

# Identification des besoins en information géographique

Cristelle Roux<sup>1</sup>, Alain Lefèvre<sup>2</sup>

<sup>1</sup>GFI Benelux S.A. – 183, route de Luxembourg – 8077 Bertrange – Luxembourg

<sup>2</sup>GFI Benelux S.A. – 54, Chaussée des Collines – 1300 Wavre – Belgique  
cristelle.roux@gfi.be, alain.lefevre@gfi.be

## Abstract

The public producers of geographic information must take into account a lot of criteria to lay down their pricing policy for the diffusion of their data. The usage of this information is one of the most difficult criteria to be quantified. GFI Benelux has developed a methodology for identifying the explicit requirements in geographical information for different potential users. This method is based on semantic and bibliometric analyses using the Tropes software in particular. This approach can be seen as the beginning of one ontology.

## Résumé

Les producteurs publics d'information géographique doivent prendre en considération de nombreux critères pour définir leur politique tarifaire de diffusion de leurs données. La nature de l'usage d'une information géographique est un des critères les plus difficiles à quantifier. GFI Benelux a développé une méthodologie d'identification des besoins explicites ou non en information géographique par des usagers potentiels. Cette méthode est basée sur des analyses sémantiques et bibliométriques utilisant notamment le logiciel Tropes. Cette approche peut être le début d'une élaboration d'ontologie d'usage d'un domaine.

**Mots-clés :** usage, information géographique, ontologie, analyse sémantique, logiciel Tropes

## 1. Contexte général

### 1.1. La directive

Les administrations produisent et rassemblent, dans le cadre de leurs missions, de nombreuses données localisées ou localisables. La directive du Parlement et du Conseil européen concernant la réutilisation et l'exploitation commerciale des documents du secteur public (2003/98/CE) demande aux services publics d'établir des principes clairs de diffusion de leurs données, et d'externaliser cette diffusion autant que possible. La proposition de cette directive laisse le soin aux États membres et à ces organismes de définir leurs propres politiques tarifaires.

### 1.2. La problématique : *accessibilité de l'information géographique, expression du besoin*

La préoccupation actuelle des organismes publics est donc de rendre accessible à tout public les données géographiques et de définir leurs coûts tarifaires. Les études réalisées mettent régulièrement en évidence l'inadéquation des produits d'information géographique aux besoins des usagers, ce qui rend la détermination de leur valeur (et de leur prix) très complexe.

Dans le cadre de ce colloque nous limiterons la présentation de nos travaux à la notion d'accessibilité de l'information, la détermination de sa valeur économique sortant de la thématique des JADT.

Tel que le précise le projet MADAME, même si les termes « accès aux données » et « diffusion » sont souvent utilisés de manière interchangeable, une certaine différence existe entre les deux. L'accès aux données peut être envisagé comme une politique plus réactive : si un utilisateur potentiel demande à accéder à un ensemble de données, sa requête est prise en compte et l'accès peut lui être accordé moyennant une série de conditions et de frais éventuels. La diffusion de données peut être considérée comme une politique plus proactive dont l'objectif est d'encourager l'utilisation des données. Elle comprend dès lors l'accès aux données mais elle s'étend avec un effort plus positif, par exemple, à la documentation de ces données, à la fourniture d'applications, au développement de services d'assistance aux utilisateurs et elle cherche à élargir la base des utilisateurs.

D'autre part, l'accès d'utilisateurs aux données géographiques, c'est-à-dire au potentiel d'information dérivable des données géographiques, pose des problèmes au niveau de l'expression du besoin d'information géographique d'un utilisateur (Bucher, 2002). Celle-ci est très loin de la description de l'utilisation de données géographiques pouvant dériver des SIG. L'expression du besoin renvoie souvent à des concepts utilisateurs, c'est-à-dire au contexte du besoin, et non aux ressources utiles pour y répondre. Or les utilisateurs ont leurs propres représentations de l'espace géographique qui sont rarement celles utilisées par le producteur de données. Ils expriment leurs besoins avec leurs propres termes spécifiques. L'auteur (Bucher, 2002) parle ici des utilisations finales des données géographiques et non des utilisations qui consistent à construire un service de dérivation d'information pour des utilisateurs finaux potentiels.

Les applications des données géographiques sont classiquement résumées en une liste de questions élémentaires auxquelles ces applications apportent des réponses. Les deux principales questions sont :

- Où ? : La localisation. Exemple : où se trouve l'Atomium ?
- Qu'y a-t-il là ? : L'adressage. Exemple : quels sites remarquables y a-t-il dans ce pays ?

D'autres questions, classiquement identifiées dans la littérature, sont :

- Et si ? : La simulation. Exemple : Si Bruxelles n'existait pas, quels seraient les flux entre les principales villes belges ?
- Quand ? : Qu'est-ce qui a changé ? : L'analyse du changement. Exemple : quand a-t-on construit ce pont ?
- Comment ? : L'analyse spatiale. Exemple : Comment puis-je me rendre là ? Comment est structurée (géographiquement) cette ville ?

Les informations permettant de répondre à ces dernières questions nécessitent des analyses des données géographiques et des documents textuels composites résultant d'études préalables sur le même sujet ou bien sur un sujet connexe. Le SIG (Système d'Information Géographique) peut fournir les données géographiques mais il n'est pas fait pour gérer la production des documents composites.

## **2. Mise en évidence des usages de l'information géographique**

### **2.1. Choix de la méthode**

Rendre accessible une information géographique – qui découle d'une donnée géographique ou d'une combinaison de données géographiques – revient à identifier les besoins en information géographique par des usagers potentiels. Identifier ces besoins signifie également prendre en considération les caractéristiques mentionnées ci-dessous :

- La **capacité d’atteindre l’information géographique** par un utilisateur potentiel. Le niveau « d’appréhension » de l’information ne sera pas le même pour un géomètre que pour un citoyen. Il va donc dépendre d’une catégorie d’utilisateurs. Ainsi un produit fini pour un utilisateur intermédiaire sera encore inaccessible pour un citoyen.
- La **nature de l’usage** de cette information va influencer sur le choix des données et des traitements nécessaires pour répondre au besoin. La nature de l’usage peut se décliner en domaine sectoriel (agriculture, urbanisme, tourisme, santé etc.) Or une difficulté majeure est l’écart entre le vocabulaire des utilisateurs et celui des bases de données géographiques. Par exemple un transporteur voulant effectuer un calcul d’itinéraire, un agriculteur voulant optimiser ses épandages, ne peuvent pas exprimer leur besoin de données géographiques à l’aide de concepts utilisés pour décrire des bases de données géographiques.

Pour mettre en évidence les besoins explicites ou non pour l’usager, nous souhaitons mettre en place une méthode ré-exploitable et objective (sans a priori) en terme de détection des besoins. Pour cela l’analyse de discours nous a semblé être la plus appropriée. Identifier un besoin par une analyse sémantique revient alors à mettre en évidence des relations sémantiques entre des acteurs et des données géographiques au sein d’un même texte ou d’un même paragraphe.

Pour soutenir ces travaux, nous avons choisi comme outil d’analyse sémantique la technologie de Tropes. Tropes est un logiciel de recherche en langage naturel et d’analyse documentaire fondé sur la compréhension du contenu qu’il a à traiter. Chaque terme significatif est inscrit dans une chaîne sémantique : un mot suffit pour retrouver tous les documents le contenant, lui et ses équivalents. Par exemple, une recherche sur « élection » retrouvera des textes contenant « voix », « urne » et « vote », même si le mot « élection » ne se trouve pas explicitement dans le texte. Tropes propose de créer des classifications de termes qui peuvent ensuite permettre d’interpréter le contenu du texte (cf. [www.acetic.com](http://www.acetic.com)).

A partir de ces objectifs et en fonction de l’outil choisi, nous avons alors décliné ces travaux selon les étapes suivantes :

- Constitution de classes :
  - De termes propres « au monde » de l’information géographique, termes géomatiques
  - D’usagers potentiels de l’information géographique
  - De termes thématiques : le domaine de l’agriculture a été choisi pour limiter notre démarche dans un premier temps.

Remarque : pour chaque classe une validation par des experts est nécessaire. Ce processus long prendra du temps, mais un processus d’amorçage est aujourd’hui nécessaire.

- Mise en relation des termes géomatiques et des catégories d’usagers afin de mettre en évidence les usages (et donc les besoins) de l’information géographique
- Proposition d’une classification en thématique des besoins détectés

## **2.2. Constitution des dictionnaires et classification des termes**

Pour constituer ces glossaires, notre démarche a été double : identification de glossaires existant sur internet ou fournis par les correspondants des administrations de la Région wallonne et identification de nouveaux termes par indexation de sites Internet.

### *2.2.1. Recensement de glossaires de termes techniques en SIG et classification*

Les sources exploitées sont le dictionnaire de données de la Région wallonne (projet META-

TER), le fond de plan cartographique à grande échelle de la Région wallonne, le lot de données PLI (Plan de Localisation Informatique), le glossaire proposé par le site internet du Centre Régional de l'Information Géographique de PACA et de l'université de Laval. Lors du recensement des différents glossaires, nous constatons que les termes sont « mélangés ». On y trouve à la fois :

- Des termes très techniques purement experts : vecteurs, sections, système d'arpentage des terres, triangulation ...
- Des termes techniques de la géomatique mais qui s'approchent des applications : bathymétrie, topographie, cartographie, télédétection, géomorphologie
- Des termes qui constituent des objets de cartes compréhensibles par tous : vallées, routes, forêts
- Des termes géomatiques à connotation applicative : carte maritime, pédologique, etc.

Nous proposons la classification suivante :

#### TERMES GEOMATIQUES

- Cartographie
  - Cartographie
  - Echelle de carte
  - Légende
  - Topographie
- Géodésie
  - Géomatique
  - Géoréférencement
  - Géoscience
  - Télédétection

- Objet géographique
  - Adresse postale
  - Information touristique
  - Objet topographique
  - Occupation du sol (Corine)
- Lots de données Métater

#### GEOGRAPHIE

- International
- Nation
- Pays
- Echelle géographique

#### 2.2.2. Recensement et classification de la liste des ACTEURS

Les sources pour constituer la liste des acteurs proviennent du code métier (ROME) de l'agence nationale pour l'emploi française, du vocabulaire déjà existant au sein de la classe « Gens » de Tropes, du guide 2000 des administrations de la Région wallonne, de sites internet d'organismes européens SIG. Nous avons classé la liste de ces acteurs en tenant compte de l'aspect économique qui peut intervenir lors de l'usage d'une information géographique. Pour élaborer cette classification, nous nous sommes ainsi basés sur les travaux de l'ASIT-VD qui constate que, même si le marché de l'information géographique est diversifié, il en ressort deux tendances :

- **Des utilisateurs professionnels** qui ne peuvent se passer de l'information pour la réalisation de leurs activités (p.ex. services des administrations publiques, gestionnaires de réseaux, autorités locales, géomètres).
- **Des intermédiaires.** Ces derniers se composent des catégories suivantes :
  - agents qui fournissent l'information originale sans valeur ajoutée
  - éditeurs qui travaillent la donnée afin de fournir des produits et services avec valeur ajoutée
  - consultants, mandataires qui achètent et traitent l'information dans le cadre d'un mandat.

Ces intermédiaires bénéficient en général d'un pourcentage sur la vente de la donnée originale et facturent principalement la valeur ajoutée.

- **Les autres utilisateurs** sont plus dispersés et sensibles au prix de l'information géographique car elle n'est pas indispensable aujourd'hui mais le deviendra certainement pour la réalisation de leurs activités. Par exemple, dans le cadre des lignes de vies prônées par les instances opératrices de e-Gouvernement.

Nous proposons donc la classification suivante :

#### FOURNISSEUR PUBLIC

- Organisation SIG

#### INTERMEDIAIRE

- Consultant
- Editeur

#### STRUCTURE

- Instance non gouvernementale

#### UTILISATEUR

- Instance gouvernementale de la Région Wallonne
- Autorité locale
- Citoyen
  - Qualité du citoyen
  - Groupe social
  - Profession
- Secteur privé

### 2.3. Constitution de corpus pour identifier les besoins en information géographique

La source la plus accessible en terme de gratuité et de facilité d'accès étant l'internet, nous avons constitué un corpus (ensemble de textes) à partir de plus de 500 sites internet :

- **Par acteur** : nous avons privilégié la constitution d'un sous-corpus sur les bureaux d'étude car ils nous semblaient être les interlocuteurs intermédiaires les plus à même de mettre en évidence les liens entre usagers et termes géomatiques. Nous avons également analysé des sites web d'entreprises wallonnes et d'institutions belges ;
- **Par thématique** : afin d'évaluer l'efficacité de la méthode, nous avons traité des sites internet sur la thématique de l'agriculture et plus particulièrement sur les informations que fournit l'usage de la donnée « parcelle agricole » c'est-à-dire ses caractéristiques topographiques ainsi que le type de culture.

## 3. Résultats

### 3.1. Premiers retours

#### 3.1.1. Adaptation de la méthode

Les premiers résultats ont démontré que cette mise en relation existe au sein des textes mais qu'elle n'est pas toujours évidente car les termes géomatiques restent encore trop « experts », trop techniques : ils ne permettent pas de définir le ou les usages potentiels finaux. Nous avons alors élargi les possibilités de mise en évidence des préoccupations de positionnement d'usagers en information géographique par les 2 procédés suivants :

- Définir les données ou informations géomatiques à partir d'un langage commun encore plus simple, plus « compréhensible » par les utilisateurs. Nous avons ainsi créé une nouvelle sous-classe qui regroupe des termes génériques tels que « localisation, emplacement, site, lieu ».
- Mettre en évidence les relations entre acteurs et termes géomatiques via le domaine d'activité étudié.



La constitution et la classification d'une liste de termes propres à l'agriculture a été réalisée à

partir du vocabulaire déjà existant au sein de la classe « agriculture » de Tropes, de l'identification de nouveaux termes provenant de sites Internet en agriculture.

### 3.1.2. Développement informatique

La technologie de Tropes met en évidence les liens existants entre des termes mais elle ne permet pas de choisir des classes de termes à mettre en relation. Par conséquent, la réalisation technique de ces croisements a nécessité un développement informatique spécifique. Nous avons associé pour cela le traitement sémantique de Tropes à des méthodes de co-occurrence. Cette mise en relation des classes de termes peut être réalisée à l'échelle du texte ou du paragraphe.

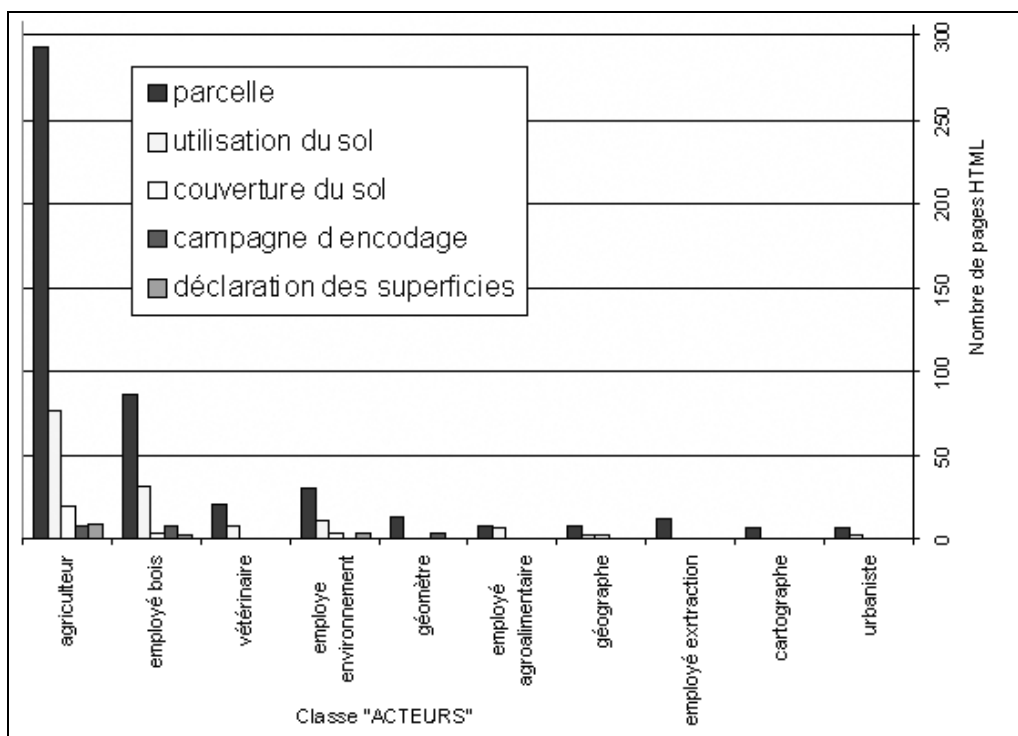


Fig. 1. Nombre de pages html contenant au sein de leur texte des termes de la classe « ACTEURS » et des termes de la classe « GEOMATIQUE »

## 3.2. Mise en évidence des usages de l'information géographique

### 3.2.1. Exemples de besoins identifiés

L'analyse sémantique du corpus a permis de mettre en évidence des usages des informations liées aux parcelles agricoles. Les textes suivants présentent quelques extraits des pages internet qui explicitent ces usages.

#### Réduction des largeurs et surface minimale des parcelles en jachère au bord des lacs et cours d'eau

Un règlement de la Commission européenne du 27 décembre 2000 précise que les parcelles de jachère d'un seul tenant situées le long de cours d'eau et de lacs pérennes peuvent avoir une largeur minimale de 10 mètres au lieu de 20 et une surface minimale de 10 ares au lieu de 30.

#### Institut Technique des Céréales et des Fourrages

Pour anticiper et adapter vos conseils techniques tout au long de la campagne, l'ITCF vous fournit des informations et des conseils de culture personnalisés (blé tendre, orge de brasserie, blé dur ?).

**Repères® Conduite s'appuie sur un observatoire de parcelles suivies localement et la mise en oeuvre de modèles agrophysiologiques** utilisant les bases de données ITCF et la météorologie locale

A chaque étape-clé de la culture, l'ITCF observe, estime, analyse les informations et rédige une synthèse comprenant : l'analyse climatique de l'année, l'évaluation des risques climatiques et leurs conséquences sur la culture, **l'état des cultures**, la prévision des stades, **la tendance du rendement** et de la teneur en protéines de la récolte.

### 3.2.2. Évaluation quantitative des besoins par les catégories d'utilisateurs

Par les combinaisons de mise en relation entre classes présentées précédemment, on détecte les textes susceptibles de mettre en évidence des usages de l'information géographique par des catégories d'utilisateurs. Nous traduisons ces résultats en terme quantitatif sous forme d'histogramme (fig. 1). Ces résultats apportent également une meilleure appréciation des classifications.

### 3.2.3. Proposition d'une classification des usages identifiés

Le recensement des usages démontre la multiplicité des usages de l'information liés aux parcelles agricoles. Avec l'aide d'experts du domaine, nous avons construit la classification suivante :

- **Gestion des sols** (ex : Création d'une cartographie de la végétation naturelle)
- **Gestion des territoires agricoles** (ex : Choix d'emplacement du verger)
- **Remembrement** (ex : rétablir les limites des parcelles agricoles)
- **Production-Productivité**
  - *Agriculture de précision* (ex : *Optimiser les engrais de ferme dans le plan de fumure*)
  - *Les récoltes* (ex : Détection de la date de semis des betteraves)
  - *Traçabilité* (ex : Traçabilité des origines des céréales)
- **Aménagement paysager** (ex : Mesurer l'évolution entre les espaces artificialisés, les espaces agricoles et les zones « naturelles » (forêts, zones humides etc.)
- **Gestion de l'irrigation et du drainage** (ex : Aide à la gestion de grands périmètres irrigués)
- **Gestion des subventions et règlements** (ex : Obtenir prime pour mise en jachère)
- **Administration** (ex : Aide à la comptabilité, au suivi de la législation agricole, des directives de subvention, des règlements d'imposition, de l'assurance récolte.)
- **Luttes contre les maladies** (ex : Lutter contre le risque d'infiltration de la bactériose vasculaire sur les cultures de pommes de terre)
- **Risque et Sinistre** (ex : Évaluation de déclaration de sinistre)
- **Transport** (ex : Gestion de réseau de distribution pour des entreprises agricoles)
- **Impact sur environnement**
  - *Contrôle de la pollution* (ex : gérer les plans d'épandages)
  - *Les nitrates* (ex : Limiter les émissions de N2O)
  - *Gestion de l'eau* (ex : améliorer la qualité des eaux agricoles)
  - *Gestion des déchets* (ex : Droits internes)
  - *Gestion du sol* (ex : Eviter le ruissellement)
  - *Epidémiologie* (ex : l'innocuité des nitrates)
  - *Gestion des ressources naturelles* (ex : Aide à la gestion des ressources naturelles)
  - *Gestion de la faune* (ex : plans de rétablissement des espèces en péril)
- **Formation**

- *Formation*
- *Démarche didactique*

#### 4. Conclusion

L'évaluation des résultats est, au stade de nos travaux, essentiellement d'ordre qualitatif. La méthode proposée permet bien de détecter des besoins en information géographique par type d'utilisateurs potentiels. Les premiers résultats nous démontrent que la liste des termes appartenant à la classe du domaine étudié ne doit pas être exhaustive. Nous avons en effet obtenu à peu près le même nombre de pages html contenant un usage à partir d'une classe moins détaillée.

Cependant la performance de la méthode est à améliorer. Même si la combinaison des mises en relation des classes permet de ressortir les textes qui émettent des besoins, l'interprétation de ces besoins reste manuelle. L'attribution de la qualité d'un besoin tel que nous pourrions le proposer selon la classe USAGE ne pourra se faire automatiquement que par de nouveaux développements informatiques. Nous pourrions pour cela envisager d'attribuer à chaque besoin les termes qui lui sont les plus fréquemment associés. Nous identifierions ainsi une liste de termes « contextuels » de l'utilisation de la donnée géographique. Cette démarche correspond en fait aux premières étapes nécessaires pour l'élaboration d'ontologie du domaine. Toute la difficulté est par la suite de mettre en évidence de manière automatique la logique qui lie les termes pour chaque usage (ou besoin) étudié.

#### Références

- Bucher B. (2002). *L'aide à l'accès à l'information géographique : un environnement de conception coopérative d'utilisations de données géographiques.*
- CNIG. (2002). <http://www.cnig.fr>
- COSIG.(2001). *Structure et politiques de tarification des données géographiques au sein de l'administration fédérale : étude comparative et propositions d'actions.*
- CPDT. (2002). <http://lepur03.geo.ulg.ac.be>
- Eigenheer K. *et al.* (2001). Étude pour la définition d'une politique tarifaire des données géographiques. *ASIT-VD*
- EUROPEAN COMMISSION. (2002). *Directive du parlement européen et du conseil concernant la réutilisation et l'exploitation commerciale des documents du secteur publics (2002/0123 (COD))*
- Heymans L. *Le fond de plan cartographique à grande échelle de la région wallonne.*
- INSPIRE. (2002). *Environmental topic user needs.*
- Lame G. (2000). *Acquisition de connaissance à partir de textes, vers l'élaboration d'une ontologie de droit.*
- Ministère des ressources naturelles et secrétariat du conseil du Trésor du Québec. *Mise en place de l'infrastructure gouvernementale : une opportunité pour favoriser la réutilisation géographique afin d'améliorer la performance de l'état et les services offerts aux citoyens et aux entreprises »*
- Projet MADAME (Method for Acces to Data and Metadata). (INFO 2000). Practitioners' forum on the economics of geographic information and pricing strategies.
- Université de Liège LEPUR. (2002). *Le dictionnaire de données de la Région Wallonne (projet METATER).*