

Cytise : une application web couplée à une base de données démographiques et socio-économiques des communes wallonnes

AUTEURS

Luc DAL, Université catholique de Louvain (Belgique)
Thierry EGGERICKX, Université catholique de Louvain (Belgique)
Jean-Paul SANDERSON, Université catholique de Louvain (Belgique)

RÉSUMÉ

Cytise communes est une application web connectée à une base de données qui couvre un ensemble de thèmes démographiques et socio-économiques, dans les dimensions spatiale et chronologique. Sa principale caractéristique est d'être un outil utile, convivial et orienté utilisateur. La construction de la base de données est détaillée et les différents types d'interrogations possibles sont présentés. Quelques exemples d'utilisation sont proposés. Développée initialement pour répondre à une demande d'appui scientifique à la prise de décision politique, elle a vu au fil du temps son audience déborder du cadre des décideurs politiques (communaux et régionaux) pour concerner aussi les étudiants, les chercheurs, les enseignants du secondaire et le monde associatif. L'objectif de cette communication est donc de présenter le contenu de cet outil et de mettre en évidence ses potentialités dans le domaine de l'aide à la décision.

MOTS CLÉS

Application web, démographie wallonne, migrations, ménages, logements, cartes interactives

ABSTRACT

Cytise communes is a web application connected to a database that covers a set of demographic and socio-economic data, in spatial and chronological dimensions. Its main feature is to be a useful, user-friendly and user-oriented tool. The construction of the database is detailed and the different types of query are presented as well as some outputs produced by the application. Initially developed in response to a request for scientific support to policy-making, it has seen over time his audience growing beyond the scope of policy-makers (municipal and regional) to involve students, researchers, high school teachers and associations. The purpose of this communication is to present the content of this tool and to highlight its potential in the field of decision support.

KEYWORDS

Web application, Walloon population, migrations, households, housing, interactive maps

1. LES OBJECTIFS ET LE CONTEXTE

Cytise communes trouve son origine dans une demande des autorités politiques régionales wallonnes qui souhaitaient développer un outil scientifique d'aide à la prise de décision. Dans les années 1980, Michel Poulain développa un logiciel de comptabilité démographique applicable à toute population spécifique (commune, quartier, population étrangère...) à partir des données individuelles du Registre national des personnes physiques. Ce Registre national, opérationnel légalement depuis 1988, est la version centralisée et informatisée des registres de population communaux qui existent en

Belgique depuis 1846. C'est donc un gigantesque fichier informatique centralisé à finalité administrative qui reprend, pour chaque individu présent sur le territoire belge, sa situation (sexe, date de naissance, nationalité, adresse précise, caractéristiques du ménage auquel il appartient...), ainsi que ses mouvements (naissance, décès, migrations, changement de nationalité, changement d'état civil). Il est actualisé en temps (presque) réel par les communes qui enregistrent tout mouvement ou modification de situation individuelle. Cette exploitation statistique inédite de données purement administratives, dans une perspective scientifique, permet entre autre l'élaboration des « atlas géostatistiques des quartiers » des villes de Namur et de Charleroi au début des années 1990. Une seconde étape fut franchie en 1992, dans le cadre d'un programme initié par le ministère de la Région wallonne qui souhaitait créer un outil d'aide à la décision en matière de gestion démographique. L'équipe du Centre de recherche en démographie et sociétés de l'université catholique de Louvain (DEMO-UCL) entreprit le développement scientifique et informatique de la première version de *Cytise* à l'échelle des 262 communes de Wallonie ; elle n'était à l'époque distribuée que de manière « confidentielle » sur CD. Au début des années 2000, afin de résoudre les problèmes liés à la nature même du support (difficulté de mise à jour des données, incompatibilité de l'application entre les différents systèmes d'exploitation et leurs versions successives, manque de visibilité), il fut décidé, en accord avec les autorités de la Région wallonne, de diffuser l'information au moyen d'un site web : l'objectif était de permettre à tous les services de la Région wallonne d'accéder à cette base de données inédite. Dans un premier temps, le site ne fut accessible qu'à l'administration wallonne et, depuis 2005, il est ouvert au monde extérieur et peut être consulté librement et gratuitement par tout un chacun. D'une simple application tournant sur PC, *Cytise communes* est donc devenue une application web ¹.

L'objectif principal de *Cytise communes* est donc de fournir une connaissance détaillée et comparative d'une thématique donnée, dans un espace choisi, que ce soit à un instant donné ou au fil du temps. Les indicateurs, tableaux et graphiques proposés sont donc le fruit d'une réflexion visant à fournir l'information la plus pertinente qui soit et qui puisse être utilisée pour répondre parfaitement aux questions que peuvent se poser les utilisateurs.

2. LE CONTENU DE *CYTISE*

Les thématiques abordées

Dans le temps, *Cytise communes* a largement débordé du cadre strictement démographique pour intégrer l'ensemble des données sociales, économiques et environnementales disponibles au niveau des communes et des quartiers, le tout dans les dimensions à la fois chronologique et spatiale.

Les données démographiques concernent aussi bien la population dans ses différentes structures (âge, sexe, nationalité, état matrimonial) que les modifications subies par cette population, par suite de naissances, décès, changements d'état-civil ou de nationalité, migrations internes et externes. La dimension ménage, variable fondamentale de par son implication en termes de logements, est également développée. La richesse des données extraites du Registre national et sa profondeur temporelle (plus de 20 ans) font que la thématique démographique est celle qui est la plus détaillée dans *Cytise*.

1 www.cytise.be.

Les données « extra démographiques » résultent de l'exploitation des données individuelles du recensement de la population et des ménages du 1^{er} mars 1991 et de l'enquête socio-économique du 1^{er} octobre 2001. Elles couvrent des aspects plus socio-économiques : le logement (superficie, type, occupation, commodités, isolation...), le niveau d'instruction, l'activité professionnelle, la santé subjective, l'appréciation des répondants quant à la qualité de leur logement et les conditions de vie dans leur quartier... D'autres sources de données permettent enfin d'enrichir ces données : on mentionnera la fiscalité, l'occupation du sol, le chômage de la population dont la profondeur temporelle est variable. Ces données sont produites par la Direction générale statistique et information économique (DGSIE) et l'Institut wallon de l'évaluation, de la prospective et de la statistique (IWEPS).

Les niveaux de découpage spatial

Cytise communes ne porte que sur l'espace wallon et le niveau de découpage spatial retenu est celui de la commune (262) et des entités de niveau supérieur : arrondissements (21), provinces wallonnes (5), ainsi que les trois régions du pays. *Cytise quartiers* concerne les quelque 3 000 quartiers wallons.

La dimension temporelle

Les données démographiques ont une profondeur de plus de 20 ans, ce qui permet d'avoir une vision à plus long terme de certains phénomènes, en particulier le vieillissement de la population et les comportements migratoires. Dans le cas des données extraites des recensements, nous ne disposons actuellement que de deux « photos » en 1991 et en 2001. Les autres données ont un recul qui est variable selon les cas.

3. LES SOURCES DE DONNÉES ET LE RESPECT DE LA VIE PRIVÉE

La base de données est construite soit à partir de données agrégées (emploi, occupation du sol, fiscalité...) disponibles au niveau communal et produites par différentes administrations fédérales ou régionales, soit à partir des données individuelles (ces dernières sont « anonymisées » afin de respecter la loi sur la protection de la vie privée). Il s'agit principalement des informations démographiques qui sont comptées sur la base des données individuelles extraites du Registre national à chaque 1^{er} janvier, depuis 1991 jusqu'en 2011 ². Ces données (plus de $2,5 \cdot 10^9$ « records ») reprennent la situation de chaque individu présent sur le territoire belge à chacune des 20 dates : sexe, date de naissance, lieu de naissance, nationalité, type et taille de ménage auquel l'individu appartient. En outre, nous disposons de l'ensemble des mouvements de la population depuis le 1^{er} janvier 1991 jusqu'au 31 décembre 2010. Un identifiant unique permet de connecter les différentes bases de données individuelles entre elles. Un ensemble de procédures spécifiques (écrites en SAS) permettent de constituer, à partir de ces données (individuelles ou agrégées), les tables qui alimentent la base de données *Cytise* : cette dernière est une base MySQL.

Outre les métadonnées et les fonds de cartes, cette base de données ne stocke donc que des données agrégées et en aucun cas des données individuelles. L'utilisation de ces données implique néanmoins le respect de certaines règles qui en conditionnent l'accès. Ces règles, assez strictes, permettent de concilier le respect de la vie privée et la volonté

² Des données relatives à la situation au 1^{er} janvier 2012, ainsi que les mouvements de la population des années 2009, 2010 et 2011, devraient être accessibles sous peu.

scientifique de fournir des informations précises et pertinentes.

1. Comme signalé plus haut, ces données sont « anonymisées » : d'une part, un identifiant individuel crypté remplace les noms et prénoms des individus et, d'autre part, nous ne disposons pas de l'adresse des individus qui permettrait leur localisation spatiale précise³. Cette anonymisation est réalisée directement au sein de l'administration préalablement à toute transmission de données.
2. L'accès à ces données est conditionné par l'autorisation de la Commission de la vie privée qui, en Belgique, veille au bon usage des données administratives.
3. Les données individuelles « anonymisées » ne peuvent en aucun cas être transmises à des tiers, qu'il s'agisse d'administrations, de bureaux privés ou même d'équipes de recherche universitaire.
4. Ces données ne peuvent servir à des fins commerciales et ne peuvent être utilisées que dans le cadre de projets impliquant des administrations publiques.
5. Les données accessibles aux utilisateurs de *Cytise* sont uniquement des données agrégées. De plus, dans la version « quartier » du logiciel (voir *infra* partie 6), si une variable compte moins de 5 observations, aucune donnée n'est présentée. Enfin, toujours à l'échelle des quartiers, la plupart des données disponibles sont des chiffres relatifs, exprimés sous la forme de rapports, de proportions. Les seuls chiffres absolus livrés portent sur des données de base (nombre d'habitants, nombre de ménages...) ne permettant aucune identification individuelle.
6. Les données individuelles sont conservées, cryptées, sur un serveur sécurisé de l'UCL.

L'ensemble de ces règles a été conçu pour éviter tout risque de dérapage. Elles conditionnent notre accès aux données et permettent de maintenir un équilibre entre respect de la vie privée, objectifs scientifiques et aide à la prise de décision politique.

4. L'INTERFACE D'INTERROGATION

Elle a été conçue pour que l'utilisateur puisse interroger la base de données sans aucune connaissance informatique : aucune ligne SQL n'est à écrire et les choix (type d'analyse souhaitée, thème, indicateurs, espaces(s), date(s)) sont opérés intuitivement au moyen de menus, de listes de sélection, de boutons à cliquer.

En premier lieu, l'utilisateur choisit un des 3 types d'analyse suivants : comparaisons spatio-temporelles, synthèses (sous forme tabulaire, sous forme d'évolutions chronologiques ou sous forme de cartes thématiques interactives) ou encore statistiques (univariées ou bivariées) *via* un menu déroulant.

Le type d'analyse étant fixé, on choisit un thème parmi les suivants : population, mouvements de la population, ménages, logements environnement, socio-économique, perception de la santé et de l'environnement.

Selon le type d'analyse et la thématique, un ensemble de tableaux ou de variables est proposé. La sélection d'un espace de référence (quartier, commune, arrondissement, province, région) et de celui de comparaison (autre commune, même commune à une autre date, province...) ainsi que des dates (ou des périodes) de référence et de

3 Les données de localisation spatiale dont nous disposons sont la commune et le secteur statistique, une unité d'observation administrative infra-communale.

comparaison finalisent la définition de la requête. Le bouton « envoyer la requête » lance celle-ci vers le serveur qui va renvoyer la sortie appropriée et correspondant au choix de l'utilisateur.

Les comparaisons spatio-temporelles

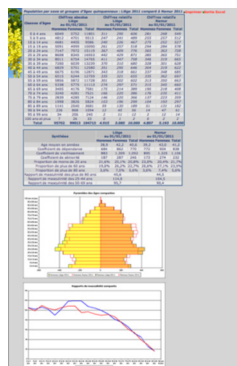
Le contenu de chacun des tableaux de comparaisons et des graphiques associés est le résultat de réflexions visant à fournir des informations essentielles aux utilisateurs, sans alourdir inutilement ces sorties avec des informations qui, à l'usage, se révèlent peu utiles ou difficilement interprétables : le souci du public cible a toujours été présent à l'esprit lors des étapes de développement. Ainsi, par exemple, toutes les sorties peuvent être exportées directement dans un tableau.

À titre d'illustration, la figure 1 reprend l'interface telle qu'elle se présente, pour produire le tableau comparatif de la population par groupes d'âges quinquennaux, pour Liège en date du 1^{er} janvier 2011 comparé à Namur à la même date. La sortie de cette requête fait l'objet de la figure 2.

Figure 1. Interface de comparaison spatio-temporelle

The screenshot shows the 'CYTISE COMMUNES' web application interface. At the top, there are logos for UCL and MEPS. The main navigation bar includes 'Accueil', 'Comparaisons spatio-temporelles', 'Synthèses', 'Statistiques', 'Nouvelles données', and 'Manuel à télécharger'. The central heading is 'Comparaisons spatio-temporelles'. Below this, there is a red text box stating: 'Cette interface vous permet de comparer un ensemble d'indicateurs d'un thème choisi, pour une commune donnée, à une date précise. Les informations du tableau seront parfaitement comparables avec celles de la commune de comparaison à la date choisie.' There are two columns of search options: 'Référence' and 'Comparaison'. Under 'Référence', 'Liège' is selected. Under 'Comparaison', 'Namur' is selected. The date is set to '2011'. A 'Tableaux disponibles' section lists various themes, with 'Population par sexe et groupes d'âges' selected. At the bottom, there is a button labeled 'Envoyer la requête'.

Figure 2. Comparaison de la population de Liège et de Namur le 1^{er} janvier 2011



Les synthèses

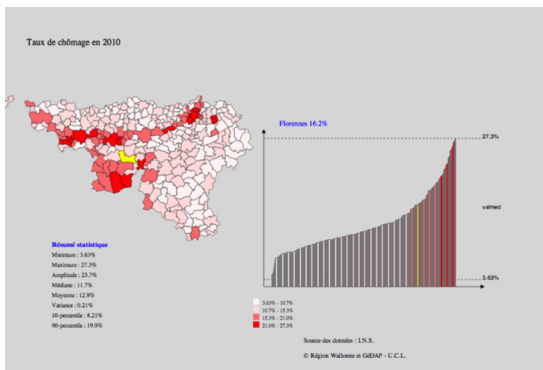
La première option des synthèses permet de choisir des tableaux qui reprennent un ensemble cohérent d'indicateurs pour un thème donné et ce pour les 262 communes. La sortie permettra de comparer directement plusieurs indicateurs.

La deuxième option permet de sélectionner un seul indicateur et de suivre sa variation au cours du temps, pour deux espaces sélectionnés.

La troisième option permet de visualiser la variation spatiale d'une variable au niveau communal (ou arrondissement ou province) au moyen d'une carte choroplèthe interactive, la méthode de discrétisation étant celle de Jenks (par défaut) ou celle des quantiles. Cette carte est présentée simultanément avec un histogramme interactif, ce qui permet, en déplaçant le pointeur sur l'écran, de localiser aisément toute commune dans l'espace wallon et dans la distribution statistique de la variable.

Par exemple, la carte 1 présente le taux de chômage communal pour l'année 2010. La commune de Florennes, avec un taux de chômage de 16 %, se situe dans le quartile supérieur.

Carte 1. Taux de chômage en 2010



Les statistiques

Un module permet de produire les statistiques élémentaires habituelles d'un seul indicateur calculé pour les 262 communes ou de croiser deux variables et de produire une analyse de régression linéaire.

5. L'AUDIENCE

Au fil des années, l'audience de *Cytise* s'est considérablement élargie : initialement destiné aux décideurs publics et l'administration de la Région wallonne, le public actuel s'est largement diversifié : par exemple, les administrations communales sont un « consommateur important » de *Cytise communes* car elles peuvent y trouver des informations précises et validées les concernant directement ; il leur serait impossible de trouver ces chiffres ailleurs ou alors éventuellement au prix de démarches administratives lourdes (et décourageantes) ; les étudiants de l'enseignement supérieur et des élèves du

secondaire font appel régulièrement à *Cytise communes* dans le cadre de leurs travaux scolaires. Les mondes associatif et citoyen (associations de personnes âgées, associations citoyennes...) puisent également beaucoup de données dans *Cytise communes* et sont des « clients » réguliers. Enfin, il faut mentionner les bureaux d'urbanisme et d'architectes qui font un usage intensif de l'application. En moyenne, le site reçoit quotidiennement une dizaine de visiteurs principalement de Belgique mais également des autres pays de l'UE (France, Suisse, Pays-Bas), voire de pays extra-européens (Canada e.a.).

6. CYTISE QUARTIERS

Une version analogue à *Cytise communes* a été développée pour les quartiers wallons. Il s'agit d'un outil essentiel et inédit au niveau de la prise de décision car il permet, par exemple, aux autorités locales d'avoir une connaissance fine et actualisée de la diversité et des inégalités socio-démographiques à l'intérieur de leur commune. Le territoire belge est partitionné actuellement en 19 781 secteurs statistiques. Ces derniers, élaborés dès le recensement de 1970 et adaptés au fur et à mesure des recensements successifs, permettent à la DGSIE de diffuser des informations à un niveau de découpage plus fin que le niveau communal, ce dernier masquant souvent des hétérogénéités importantes. Tout point du territoire appartient à un seul secteur et un secteur appartient à une seule commune. Dans la plupart des cas, ils sont simplement connexes (en un seul morceau).

La Wallonie compte à elle seule 9 877 secteurs statistiques. La production de statistiques à ce (trop) fin niveau de découpage est indigeste pour l'utilisateur « moyen ». Fort de ce constat, DEMO a procédé au découpage de chacune des 262 communes en quartiers : ils sont construits sur la base d'un certain nombre de critères objectifs (regroupements de secteurs statistiques, taille de population minimale, connexité, superficie comparable, forme la plus compacte possible...) et subjectif (sentiment d'appartenance de la population à tel ou tel quartier) et ont été validés par les autorités communales.

Cytise quartiers offre des possibilités analogues à celles de *Cytise communes*, les variables étant en quantité moindre. Signalons qu'il est possible, dans cette application, pour chaque utilisateur de définir ses propres regroupements de quartiers. Dans ce cas, chaque indicateur est recalculé exactement et n'est donc pas une moyenne de valeurs individuelles.

7. LES DÉVELOPPEMENTS ACTUELS ET FUTURS

Actuellement, deux applications et deux bases de données différentes coexistent : cet état, héritage du passé, n'est pas optimal. *Cytise* est donc en cours de refonte complète, l'objectif final étant de travailler avec une seule base de données et une seule application écrite dans un langage plus moderne (PHP) que celui utilisé actuellement (ASP). Cette opération permettra de corriger un certain nombre de « bugs » informatiques et d'enrichir, outre les données, les fonctions de la nouvelle application : on mentionnera l'extension d'un module d'édition d'agrégats personnalisés aux communes, la possibilité d'exporter les cartes thématiques et l'ajout d'un module d'analyse statistique multivariée.

RÉFÉRENCES

- Costa R., Eggerickx T., Sanderson J.-P., 2011, « Les territoires de la fécondité en Belgique au 20^e siècle. Une approche longitudinale et communale », *Espace, Populations, Sociétés*, n° 2, pp. 353-375.
- Lord S., Gerber P., Sohn C., Eggerickx T., Hermia J.-P., Kesteloot C., De Maesschalck F., 2011, "Temporal

and spatial analysis of social inequalities: 'An innovative method to grasp social inequalities evolution on the territory' ", CEAPS/GEODE Working Paper, 44.

Sanderson J.-P., Eggerickx T., 2010, « Des outils d'aide à la décision : les indicateurs de condition de vie et de mixité socio-démographique des communes et des quartiers en Belgique », *Cahiers de Démographie Locale 2009*, n° 2, pp. 115-158.

Eggerickx T., 2010, « Différences et inégalités socio-démographiques : l'approche par le local », *Espace, Populations, Sociétés*, n° 1, 2009, pp. 3-6.

Poulain M., Eggerickx T., 2007, « La démographie des populations locales peut-elle ne pas être, faute de données appropriées ? », actes du XI^e colloque national de démographie *Les populations locales*, CUDEP, Strasbourg, mai 1999, Bordeaux, pp. 9-23.

Capron C., Oris M., Eggerickx T., Poulain M., 2007, « Un quartier pour chaque âge de la vie ? Pour une nouvelle lecture des différenciations urbaines. Namur (1991-1998) », actes du XI^e colloque national de démographie *Les populations locales*, CUDEP, Strasbourg, mai 1999, Bordeaux, pp. 281-295.

Eggerickx T., Poulain M., 2007 « Une logique démographique pour le suivi du développement durable de la société, dans un cadre local », actes du XI^e colloque national de démographie *Les populations locales*, CUDEP, Strasbourg, mai 1999, Bordeaux, pp. 327-338.

Eggerickx T., 2003, « La démographie comme aide à la gestion locale », actes du séminaire de L'AIDELF en Calabre « La démographie appliquée à la gestion publique et des entreprises » (Cosenza, avril 1995) et de la session spéciale du congrès de l'EAPS (Cracovie, juin 1997), Paris, AIDELF, pp. 55-70.

Dal L., Eggerickx T., Hermia J.-P., Poulain M., 2001, *Éléments statistiques de base pour chacune des communes wallonnes dans le cadre de l'ancrage communal de la politique du logement du ministère de la Région wallonne*, GÉDAP, Université catholique de Louvain.

Dal L., Eggerickx T., Goffin E., Poulain M., 1999, *Population, logement, aménagement du territoire à l'échelle des 262 communes de la Wallonie (1991-1998)*, projet Région wallonne–GÉDAP, Université catholique de Louvain.

Goffin E., Dal L., Eggerickx T., Poulain M., 1999, « Ménages et logements : une adéquation selon le milieu d'habitat », European Population Conference, La Haye, 30 août-3 septembre 1999.

LES AUTEURS

Luc Dal

Centre de recherche en démographie et sociétés (DEMO)
Université catholique de Louvain
luc.dal@uclouvain.be

Thierry Eggerickx

Centre de recherche en démographie et sociétés (DEMO)
Université catholique de Louvain
thierry.eggerickx@uclouvain.be

Jean-Paul Sanderson

Centre de recherche en démographie et sociétés (DEMO)
Université catholique de Louvain
jean-paul.sanderson@uclouvain.be