

Le référentiel toponymique

La CNT « a notamment pour mission (...) d'assurer des actions de spécification, de normalisation, de coordination dans la réalisation et la mise à jour des bases de données toponymiques » (arrêté du 1^{er} août 2000 fixant les missions, l'organisation et le fonctionnement de la CNT). Un modèle conceptuel de données a donc été élaboré en 2006. Il reste à développer pour aboutir à une norme homologuée et pour servir de base au développement du référentiel toponymique nécessaire à la constitution de la composante « adresse » du référentiel à grande échelle (RGE), intégrée à la base de données unifiée (BDUni) de l'IGN.

La présente note vise à présenter ce modèle conceptuel de données, qui repose, afin de répondre aux besoins principalement géographiques et linguistiques des différents utilisateurs de la toponymie :

- sur un système d'identification combinant les deux entrées géographique et linguistique avec un numéro d'ordre non significatif ;
- sur une table principale attribuant à chaque objet géographique recensé une vingtaine d'attributs uniques de nature géographique, linguistique ou autre ;
- sur des tables annexes géographiques et linguistiques, où sont notamment rejetés les autres noms, français ou étrangers, modernes ou anciens, attribués à chaque objet géographique.

Le système d'identification

Le système d'identification repose sur un numéro d'ordre numérique non significatif, associé :

- à un identifiant géocodé alphanumérique, fondé sur les découpages juridiques en vigueur, qui peut donc varier dans le temps, au gré des redécoupages politiques ou administratifs. Il se compose ainsi :
 - pour les objets terrestres :
 - du code ISO alpha 2 (sur deux positions alphabétiques) du pays,
 - pour l'Union européenne, du code de la nomenclature des unités territoriales statistiques (NUTS),
 - pour la France, de la partie du code de l'INSEE relative aux communes,
 - pour les détails géographiques, du code de l'IGN,
 - pour les objets hydrographiques :
 - pour les espaces maritimes, du code de l'Organisation hydrographique internationale (OHI), commençant par une lettre unique suivie de chiffres, tous séparés par des points, complété par un code des mers fermées,

136 bis rue de Grenelle - F-75700 PARIS 07SP

- pour les cours d'eau et pour les lacs :
 - du code de l'espace maritime où il se jette ou du cours d'eau qu'il alimente,
 - d'un code lié à son embouchure dans cet espace maritime, ou à son confluent avec ce cours d'eau, ou à son déversoir ;
- au nom courant, unique nom français recommandé, choisi et orthographié conformément aux recommandations de la CNT (*Paris, mont Blanc, estuaire de la Gironde.*).

Ce système peut donc être interrogé en toute indépendance par l'une ou l'autre entrée, géographique ou toponymique. En outre, les attributs de la table principale permettent à l'interface d'interrogation de retrouver un objet géographique selon d'autres clefs.

La table principale

La table principale comprend un enregistrement par objet géographique recensé. Elle comprend pour chacun les trois types d'identifiants déjà définis (numéro d'ordre, identifiant géocodé et nom français) et une vingtaine d'attributs, géographiques, linguistiques ou autres. Naturellement, un objet géographique peut être enregistré avec une partie seulement de ces attributs, voire sans aucun attribut, mais ceux qui sont listés ci-dessous sont apparus comme utiles pour la plupart des utilisations d'une base de données toponymique.

Les attributs géographiques

La nature de l'entité géographique

Il s'agit de deux champs distincts, permettant chacun d'identifier par un pur code numérique à trois positions :

- une nature géographique, pour les objets naturels ;
- une nature juridique, pour les objets artificiels, incluant les formes juridiques anciennes (provinces de l'Ancien Régime, etc.).

Un objet géographique naturel peut en effet être en même temps un objet juridique, comme une île formant commune ou département.

La géolocalisation

Il s'agit de trois données numériques :

- la longitude ;
- la latitude ;
- la précision de cette géolocalisation.

Elles sont utiles notamment pour résoudre des doublons toponymiques. Il reste pour chaque type d'objet de convenir du point visé pour la géolocalisation, comme le clocher pour un village, le barycentre géographique pour un objet surfacique ou le débouché pour un cours d'eau.

L'altitude

Il s'agit de deux données numériques :

- l'altitude ;
- sa précision.

Le même type de convention reste à définir que pour la géolocalisation, comme le sommet pour un accident orographique, l'altitude moyenne pour un objet surfacique, etc.

La dimension

Il s'agit d'une donnée numérique :

- la superficie pour les seuls objets surfaciques ;
- la longueur pour les objets linéaires (cours d'eau, ligne de côte).

La référence cartographique nationale

Compte tenu de la richesse de la source cadastrale en matière toponymique, il s'agit pour la France du numéro alphanumérique de la ou des sections cadastrales (le pluriel est nécessaire pour les cours d'eau).

Les attributs linguistiques en français

Les attributs linguistiques de base sont tous relatifs au nom français. Ils ont tous une nature alphabétique.

La prononciation

Certains toponymes, même français, nécessitent un guide phonétique (*Papeete*). Deux champs doivent permettre de porter cette précision, au moins quand elle est nécessaire :

- dans le système de l'Association phonétique internationale (API) ;
- dans un système de prononciation « à la française ».

La nature grammaticale

Elle peut consister bien sûr en un nom, dont il faut alors préciser le genre, mais aussi en un adjectif (*Méditerranée*), ou éventuellement d'autres mots encore.

La date d'apparition

Il s'agit de la première apparition dans le cas où l'objet aurait connu une éclipse (notamment pour les objets juridiques).

L'étymologie

Cette information est la plus recherchée du grand public et doit être prévue, même si elle n'est donnée que quand elle est bien attestée et seulement pour les toponymes les plus importants.

La forme officielle

De nombreux toponymes sont des formes courantes de noms officiels plus longs (*la République française*). Bien que ce nom officiel puisse être considéré comme un nom juridique plutôt que comme un toponyme, sa mention reste nécessaire.

La forme cartographique

Dans certains cas, les contraintes cartographiques amènent à mentionner sur les cartes des formes spécifiques des toponymes. Ces formes doivent être recensées.

Exemples : Paris, Mont Blanc, Estuaire de la Gironde.

La syntaxe associée

L'usage d'un toponyme dans un texte peut faire varier sa forme, ou commander l'usage de tel ou tel mot-outil (article, préposition). Deux champs doivent permettre de préciser cet usage :

- en fonction de complément de nom ;
- en fonction de complément de lieu.

L'adjectif associé

L'usage d'un toponyme comprend aussi sa dérivation pour en tirer un adjectif désignant ce qui y est relatif. Le cas échéant, cet adjectif peut être substantivé pour former, avec une majuscule, le nom des habitants (ou *gentilé*), et avec une minuscule, le nom de leur langue.

Les autres attributs

Les codes

Les objets géographiques font l'objet de plusieurs codifications, dont certaines ont un usage suffisamment répandu pour constituer des attributs qu'on ne peut ignorer. C'est notamment le cas :

- du code alphabétique défini par l'ISO pour certains objets géographiques. Pour les pays, le code retenu est le code alpha 2 de la norme ISO 3166 ;
- du code NUTS, qui, bien qu'intégré à l'identifiant géocodé, doit aussi être géré comme attribut, pour l'Union européenne, sous forme alphanumérique ;
- du code national éventuel : pour la France, le code INSEE des départements et des communes ; pour le Québec, les codes des comtés et des provinces...

Le ou les chefs-lieux

Un même lieu peut servir de chef-lieu à plusieurs objets surfaciques (Paris est chef-lieu de département, de région et de pays). Inversement, pour les objets surfaciques, un ou deux chefs-lieux (jamais plus, semble-t-il) peuvent être définis juridiquement par les autorités locales ou politiquement par la communauté internationale. Cependant, la distinction entre les fonctions de ces deux chefs-lieux éventuels n'obéit à aucune règle générale susceptible de donner une homogénéité à une dualité de champs.

Le ou les chefs-lieux sont donc gérés pour chaque objet surfacique, le cas échéant sous forme de répétition de leurs identifiants dans un champ unique.

Les langues administratives actuelles

Notamment pour guider les recherches dans la base annexe linguistique, il est nécessaire de mentionner dans la base principale la ou les langues administratives actuellement en usage au lieu même de l'objet géographique. Ces langues sont mentionnées sous forme alphabétique par le code alpha 3 de la norme ISO 639-2/T (exemple : *fra* pour le français).

Les remarques

Un champ alphabétique libre permet enfin notamment de mentionner des éléments d'information qui ne peuvent être normalisés, par exemple sur les chefs-lieux.

Les tables annexes

Deux types de données ne peuvent être enregistrées que dans des tables annexes : les données linguistiques et les contiguïtés territoriales.

La table linguistique

La table principale comprend un enregistrement par objet géographique ; la table linguistique en comprend un par toponyme. En effet, un même objet géographique peut être dénommé par plusieurs toponymes, selon l'usage (noms régionaux, surnoms...), selon la langue (exonymes, translittération...), selon la date (évolutions linguistiques, changements de nom...). Ces variations constituent un matériau essentiel pour les linguistes. Leur recensement doit s'efforcer de fournir les informations suivantes.

Le toponyme

Le toponyme recensé doit être orthographié aussi précisément que possible, avec tous ses caractères (lettres, traits d'union, espaces...) et ses signes diacritiques (accents, cédilles...), y compris sur les majuscules :

- dans le système d'écriture correspondant au toponyme lui-même (alphabet, syllabaire, idéogrammes) ;
- dans une forme romanisée, c'est-à-dire translittérée en alphabet latin.

Ainsi, par exemple, les noms grec et chinois de la France devraient être enregistrés, avec leurs romanisations.

La langue

Chaque toponyme appartient à une langue, nationale ou régionale (français, provençal, basque...), locale ou étrangère (anglais, espagnol, allemand...), morte ou vivante (grec, latin, ancien français...), dont il faut préciser :

- l'identité, par le code alpha 3 de la norme ISO 639-2/T pour les principales langues vivantes (exemple : *fra* pour le français) ;
- le caractère actuellement administratif ou non, par un codage numérique binaire.

L'usage

L'indication de la langue peut être complétée par celle de l'usage : le toponyme peut être une simple variante d'usage dans une langue donnée, ou revêtir un caractère particulier, diplomatique par exemple pour certaines variantes en usage à l'ONU.

Les sources

Les sources attestant l'usage du toponyme doivent être décrites, avec notamment trois informations :

- le principal document ;
- la date de première attestation de la forme ;
- la date de dernière attestation de la forme.

Naturellement, ces deux dates peuvent coïncider en cas d'hapax.

L'adjectif associé

Enfin, notamment pour les formes étrangères, l'adjectif associé au toponyme dans la langue correspondante est une information utile pour les traducteurs et interprètes.

La table des contiguïtés

Il est apparu utile dans quelques cas de connaître les contiguïtés binaires entre objets surfaciques, terrestres ou hydrographiques. Elles ne peuvent être recensées dans l'une des deux tables précédentes. Une simple table à deux données mettrait donc en relation les identifiants de l'un et de l'autre objet contigu.

*
* *