se contente de réactualisations en cas de nouveaux produits. Cela implique d’avoir des réflexions suffisamment approfondies pour obtenir une liste qu’on puisse considérer comme définitive (et consensuelle)
* 2 : on va pour objectif une liste minimale en ne sélectionnant que les jeux de données dont on est à on est 100% sûr que ce sont des données de référence ; on prévoit de faire une révision et des enrichissements éventuels d’ici quelques mois selon l’expérience des autres Etats membres, des retours éventuels de la Commission, …. Cela implique d’avoir des critères exigeants de sélection.
* 3 : on a pour objectif une liste « maximale » de jeux de données dont on exclut seulement ceux dont on est sûr à 100% que ce ne sont pas des jeux de référence à couverture nationale ou régionale. On prévoit de faire une révision d’ici quelques mois selon l’expérience des autres Etats membres, des retours éventuels de la Commission, …. et de retirer éventuellement certains jeux de données. Cela implique d’avoir des critères plus permissifs de sélection.

Proposition : adopter la deuxième option => liste a minima de jeux de données de référence.

LBT : Cela me paraît cohérent avec ce que la commission RMO a décidé en décembre : « (...) la meilleure solution serait mixte, avec des règles et une analyse par thème, et surtout un mode itératif. Il est souhaité de traiter d'abord les cas simples, au niveau national comme au niveau régional, et de traiter ensuite les cas compliqués. ». J’espère que nous pourrons régler cela dans l’année.

En revanche, les retours d’expériences d’autres Etats membres ou d’éventuelles remarques de la Commission européenne ne viendront que dans plusieurs années. A mon avis, il ne se passera rien sur ce plan avant 2025 (point à mi-parcours de la prochaine directive révisée).

La première option serait peut-être la plus souhaitable mais risque d’être difficilement faisable. Quant à la troisième option, elle demanderait à certains producteurs d’ajouter puis de retirer un mot-clé, ce qui serait source de mécontentement et de confusion.

Critères de sélection

Il s’agit de déterminer les «données de référence à couverture nationale ou régionale ». Il y aura sûrement quelques interrogations sur l’emprise géographique, mais la principale question portera sur comment définir une **donnée de référence**.

**Analyse de la notice explicative sur les données de référence**

Le premier critère devrait être le fait d’être ainsi désigné par une loi française, comme dit dans le document martyr que j’ai proposé : « Il s’agit du plan cadastral informatisé et de son bâti (DGFiP) et du service public de la donnée[[1]](#footnote-2) : Base adresse nationale, le Code officiel géographique (INSEE), le Registre parcellaire graphique et le Référentiel à grande échelle (IGN) ».

Il est également proposé: « Dans le cadre français, les critères principaux pourraient être les suivants :

- pérennité des services conçus en aval de la donnée de référence : stabilité du

producteur dans le temps, niveau de couverture, fréquence de mise à jour.

- richesse de l’information : niveau de détail, précision, complétude.

- accès à la donnée et à sa métadonnée correctement remplie *via* des services en réseau.

Ces critères sont à débattre. Le dernier point ne pose en général pas de difficulté en

France pour le domaine considéré. »

A mon avis :

* le dernier critère est en grande partie inutile car non discriminant.
* Je pense qu’il faut traiter le niveau de couverture à part ; c’est un critère à lui tout seul.
* Il manque «  l’adhérence » aux modèles INSPIRE ; la richesse d’information doit se mesurer comparativement à ce qui est attendu d’INSPIRE

LBT : En effet, et c'est volontaire pour deux raisons. Sur le plan politique, la position française reste que la mise en œuvre de ces modèles n’est pas réaliste dans l'annexe III. A court terme, seuls les géostandards français sont considérés comme pertinents. Sur le fond, les données françaises sont par construction plus riches que ce qui est défini par INSPIRE. Cela conduit à valoriser d’abord la donnée dans toute sa richesse de production.

Certaines données sont considérées comme de référence dans le contexte français ; cela n’implique pas forcément – à mon avis – qu’elles doivent être aussi de référence dans le contexte INSPIRE qui est sensiblement différent.

LBT : l’utilisatrice qu’est la Commission européenne a tranché pour le niveau européen : les données de référence dans le contexte INSPIRE sont les données de rapportage environnemental. Les Etats membres, dont la France, défendent, quant à eux, que les données des annexes I & II sont les références par construction : c’est ce qui a justifié leurs calendriers de mise en œuvre.

**Propositions :**

1. Cas général

Les critères utilisés par les Pays-Bas pour décider du responsable d’un thème  sont globalement pertinents pour décider des données de référence:

* Critères 1a, 1b, 1c : vérifier que le jeu de données est dans le périmètre d’INSPIRE
* Critères 2 à 10 : voir quel est le jeu de données le plus adapté (classes et attributs INSPIRE, échelle, mise à jour, …)
* Critère 11 : vérifier que le producteur a la capacité et la volonté de rendre ses données conformes à INSPIRE.

En particulier, la deuxième série de critères me paraît essentielle : cela me semble très important de **ne retenir que les jeux de données qui ont une intersection forte avec les modèles INSPIRE**, **ceux que l’on peut rendre interopérables**.

LBT : La question du niveau d’interopérabilité devrait être séparée de la notion de données de référence pour deux raisons :

- La priorité pour la Commission elle-même est d'avoir accès à la donnée 'as-is », l'interopérabilité étant un enjeu littéralement secondaire. J’ai entendu les mêmes remarques des assureurs français et des entreprises croisées ça et là.

- un renforcement des contraintes pourrait, comme dit plus haut, s’appliquer dans le futur aux données de référence. Sur un nombre plus réduit de données, j’estime que nous pourrons à terme aller plus loin que dans le règlement en vigueur. Il me paraît donc aventureux de bâtir un critère sur le modèle actuel.

Un exercice intéressant serait, pour un échantillon de thèmes, d’établir une liste assez large de candidats et de faire une sélection => **avoir des critères de sélection et/ou d’exclusion.**

LBT : c’est ce que j’ai essayé de lancer avec le tableau Je le joins à nouveau, en .xls cette fois. Il prend en compte certaines remarques du présent document.

1. Cas des modèles génériques

Globalement, je trouve ces critères (ceux des Pays-Bas) relativement pertinents pour décider des données de référence. Ils risquent par contre d’être insuffisants, en cas de modèle générique où l’information principale vient de la « variable » à représenter plus que du modèle de données (AM, SD, OF, EF, …).

1. Données transformées – données source

Au cas où des données transformées pour être conformes à INSPIRE, il faudrait taguer uniquement le jeu de données transformées (et pas le jeu de données source). Par exemple,

l’IGN a transformé pour INSPIRE les thèmes AU, AD, GN, HY, TN et BU. Pour ces thèmes, on a donc 2 jeux de données : les données source (BD UNI GE ou BD TOPO) et les données INSPIRE. Ces 2 jeux de données sont à emprise nationale mais seules les données transformées devraient être considérées « de référence » pour INSPIRE.

LBT : si ces données n'ont eu aucun utilisateur, comme je l’ai compris, cela semble prouver que cette hypothèse n'est pas validée.

1. Données à producteur unique ou multiple

Il y a au moins 2 cas à considérer :

* 1. Thème ou sous-thème à producteur unique (ex : Géologie, Météorologie, Océanographie) : on pourrait demander au producteur de proposer une **sélection argumentée** de ses données de référence => contribuer au débat sur les critères de sélection. Dans une deuxième phase, le producteur doit revoir sa liste de façon à respecter – en gros- les critères décidés en communs mais peut avoir un certain degré de liberté
  2. Thème ou sous-thème à producteurs multiples : les décisions sont à prendre en commun.

LBT : cette proposition de catégorisation me semble être une avancée bienvenue. Toutefois, selon moi, les données de référence sont d’une part, celles désignées comme telles par un acte règlementaire, et d’autre part celles qui feront l’objet d’un consensus au sein du CNIG. Que le producteur désigne lui-même parmi ses productions « ses » données de référence me semble être à contre-courant des évolutions en cours.

Un autre enjeu est celui des données à plusieurs producteurs. Le choix des critères de sélection risque d’être plus délicat. A mon avis, il faudrait éviter (le plus possible) de déclarer des doublons en données de référence ; par contre, il peut y avoir plusieurs producteurs pour un même thème, si les données sont disjointes, par exemple, selon le sous-thème (réseaux eau / électricité / …) ou selon la zone géographique (Terre/ mer).

LBT : en effet. Sur une zone donnée, il ne devrait pas y avoir deux données de référence comparables !

A mon avis, si vu au travers des modèles INSPIRE, un jeu de données A n’est qu’un sous-ensemble d’un jeu de données B, alors seul le jeu de données B doit recevoir le tag « national » (ou « régional »).

LBT : d’accord sur le principe. Toutefois, cela pourra dépendre de la richesse du jeu A. Si une donnée régionale A était appauvrie lors de son intégration dans une série de donnée nationale B (notamment pour des raisons de volumétrie), alors les deux resteraient étiquetées.

Commentaires sur le fichier Excel « Couches de référence nationales ou régionales »

Je suis partie de l’hypothèse minimaliste où l’on ne met que les jeux de données que l’on veut rendre interopérable ; pour l’IGN, il s’agit de la grande échelle.

LBT : je suis d’accord que les utilisateurs réclament d’abord de la grande échelle. Ce sont ces données qui sont prioritairement candidates au statut de données de référence.

Commentaires par thème

Thème GN :

* Le COG de l’INSEE est une source pour les noms officiels de données(en particulier pour les données IGN) ; c’est un sous-ensemble des données IGN  mais pas un jeu de référence. Le COG ne contient pas de géométrie qui est obligatoire dans les modèles INSPIRE => on ne pourra pas le rendre conforme à INSPIRE.

LBT : Les modèles INSPIRE ne définissent en aucun cas ce qui est dans INSPIRE ou pas. Au sens de la directive INSPIRE, le COG est une donnée géographique. C’est la donnée de référence française (SPD) et euopéenne (NUTS) pour les noms officiels. Les données IGN dont tu parles appartiennent à un autre thème (« limites administratives »), pour lequel je rejoins ta proposition de les classer en données de référence.

* Par contre, le SHOM a peut -être des données de toponymie non présentes dans les produits IGN
* Pour les produits IGN, faut-il plusieurs produits de référence, certains toponymes n’apparaissant qu’à moyenne voire petite échelle ?

LBT : je ne sais pas. N’y a-t-il pas un fichier global avec le niveau d’emploi selon l’échelle ?

Thème AU :

* A mon avis, les agglomérations (Noise directive) ne relèvent pas du thème AU mais du thème AM LBT : c’est en effet sujet de discussion. Toutefois, la Commission a tranché en ce sens et je ne vois pas l’intérêt d’essayer de la faire changer d’avis.
* COG : idem GN LBT : en fait, à la réflexion, le COG ne devrait pas être dans le thème « Unités administratives ».. Je le supprime de la liste.
* La BD TOPO et les zonages maritimes du SHOM sont effectivement des jeux de référence.

Thèmes AD et CP :

* D’accord pour BAN et PCI

Thème TN :

* Routier : l’IGN a une grande partie des informations attendues par INSPIRE ; a priori, peu d’intérêt de déclarer en supplément les routes principales de la DGPR.
* LBT : comme les deux lignes suivantes, la Commission l’exige dans le cadre de la directive Bruit (voir colonne A du tableau). Cela ne dépend pas de nous. Je pense que la source est une base IGN.
* Ferré : idem (mais à vérifier, l’IGN étant moins riche que sur le routier)
* Aérien : l’IGN n’a que les aéroports et les pistes ; la DGAC a probablement beaucoup plus de données routes aériennes, procédures d’approche, …)  ; peu d’intérêt de déclarer à la fois les aéroports de l’IGN et ceux de la DGPR.
* Eau (fluvial) : les informations de navigabilité étaient dans la BD CARTHAGE ; je crois qu’elles n’ont pas été conservées dans la BD TOPAGE (à confirmer)
* Eau (maritime) : l’IGN a les ports ainsi que les balises mais c’est une toute petite partie du modèle INSPIRE ; il faudrait vérifier si le SHOM ou un autre organisme a les données manquantes.
* Câble : l’IGN a des données en BD TOPO
* Les données IGN du thème TN ont été transformées pour INSPIRE = > ce sont les candidates au tag

Thème HY

* La BD Topage couvre, avec des données à grande échelle et de façon relativement complète les informations demandées par les modèles INSPIRE PhysicalWaters et HydroNetwork ; à mon avis, ça devrait être le seul jeu de données de référence
* Je ne connais pas bien les données de l’AFB mais j’ai l’impression que les données listées correspondent, soit à des sous-ensembles de la BD TOPAGE (ex : eaux de surface), soit à d’autres thèmes INSPIRE (Natura 2000 => PS, masses d’eau DCE => AM, stations de mesure => EF, …)

Thème PS

* Dans le modèle INSPIRE, il y a 6 sous-catégories (Natura2000, Ramsar, biosphere, World Heritage, national monuments, IUCN) que l’on retrouve seulement partiellement dans le fichier Excel.

LBT : à ma connaissance World Heritage et national monuments correspondent à la ligne « historique » ; IUCN est très lié à CDDA etc. A contrôpoler par la DGALN ou l’AFB.

Thème EL

* Cas de Litto 3D en zones côtières ?
* Le trait de côte est cité dans la Directive sous le thème EL ; en pratique, la classe « Shoreline » se trouve dans le modèle du thème SR (Régions marines). La DGALN a lancé la constitution de 2 traits de côte : l’un destiné à servir de référence aux zonages administratifs et règlementaires, l’autre destiné au suivi de l’érosion. A mon avis, les 2 peuvent avoir le tag « national ».

Thème LC :

* Je ne suis pas sûre qu’il faille mettre la couche végétation de la BD TOPO (redondance avec BD Forêt ?) et encore moins la couche OCS de la BD CARTO dont l’actualité est vraiment très mauvaise.
* A court terme, il faut effectivement se contenter de thématiques spécialisées (forêt, zones agricoles) sur tout le territoire et de thématiques complètes (OCS GE) sur des emprises régionales.

Thème OI

* la BD ORTHO à 5 m ne semble pas pertinente ; il faut une approche cohérente pour décider des niveaux de détail. Avoir une règle à appliquer de façon générale à tous les thèmes. Pour les thèmes vecteurs, le tableau ne propose que les données les plus détaillées (BD TOPO), ce qui correspond aussi à la proposition IGN.

Thème GE

* A voir avec le BRGM ; même si le BRGM est le seul producteur, ce serait bien d’avoir des règles générales appliquées par tous et que la France puisse justifier, expliquer.

LBT : en effet : je n connais pas assez bien les bases de données géologiques du BRGM

Thème SU :

* La liste n’est pas forcément restreinte aux IRIS. Dans les données socle de l’ONU, on a considéré les NUTS-LAU, la grille kilométrique (ces 2 produits pouvant être constitués au niveau Européen) et les zonages urbains.

LBT : les NUTS sont d’abords des découpages administratifs (régionaux et plus petits), comme LAU. En France, je ne connais pas qu’on utilise ces classifications.

En revanche, je te rejoins sur les grilles (c’est le thème «  Systèmes de maillage géographique » que nous pourrions ainsi peupler.

* Les zonages urbaines sont en effet une classification INSEE,. Toutefois, si le principe est de placer les agglomérations diverses (Bruit, Eaux résiduaires urbaines...) en « Unités administratives », ne vaudrait-il pas mieux de les y mettre toutes ?

Thème BU

* Conflit fort entre la règlementation française (avec les bâtiments de la DGFIP) reconnus comme référence nationale et l’intérêt européen : les données de l’IGN sont plus riches que celles de la DgFIP, elles sont déjà transformées en INSPIRE. En outre, il n’est pas certain que la DGFiP continue à entretenir les données sur ce thème. **Choisir ou au moins ajouter les données de l’IGN. LBT :** En effet, si la DGFiP se désengageait du bâti, l'IGN serait la référence. Mais, pour le moment, c’est la référence juridique, et INSPIRE est d’ordre juridique.
* Les projets de bâtiments ne sont pas des bons candidats (emprise très partielle et morcelée)

Thème SO :

* Lors des séminaires de revue des thèmes des annexes II et III, j’avais effectivement compris qu’il y avait un standard national et des données produites régionalement. A voir avec l’INRA quelle est la granularité de diffusion : régionale ou nationale.

Thème LU

* Pour l’usage du sol existant, il y a l’OCS GE (production régionale) qui comprend des informations à la fois de couverture et d’usage du sol ainsi que les autres couvertures régionales.
* Pour l’usage du sol planifié, on a un standard national mais la majorité (voire la totalité ?) des données sont à échelon local => pas des candidats pour le tag « national » ou « régional »
* LBT : le SRCE (Schéma régional de continuité écologique) est un document régional de planification établi par le Conseil régional.

Thème HH

* Voir Santé Publique France ?

LBT : ma conviction est que ce thème n’avait rien à faire dans INSPIRE. A classer en « données difficiles à définir » ?

Thème US

* Pour la partie « réseaux », le plus logique serait de considérer les données des gestionnaires de réseaux, surtout dans le cas d’un seul gestionnaire par type de réseaux. Pour le réseau électrique, l’IGN n’a que les lignes aériennes. Les énergéticiens les publient en open data. Une piste si nous arrivons à les faire appliquer INSPIRE ?
* Pour la partie services publics, l’IGN a la plupart d’entre eux, sous forme de PAI/ZAI ; je ne sais pas si les données de l’Education Nationale ou des Conseils Généraux apporteraient quelque chose / modèles INSPIRE.

LBT : par facilité, je penche aussi sur des données IGN, mais qu’en disent les autres membres ?

Thème EF

* Il manque peut-être les données d’observation de la qualité de l’air (qui sont dans le thème « Conditions atmosphériques »).

LBT : en effet. A voir avec l’INERIS. A noter qu’elles serotn déjà étiquetées « données prioritaires ».

Autres thèmes :

* Je ne connais pas assez pour commenter.

1. https://www.data.gouv.fr/fr/reference [↑](#footnote-ref-2)