

*VERS UN SYSTEME D'INTELLIGENCE TERRITORIALE: INDICATEURS ET DONNEES
GEOGRAPHIQUES POUR LES ACTEURS EUROPEENS*

Breuer Christophe

Licencié en Sciences géographiques
Maître en Urbanisme et Aménagement du Territoire
christophe.breuer@ulg.ac.be, +32 4 366 57 47

Devillet Guénaël

Licencié en Sciences géographiques
D.E.S. en Urbanisme et Aménagement du Territoire
G.Devillet@ulg.ac.be, +32 4 366 53 19

Adresse professionnelle

Service d'Étude en Géographie Économique Fondamentale et Appliquée (SEGEFA)
Université de Liège
2, Allée du 6 Août, B-4000, Liège, www.segefa.be

Summary: This article deals with the realization of a territorial information gate for the European actors. It defines the sources of territorial information and the files allowing the cartography of administrative and statistical cuttings. The acquisition cost of information is also putted forward. It establishes the link between the technical realization of an interactive interface of webmapping and the problems of the data and metadata management.

Résumé : Cet article traite de la réalisation d'un portail d'information territoriale pour les acteurs européens. Il définit les sources de l'information territoriale et des fichiers permettant la cartographie des découpages administratifs et statistiques. Le coût d'acquisition des informations est également mis en avant. Il fait le lien vers la réalisation technique d'une interface interactive de webmapping et la problématique de la gestion des données et des métadonnées.

Keywords: Portal of information, territorial information, indicators, data, metadata.

Mots clés : Portal d'information, information territoriale, indicateurs, données, métadonnées.

Vers un système d'intelligence territoriale : indicateurs et données géographiques pour les acteurs européens

1. PRÉAMBULE

Dans le cadre du projet européen CAENTI (coordination action european network of territorial intelligence) du FP6 (Sixième programme cadre européen), des recherches sur les indicateurs territoriaux, leur traitement et leur mode de représentation ont été menées. Après de multiples recherches sur les indicateurs, leur accessibilité et leur définition dans l'Union européenne des vingt-sept (CHEN *et al.*, 2006 ; CAENTI 2008), il était nécessaire de traiter de la problématique de la mise en place d'un portail d'information territorial pour les acteurs, disponible sur internet. Concrètement, il s'agit de permettre la visualisation de données caractérisant l'environnement socio-économique des populations, pour accroître la compréhension des territoires, de leurs dynamiques et de leurs problèmes, et de permettre une mise en perspective des données récoltées par les acteurs auprès des populations à une échelle plus fine.

La représentation cartographique a été privilégiée pour permettre une meilleure lisibilité de l'information, la compréhension des disparités spatiales et la possibilité de comparaison avec d'autres régions. Globalement, bien que la lecture de cartes ne soit pas un acquis pour tous, il s'agit d'un outil de communication puissant.

Il a été décidé que cet outil serait mis à disposition librement sur Internet, ce qui engendre des contraintes techniques et juridiques complexes, en termes de traitements, de représentation et de diffusion de l'information. Néanmoins, c'est probablement grâce à cet atout de libre utilisation que cet outil sera utilisé par les acteurs locaux, souvent confrontés à des restrictions budgétaires et humaines importantes.

Notre travail s'est donc focalisé sur les tâches suivantes :

- l'obtention des données contextuelles et géographiques (découpages administratifs et statistiques) ;
- l'étude et la mise en œuvre des modes de représentation de l'information cartographique en ligne ;
- l'étude et la mise en œuvre des techniques de stockage de l'information et de la meta-information.

Le présent article traitera en premier lieu de la collecte de l'information territoriale et des fichiers géographiques (shapefiles¹) permettant ensemble de créer une cartographie de l'information contextuelle socio-économique et environnementale. Cette partie prendra en considération les contraintes liées au mode de représentation sélectionné et déterminera l'essentiel de la caractérisation de celui-ci. La seconde partie donnera les grandes lignes des représentations cartographiques sélectionnées pour le prototype de l'outil cartographique. La troisième partie traitera de la gestion des données et des métadonnées au sein du système. Ensuite nous concluons sur base des considérations précédentes.

2. COLLECTE DES DONNÉES

La collecte des données est une étape essentielle dans la réalisation du portail d'information territoriale : sans celles-ci, il est évident que le projet n'a pas de substance. Afin de réaliser une cartographie des indicateurs ou des données, deux types de fichiers sont nécessaires : d'une part, les données à représenter, et d'autre part un fichier reprenant le découpage spatial en entités (communes, régions, pays...). C'est à l'intérieur des entités de ce dernier que seront représentées les valeurs des indicateurs.

2.1 Données contextuelles

Les données contextuelles font partie intégrante de l'analyse par la méthode CATALYSE (CAENTI, 2006c). Afin de limiter les recherches de données, des indicateurs ont été sélectionnés sur base du guide européen (CAENTI, 2006a). Au nombre de 15, leur disponibilité auprès de l'Office européen des Statistiques (EUROSTAT) et des fournisseurs nationaux de statistiques a été étudié (CAENTI, 2008). Néanmoins, la création d'un portail d'information ne se résume pas à la simple visualisation, de nouvelles contraintes sont donc apparues.

Contraintes

Différentes contraintes ont rapidement émergé : alors que certains sites de statistiques fournissent les données au format tabulaire, d'autres ne permettent qu'un affichage et une visualisation à l'écran de l'information. Il est cependant indispensable de disposer d'une forme tabulaire pour permettre l'introduction des données dans les logiciels cartographiques. Certains sites internet contraignent de la sorte les utilisateurs à encoder de l'information, ce qui est une opération à la fois longue et coûteuse, généralement hors de portée des acteurs locaux.

Le coût d'acquisition des données reste en effet une problématique importante puisque celui-ci sera le plus souvent répercuté sur les utilisateurs finaux, ce que nous souhaitons éviter. Nous avons donc favorisé l'obtention de données distribuées gratuitement par internet. EUROSTAT rentre dans ce cas de figure, et certains sites nationaux offrent également une large offre de données socio-économiques gratuites. Néanmoins, la majorité des sites statistiques consultés demandent une rétribution pour les données les plus précises, ou les plus spécifiques. Il s'agit essentiellement des données au niveau communal, ou infra communal comme les secteurs statistiques.

La collecte de l'information est donc soumise à une double contrainte, liée au format et aux coûts d'acquisition des données. Concrètement, cela réduit considérablement la disponibilité des données pour leur insertion, alors que les relevés précédents apportaient une vision très prometteuse de l'accessibilité des données au niveau européen (CHEN *et al.*, 2006).

Demeure en outre la problématique de la mise à jour des données, non épuisée aujourd'hui. En effet, la comparaison des relevés de situation sur internet entre les années 2006, 2007 et 2008 a montré une forte variabilité des adresses permettant d'accéder aux sources d'information. Il semble donc complexe, dans un premier temps au moins, d'envisager une mise à jour automatique voire continue sans convention préalable avec les partenaires statistiques.

Résultats

La difficulté de collecte d'information est apparue dès l'utilisation des fichiers issus d'EUROSTAT. Comme mis en avant précédemment (CAENTI, 2008), EUROSTAT collecte majoritairement des informations fournies par des tiers (instituts nationaux...), ce qui engendre des lacunes d'information au sein des bases de données. Dans ce cas, il est parfois nécessaire d'utiliser des années de référence plus anciennes, en prenant en compte la dynamique de constitution de l'Union européenne et donc les disparités d'information pour les pays intégrés plus récemment. En outre, certains indicateurs de l'Office européen de la Statistique sont nommés de manière impropre, ou en créant la confusion : des indicateurs que l'on pense acquis peuvent se révéler peu pertinents suivant la méthode de calcul.

Au niveau des fournisseurs nationaux, c'est la multitude qui accroît la difficulté d'obtention des données. L'acquisition peut en outre s'avérer coûteuse en deçà de certains niveaux spatiaux très fins (par exemple pour les communes en France et les secteurs statistiques en Belgique). Les indicateurs majeurs (population totale, taux de chômage, revenu par personne...) sont souvent gratuits, même pour des niveaux spatiaux restreints. Dès que les indicateurs deviennent plus complexes, ils sont distribués moyennant des sommes pouvant être importantes. Il faut en effet prendre en considération le nombre de pays et le nombre d'entités qui démultiplient les coûts d'acquisition, sans assurer pour autant des mises à jour régulières.

Au sein de quatre pays pilotes, nous avons pu collecter plusieurs indicateurs simples (population totale, densité de population, taux de chômage, part des étrangers, revenu des ménages...) de manière continue, depuis les NUTS 1 jusqu'aux LAU2/NUTS5. Des indicateurs plus complexes présentent de manière systématique des lacunes et discontinuités, soit verticales (des niveaux spatiaux les plus bas vers les plus hauts), soit horizontales (entre entités d'un même niveau administratif).

2.2 Données géographiques

Les données géographiques constituent la base de la représentation spatiale des indicateurs. Essentiellement échangés et distribués sous forme de fichier vecteur « shapefiles », nous avons tenté de conserver l'homogénéité de format pour l'ensemble des données collectées. Les distributeurs nationaux et internationaux de données géographiques proposent le plus souvent différents formats alternatifs, bien que les shapefiles soient dominants.

Contraintes

La cartographie de différents pays dans une même interface et de manière continue peut être confrontée à différents problèmes. Ils résultent essentiellement des méthodes et caractéristiques des différents produits cartographiques. En effet, la création des shapefiles en général, et de couches d'information sur les découpages administratifs particulièrement, sous-tend un certain nombre de variables pouvant différer d'un créateur de données à l'autre. Parmi ceux-ci, nous retiendrons les niveaux de référence des digitalisations, l'échelle de base de la cartographie, l'imprécision pouvant atteindre plusieurs kilomètres ou encore les projections employées.

Nombre de ces contraintes peuvent trouver une réponse par l'exploitation des possibilités des logiciels de cartographie, qui disposent de fonction de correction ou de modification des projections. Cependant, les problèmes d'imprécision ne peuvent trouver de solution aisée. Le risque de l'assemblage de shapefiles d'origines différentes pour les pays européens est donc de créer des incohérences et des discontinuités dans la cartographie.

D'autre part, les contraintes liées au coût des données demeurent également pour les informations géographiques.

Fournisseurs nationaux

Chaque pays dispose de son fournisseur de données géographiques officiel, même si celui-ci peut sous-traiter tant la revente des informations que leur création. Dans certaines nations, un seul institut s'occupe de la digitalisation des limites administratives et de leur distribution, mais ce cas de figure reste rare.

Les données ont le plus souvent un coût, qui varie selon le pays, la zone que l'on souhaite acquérir et le niveau administratif des découpages. Les pays qui donnent gratuitement les données géographiques sont minoritaires dans l'espace européen. La plupart d'entre eux demandent un prix croissant en fonction de la précision des données et dans certains cas le format (certaines données sont gratuites en format image mais payantes en format shapefile).

D'autre part, tous les pays ne distribuent pas les fichiers géographiques. Certains instituts nationaux contactés personnellement ont confirmé ne pas mettre à la disposition de tiers les données des découpages administratifs, voire ne pas en disposer dans des formats utilisables par les logiciels de cartographie. Ces réponses sont certainement discutables, mais illustrent la difficulté d'obtention de telles informations.

Notre travail de collecte d'information a donc été rendu plus complexe par la multitude de fournisseurs, les formats de données non compatibles avec les logiciels de cartographie les plus répandus, et les coûts souvent prohibitifs des données en outre soumises à certaines restrictions légales en termes de diffusion. Par ailleurs, les craintes concernant les disparités de résolution, de référentiel et de projection se sont avérées fondées. Les incohérences constatées aux frontières sont notables et rendent les interprétations des indicateurs peu aisées.

Le seul organisme officiel fournissant des découpages administratifs pour l'ensemble de l'Union européenne pour des entités infranationale est l'Office européen de la Statistique. Son site internet met à la disposition de tous et gratuitement une carte des découpages administratifs des niveaux NUTS 1, NUTS 2 et NUTS 3. La précision de ces données est médiocre (échelle de référence 1 : 1.000.000) et ne règle pas le problème des niveaux supra-communaux, communaux et infra-communaux, niveaux les plus importants pour les acteurs locaux.

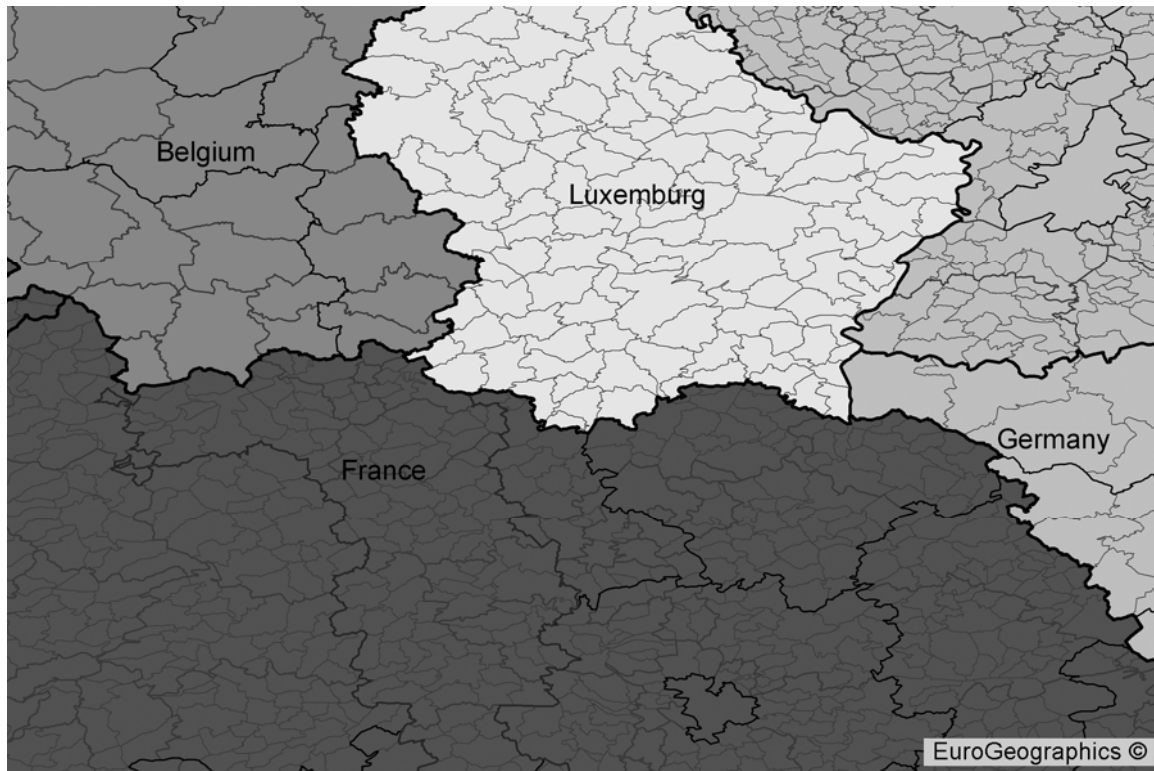
Fournisseur unique

Aucun organisme public traditionnel ne fournit des données depuis le niveau des pays jusqu'aux communes, principal objectif de notre collecte d'information, les « communes² » étant une unité de base dans tous les pays européens. Nos recherches nous ont donc amené à chercher des distributeurs privés ou parapublics pour ces données. L'utilisation de produits d'un fournisseur unique ayant par ailleurs harmonisé les données permet de s'affranchir d'un nombre important de contraintes citées auparavant.

Nous avons trouvé un seul distributeur de données pour l'ensemble de l'Union européenne et certains pays partenaires³. Ce fournisseur, EuroGeographics, met à disposition divers produits à l'échelle de l'Union dont une carte des découpages administratifs ayant pour base géométrique la plus petite les communes ou équivalent. En outre, la base attributaire de ces fichiers assure la relation entre certaines entités infra-communales et les entités supérieures. La base attributaire de ce fournisseur élude également la question de la diversité des identifiants (rencontrée pour des fournisseurs nationaux) puisque ceux-ci sont harmonisés selon la nomenclature européenne

NUTS / LAU. Chacune des 125.286 entités représentées est donc identifiée par son code unique. Les données sont régulièrement mises à jour par la société.

Afin de constituer son produit, EuroGeographics s'est associé avec les différents distributeurs nationaux. Les distributeurs nationaux restent donc détenteurs des droits de copyright mais EuroGeographics peut diffuser les produits des différents pays.



Le coût d'utilisation de ces données est prohibitif ; une licence pour un utilisateur unique doit être acquise à hauteur de 15.340 euros pour l'UE27 et une durée de 2 ans. Des charges supplémentaires doivent être prises en compte pour une diffusion sur internet. Néanmoins, EuroGeographics dispose d'accords avec l'Union européenne : les projets de recherche européens peuvent disposer des données EuroGeographics durant la durée du projet et pour une utilisation interne seulement.

L'utilisation d'un tel produit est optimale car elle permet de disposer de couches d'information uniformes dont la précision est satisfaisante au regard des objectifs de l'outil. Néanmoins son coût à terme rend son exploitation impossible pour un nombre important d'acteurs locaux aux moyens financiers réduits. Il serait intéressant de pouvoir développer dans un futur proche un produit similaire, dont l'accessibilité ne serait pas soumise à de telles contraintes financières.

3. DONNÉES ET MÉTADONNÉES

Le volume de données traitées par le système de représentation de l'information territoriale est considérable : il s'agit de distribuer des données de 27 pays, pour de nombreux indicateurs. Potentiellement, compte tenu des niveaux administratifs, cela représente plusieurs millions d'informations. Les études précédentes (CAENTI, 2008) ont par ailleurs montré que les sources des indicateurs sont nombreuses, et les conditions d'obtention particulièrement diversifiées.

Dans ce contexte, plusieurs arguments sont en faveur de la collecte et l'agrégation de métadonnées décrivant à la fois les données elles-mêmes, et les indicateurs. Premièrement, les utilisateurs finaux doivent pouvoir analyser les données en connaissant les méthodes de calcul des indicateurs, l'année des données, leur source, etc. En second lieu, les métadonnées sont utiles dans le cadre des mises à jour et des divers traitements automatiques pouvant être effectués sur la base de données attributaires.

Les recherches menées dans le cadre du programme CAENTI visent à harmoniser les métadonnées et les adapter aux données géographiques et aux indicateurs. Les recherches menées en ce sens s'appuient sur le Dublin Core et sur les normes ISO. Le travail a abouti à une proposition de grille pour les métadonnées, laquelle a été testée avec succès sur des indicateurs européens et nationaux, ainsi que sur des fichiers géographiques. Les champs seront validés par la comparaison avec les conclusions du projet INSPIRE⁴ qui doit aboutir en novembre 2008.

4. L'INTERFACE CARTOGRAPHIQUE

L'interface cartographique est l'élément du système de présentation que les utilisateurs finaux verront le plus. Celle-ci, dans sa conception, doit s'adapter à tous les environnements informatiques afin d'avoir un public potentiellement plus important. Néanmoins, il faut souligner qu'un bon portail d'information dépend moins de sa physionomie que de ce qu'il contient, mais celle-ci contribue à la visibilité et à la maniabilité de l'outil.

Il s'agit d'un outil de *web mapping* permettant de réaliser des cartes interactives. Les développements de cette interface sont précisés par l'article (voir référence). En plus de cartographier les valeurs des indices, l'interface permet également de prendre connaissance des métadonnées renseignant les utilisateurs sur les conditions d'obtention et de création des données visibles sur la carte.

L'interface cartographique utilise donc nécessairement l'ensemble des ressources préparées précédemment. Le prototype répond aux exigences que nous nous étions fixés initialement, en se basant sur des technologies *open-source*.

5. CONCLUSION

La réalisation d'un portail d'indicateurs européen comme élément constitutif du processus cognitif des territoires nécessite un travail de fond particulièrement important. Après une sélection des indicateurs représentés sur base du guide adapté à la méthode CATALYSE, des questions relatives à l'obtention des données et des métadonnées ont été soulevées. Elles ont mis en évidence la nécessité de collecter l'une et l'autre simultanément pour informer au mieux l'utilisateur final.

La collecte d'information a été également particulièrement instructive quand à la disponibilité et le coût d'acquisition de l'information géographique ou attributaire. Les disparités de diffusion de l'information démontrées au cours des recherches montrent également les difficultés rencontrées par les acteurs pour obtenir des informations territoriales abondantes et pertinentes. La centralisation de l'information et sa normalisation semblent être des prérequis essentiels à des processus de connaissance du territoire efficaces et accessibles pour les divers acteurs.

L'utilisation d'une interface cartographique interactive basée sur les technologies de *web mapping* permet d'exploiter avec pertinence les éléments mis en œuvre auparavant et de diffuser largement auprès d'utilisateurs parfois non aguerris aux méthodes de cartographie des informations essentielles à la compréhension des territoires. L'existence d'un prototype fonctionnel basé sur des données réelles et territorialisées tend à confirmer la faisabilité d'un portail à l'échelle européenne, moyennant des améliorations substantielles au niveau de l'accessibilité des données, objectif qui sera probablement atteint grâce aux engagements européens au travers de la directive INSPIRE.

¹ *Shapefiles est un format de fichier initialement mis en œuvre par la société ESRI, référence de facto dans les logiciels de cartographie et de système d'information géographique (GIS). Il est désormais utilisé par une majorité de logiciels SIG ou de cartographie. Il contient une couche d'information géographique pouvant servir de support à la cartographie de données contextuelles.*

² *Les définitions de ce niveau diffèrent cependant. Néanmoins, les structures démocratiques sont globalement similaires et correspondent à un « niveau local de base ».*

³ *Les pays partenaires sont, entre autres, certains pays candidats à l'entrée dans l'Union européenne ou des pays frontaliers.*

⁴ *Le projet européen INSPIRE répond à la directive 2007/2/CE du Parlement européen qui vise à harmoniser et améliorer la diffusion de l'information géographique dans l'Union européenne. Un volet spécifique est consacré aux métadonnées géographiques.*

Bibliographie

CAENTI, 2006a, *List of territorial indicators available on internet for comparison with CATALYSE Guide Data*, Deliverable 53, 28p.

CAENTI, 2006b, *State-of-art about of territorial information available on internet and sources in Europe*, Deliverable 24, 23p.

CAENTI 2006c, *European contents specification for a CATALYSE guide for diagnosis and evaluation*, Deliverable 51, 155p.

CAENTI 2006d, *Guidance notes for the use of CATALYSE information and tools*, Deliverable 56, 192p.

CAENTI, 2008, *Reasoned catalogue of territorial information available on internet and sources in Europe*, Deliverable 29, 83p.

CHARLIER J., DEVILLET G., BRUNETTI E., CUSUMANO C., DELVOYE J.-M., DOYEN P., URBAN A., 2007, *To a TI Community System : shared diagnostic and territorial animation of a Seraing 's area (Belgium)*, in Acts of International Conference of Territorial Intelligence, Huelva 2007 (caENTI)

CHEN D., DEVILLET G., 2006, *Territorial information, themes, indicators and sources*, in Acts of the international Conference of Territorial Intelligence, Vol 2 caENTI, Alba Iulia, October 21st-22nd, 2007

VERMEYLEN P., 2006, *Aménagement et urbanisme, L'Agenda 21 des Communes durables*, Namur : CPDT, 97p.

GIRARDOT J.-J., 2006, *Activities and prospects of research activities concerning tools of territorial intelligence for sustainable developments actors*, in Acts of the international Conference of Territorial Intelligence, Vol 2 caENTI, Alba Iulia, October 21st-22nd, 2007

LAZZERY Y., 2006, *Les indicateurs territoriaux de développement durable. Questionnement et expériences*, L'Harmattan, 323p.

SANCHEZ-LOPEZ C., GIRARDOT J.-J., 2006, *Specifications of the contents of the european guide of diagnosis and evaluation*, in Acts of the international Conference of Territorial Intelligence, Vol 2 caENTI, Alba Iulia, October 21st-22nd, 2007