



HAL
open science

Etude de faisabilité d'une observation pérenne des espaces transfrontaliers

Amel Feredj, Timothée Giraud, Claude Grasland, Marianne Guérois, Romain Louvet, Malika Madelin, Yves Michel, Ronan Ysebaert

► To cite this version:

Amel Feredj, Timothée Giraud, Claude Grasland, Marianne Guérois, Romain Louvet, et al.. Etude de faisabilité d'une observation pérenne des espaces transfrontaliers. [Rapport de recherche] CGET - Commissariat Général à l'Egalité des Territoires. 2014, 48 p. hal-03610547

HAL Id: hal-03610547

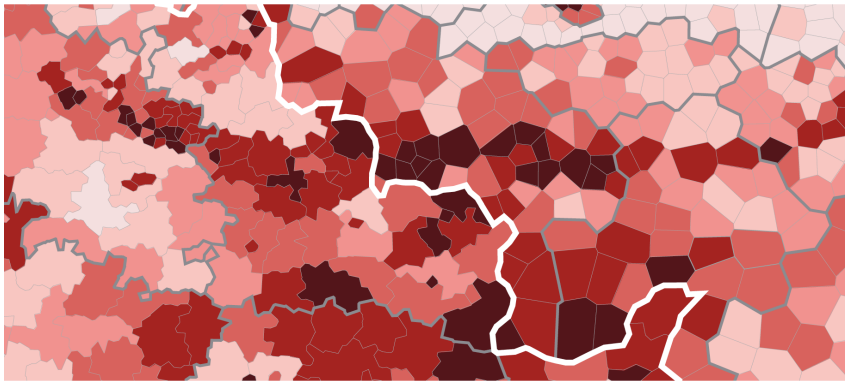
<https://hal.science/hal-03610547>

Submitted on 16 Mar 2022

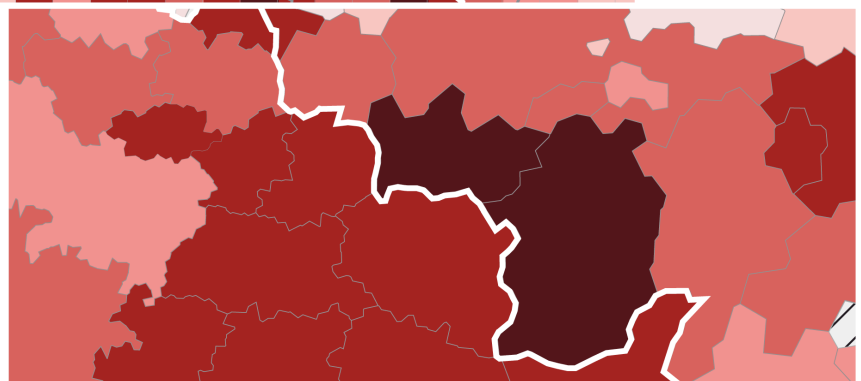
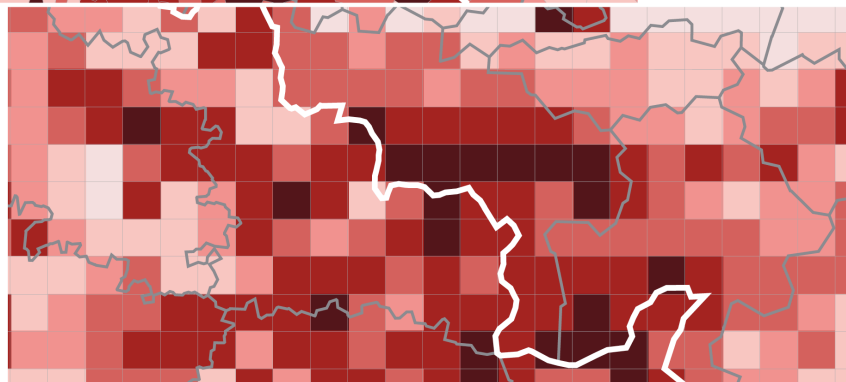
HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Etude de faisabilité d'une observation pérenne des espaces transfrontaliers



Amel Feredj
Timothée Giraud
Claude Grasland
Marianne Guérois
Romain Louvet
Malika Madelin
Yves Michel
Ronan Ysebaert



Juin 2014

Plan du rapport

Plan du rapport	3
Introduction	5
1. Enjeux méthodologiques de constitution d'un observatoire pérenne du transfrontalier	6
1.1. Approches scientifiques et institutionnelles de l'observation en contexte transfrontalier	6
1.2. Pérennisation et harmonisation des bases de données transfrontalières	8
2. Règles de bonnes pratiques pour la mise en place d'un observatoire, « volet données » : de la collecte à la diffusion des données	11
Fiche 1 - Identification des enjeux de l'observatoire, en fonction des usages et des moyens	12
Fiche 2 - Sélection des variables, des maillages pertinents d'observation, de l'échelon de collecte et du fond de carte	13
Fiche 3 – Modèle de métadonnées et modèle de données	15
Fiche 4 – Collecte des métadonnées et des données (statistiques et cartographiques)	18
Fiche 5 – Harmonisation des données entre les pays	22
Fiche 6 – Diffusion, gestion courante et mise à jour de la base	24
3. Etude de cas sur les enjeux de l'harmonisation spatiale dans l'espace transfrontalier franco-belge : des « briques de base » aux zones d'emploi	Erreur ! Signet non défini.
3.1. Objectifs, choix de l'espace d'étude et données.....	Erreur ! Signet non défini.
3.2. Harmonisation de découpages territoriaux élémentaires	Erreur ! Signet non défini.
3.3. Harmonisation de découpages territoriaux fonctionnels (zones d'emploi) ...	Erreur ! Signet non défini.
4. Propositions pour la maintenance et la gestion d'un observatoire	37
5. Synthèse des propositions	Erreur ! Signet non défini.
Références bibliographiques	45
Table des figures	47
Table des cartes	48
Table des tableaux	48
Table des encadrés	48

Cette étude a été conduite par le Collège International des Sciences du Territoires - Axe Information territoriale locale (M. Guérois & M. Madelin) et l'UMS RIATE (A. Feredj, T. Giraud, C. Grasland & R. Ysebaert), pour le compte du CGET (précédemment DATAR).

Elle a donné lieu à deux vacations :

- Romain Louvet, Assistant ingénieur - Traitement, analyse et représentation de l'information spatiale, 1 mois et demi, à mi-temps

- Yves Michel, Ingénieur d'études - Traitement, analyse et représentation de l'information spatiale, 3 mois, à mi-temps

Introduction

A l'image d'une observation menée dans un espace national, l'observation des espaces transfrontaliers se doit de collecter des données, chiffrées ou non, pour mener les analyses qui appuieront la mise en place de politiques publiques efficaces. Cette phase de recherche et de compilation de données est réalisée sur certaines frontières mais ne fait pas l'objet d'une observation partagée qui permettrait la comparaison des phénomènes à l'œuvre dans ces territoires. C'est pourquoi, depuis 2012 en France, la Délégation interministérielle à l'aménagement du territoire (DATAR) puis le Commissariat général à l'égalité des territoires (CGET) ont entrepris de porter une démarche internationale visant à promouvoir cette observation sur l'ensemble des frontières françaises, puis sur l'ensemble des frontières européennes. Les questions de coopération et de gouvernance de cette observation font l'objet d'échanges entre la France et ses partenaires frontaliers depuis 2013, au sein d'un comité stratégique et d'un groupe de travail technique. Pour alimenter ces discussions, l'Observatoire des territoires a demandé aux ingénieurs de l'UMS RIATE (Réseau interdisciplinaire pour l'aménagement et la cohésion des territoires de l'Europe et de ses voisinages) et aux chercheurs du GIS CIST (Collège Interdisciplinaire des Sciences du Territoire, axe Information territoriale locale) de produire une étude de faisabilité permettant de préfigurer la constitution d'une base de données transfrontalière. La mise en place de cette infrastructure de données suppose d'avoir au préalable une vision claire des bonnes pratiques nécessaires à la constitution d'une base de données harmonisée et pérenne. Sans prétendre recenser l'ensemble des difficultés sous-jacentes à la constitution d'une telle base, l'étude vise à en élucider les enjeux proprement méthodologiques, dans les domaines de la cartographie, de la statistique, de la géomatique et de l'analyse spatiale.

L'étude ainsi réalisée s'inscrit dans la continuité des travaux de la Mission Opérationnelle Transfrontalière (MOT) et de la Fédération Nationale des Agences d'Urbanisme (FNAU), qui ont montré que l'observation transfrontalière, qu'elle soit statistique ou plus qualitative, n'est pas une simple extension des bases de données nationales : l'architecture d'une base de données transfrontalière est nécessairement complexe, pour refléter la multitude des échelles, des définitions et des méthodes qui sont employées de part et d'autre des frontières. D'un autre côté, l'étude prolonge la réflexion sur l'harmonisation des analyses transfrontalières qui a été menée dans les deux précédents rapports coordonnés par l'UMS RIATE pour le compte de la DATAR. De plus, la coordination du projet *ESPON Database* a donné l'occasion à l'UMS RIATE de développer une expertise importante en matière de construction de bases de données pérennes, dans un contexte international. Sur la base de ces expériences, l'objectif du rapport est de proposer des éléments de réflexion sur différentes stratégies envisageables pour la constitution d'un observatoire, depuis la collecte des données cartographiques et statistiques jusqu'à la diffusion plus ou moins élaborée de l'information ainsi produite. Il sera également question des moyens humains et techniques nécessaires à cette mise en œuvre.

En préambule, nous resituerons brièvement les enjeux méthodologiques de la constitution d'un observatoire des espaces transfrontaliers parmi les autres enjeux, notamment institutionnels, d'un tel projet (Partie 1). Puis nous répondrons à la demande du CGET sur la faisabilité d'un observatoire du transfrontalier en trois temps : nous présenterons d'abord un ensemble de règles de bonnes pratiques en vue de constituer, traiter et gérer une base de données transfrontalière, en prenant en compte les objectifs et les usages potentiels de la base (Partie 2) ; dans un second temps, à partir de l'exemple franco-belge des données sur l'emploi, nous développerons un aspect particulier du traitement et de l'harmonisation spatiale des données transfrontalières, en explorant notamment la construction de zones d'emploi comparables (Partie 3). Enfin, nous proposerons trois scénarios différents de maintenance et de gestion d'un observatoire statistique et cartographique, selon différents degrés de complexité technique et de coopération internationale (Partie 4).

1. Enjeux méthodologiques de constitution d'un observatoire pérenne du transfrontalier

1.1. Approches scientifiques et institutionnelles de l'observation en contexte transfrontalier

La demande d'étude de faisabilité d'une observation pérenne des espaces transfrontaliers qui a été formulée par la DATAR auprès de l'UMS RIATE et du GIS CIST s'inscrit au croisement de trois types de démarches qui partagent formellement le même objectif mais qui, en pratique, obéissent à des logiques différentes et ne sont pas toujours faciles à concilier. Avant de préciser l'objectif propre de cette étude, il est nécessaire de brosser un portrait rapide de ces trois démarches « à l'état pur », étant bien entendu qu'en pratique elles se mélangent toujours à des degrés divers.

La démarche scientifique d'analyse comparative des espaces transfrontaliers a pour but premier la comparaison objective d'espaces transfrontalier localisés dans des contextes différents afin de tester des hypothèses générales, de dégager des lois éventuelles, et seulement en dernier lieu de mettre à jour de possibles spécificités locales. A titre d'exemple, on peut citer le cas d'une analyse comparative des discontinuités démographiques à la frontière franco-belge d'une part et le long de la frontière austro-tchéco-hongroise menée en 1992 par deux chercheurs pour l'influence des limites politiques sur les comportements démographiques (Figure 1). L'harmonisation des données (choix des maillages, des dates, des indicateurs, ...) prime ici toute autre considération puisque l'objectif premier est la comparaison objective des situations.

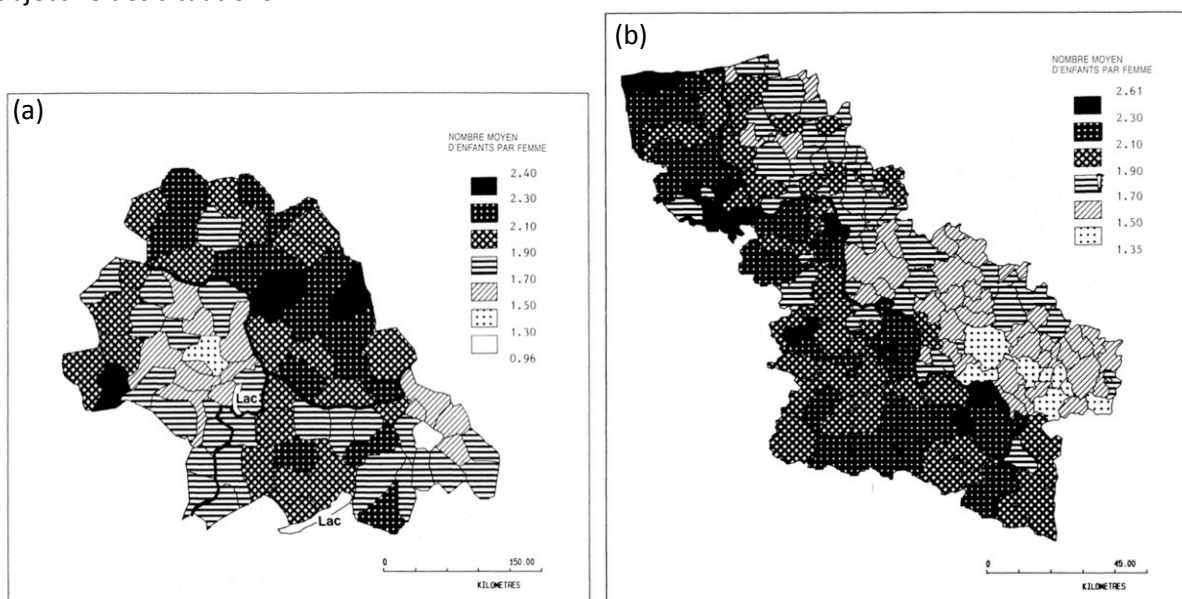


Figure 1 : Exemples d'analyse comparative des discontinuités démographiques (a) à la frontière Autriche - Tchécoslovaquie - Hongrie et (b) à la frontière France - Belgique (Decroly et Grasland, 1992)

La démarche de coopération transfrontalière portée par des acteurs locaux s'inscrit dans une logique différente de construction de liens et d'échanges de part et d'autre d'une frontière précise et singulière. La comparaison avec d'autres espaces transfrontaliers n'est pas le but premier et l'établissement d'un système statistique et cartographique commun vise en premier lieu à faciliter la réalisation de projets futurs de coopération dans tous les domaines (culture, économie, santé, ...). Il n'est pas exagéré de dire que l'échange de bases de données et la réalisation commune d'un premier atlas ont une valeur autant symbolique qu'opérationnelle. Chaque pays offrant des données établies dans une logique nationale, il y a de fortes chances que le résultat obtenu souffre d'un déficit d'harmonisation du point de vue scientifique. Mais le principe même de la coopération est celui d'une relation égalitaire entre les partenaires où il est difficile voire impossible d'imposer à l'autre son point de vue, au moins dans un premier temps.

La démarche institutionnelle de partenariat entre les instituts statistiques et cartographiques nationaux correspond enfin à un processus normatif intermédiaire entre les deux cas de figure précédents et pouvant fréquemment servir de lien avec l'un, l'autre ou les deux (Figure 2). Les instituts statistiques nationaux sont en effet les garants d'une statistique officielle de qualité imposant des contraintes strictes à la publication de résultats, notamment en termes de métadonnées, de précision, ou de marge d'erreur, ce qui les rapproche de la démarche comparatiste des chercheurs scientifiques. Mais ils sont également les représentants officiels d'une statistique nationale qui est soumise à des règles strictes de choix des définitions, des zonages, etc. L'adoption éventuelle d'un compromis avec l'organisme statistique d'un pays voisin est soumis à un processus de négociation et de validation qui met en jeu de nombreuses échelles de pouvoir (Région, Etat, Union Européenne) avec les complications que peut entraîner le degré différent d'autonomie des services statistiques locaux selon que l'on se trouve dans un pays à structure centralisée ou fédérale...

Témoignage



Joël Marty
Chargé de Mission au
Commissariat à l'Aménagement
des Pyrénées (DIACT).

« Une politique d'aménagement du territoire sur un espace transfrontalier demande une vision commune de ses caractéristiques socio-économiques »

La mise en œuvre d'une politique d'aménagement du territoire sur un espace transfrontalier comme le massif des Pyrénées, ne peut se concevoir sans l'existence d'une vision commune des phénomènes socio-économiques qui le caractérisent. En effet, la compréhension des problématiques transfrontalières s'apprécie notamment au regard d'indicateurs statistiques adaptés à la zone franco-espagnole.

Pour ces raisons, nous avons initié en partenariat avec l'Insee Midi-Pyrénées, les premiers travaux de coopération faisant intervenir l'ensemble des services statistiques concernés par cette zone de montagne. Le fruit de ce travail a donné lieu à la publication d'un atlas transfrontalier et consultable en ligne.

A partir des éléments de l'atlas qui recouvre le zonage sur lequel interviennent de nombreux acteurs publics de part et d'autre des Pyrénées, nous avons défini une analyse stratégique partagée et proposé des programmes de soutien au développement du massif des Pyrénées alimentés notamment par des crédits de l'Union européenne.

A l'avenir, les données actualisées de cet outil statistique pourront être utilisées pour l'évaluation des politiques publiques.

Figure 2 : Exemple de démarche institutionnelle de partenariat entre instituts statistiques dans le massif des Pyrénées

L'hybridation des différents types d'approche (scientifique, acteurs locaux, instituts statistiques nationaux) demeure toutefois le cas de figure le plus fréquent dans le montage des observatoires transfrontaliers, ce qui est à la fois un avantage et un inconvénient. L'avantage est évidemment la combinaison de points de vue variés et complémentaires sur les méthodes et les objectifs de l'observation transfrontalière. L'inconvénient est le coût de transaction pour se mettre d'accord sur des objectifs et des calendriers communs. Il faut signaler à cet égard le rôle de *facilitateurs* que peuvent jouer des organismes spécialisés dans l'observation transfrontalière au niveau national (i.e. la Mission Opérationnelle Transfrontalière) comme au niveau européen (le programme ESPON, en particulier à travers sa Priorité 2). Habités à gérer des négociations complexes à la fois multi-acteurs et multi-scalaires, ils peuvent permettre de concilier plus facilement les points de vue des uns et des autres.

Ainsi, l'objet de ce rapport n'est pas de recenser l'ensemble des difficultés de constitution d'un observatoire pérenne du transfrontalier, mais d'en élucider au moins les difficultés proprement méthodologiques, dans les domaines de la cartographie, la statistique, la géomatique et l'analyse spatiale. Ce travail se propose de montrer que la plupart des obstacles proprement techniques et scientifiques peuvent être levés moyennant une stratégie adaptée et la mobilisation de moyens suffisants. Mais nous ne nous prononcerons pas sur les moyens de faire adopter une telle stratégie par l'ensemble des parties prenantes de l'observation des territoires transfrontaliers.

1.2. Pérennisation et harmonisation des bases de données transfrontalières

La plupart des obstacles méthodologiques et techniques soulevés par la constitution d'une base de données transfrontalière peuvent être envisagés sous l'angle des deux objectifs complémentaires que sont **la pérennisation et l'harmonisation de la base**.

Il semble en effet qu'après une phase d'émergence des observatoires transfrontaliers régionaux à la fin des années 1980, puis une multiplication des démarches et des échelles d'observation depuis les années 2000 (MOT, 2014), l'accent porte désormais sur la nécessité d'engager la pérennisation des démarches d'observation aux frontières entre la France et ses Etats voisins. Le terme de **pérennisation** auquel il est fait référence recouvre plusieurs significations qu'il est important de distinguer et qui jouent à différents niveaux dans l'inscription de l'observation dans la durée. D'un côté, ce terme renvoie au **renforcement d'une vision politique partagée** des espaces transfrontaliers, sur laquelle peut s'appuyer un programme de travail d'observation pluriannuel. Il évoque également **l'approfondissement de la coopération entre différents acteurs**, notamment entre instituts statistiques nationaux ou régionaux, ainsi qu'une **garantie de moyens financiers et humains** suffisants. Aux côtés de ces aspects politiques, juridiques et financiers, la pérennité de l'observation dépend fondamentalement de la **manière dont sera pensée et organisée la base de données** qui constituera le socle du projet. Par exemple, cet objectif peut amener à restreindre le nombre et la nature des informations traitées, de manière à assurer plus facilement la mise à jour et le suivi de cette information ; l'exigence de traçabilité des données et de reproductibilité des méthodes employées peut orienter fortement le format des métadonnées, etc. Ces éléments, qui seront développés plus précisément dans la suite de ce rapport, sont à prendre en compte et à anticiper au cours des discussions plus générales sur le fonctionnement de l'observatoire et les moyens nécessaires à sa mise en œuvre.

Par ailleurs, quel que soit l'horizon temporel des observatoires mis en place, la question de la **mise en cohérence de données par nature hétérogènes** tout aussi fondamentale. Les travaux de la MOT-FNAU (2012) ainsi que les précédents rapports coordonnés par RIATE pour le compte de la DATAR (de Ruffray *et al.*, 2008 et 2011) ont insisté sur la nécessité de conduire, avant toute étude transfrontalière, un **travail d'harmonisation des données statistiques à une échelle fine**. En effet, des artefacts peuvent résulter aussi bien de différences dans les modes de collecte de l'information statistique entre deux pays (ou deux régions) frontaliers que de l'hétérogénéité des unités spatiales d'observation. Par exemple, sur les extraits de cartes de la Figure , il est difficile de savoir à première vue si les fortes disparités observées de part et d'autre de la frontière (ici représentée par un tracé orange) illustrent des ruptures réelles dans la répartition d'un phénomène donné ou si elles sont liées à des nuances nationales dans la définition du phénomène mesuré, à des différences de dates de recensement ou encore à des maillages hétérogènes.

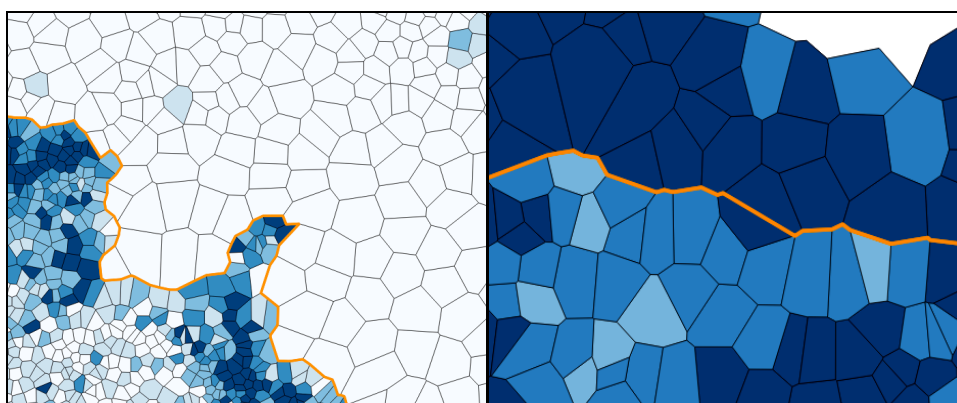


Figure 3 : Exemples de répartitions spatiales différentes de part et d'autre d'une frontière

Ainsi, l'analyse des espaces transfrontaliers repose sur des choix méthodologiques forts, portant sur la mise en cohérence des données de différents pays, selon **trois orientations** principales : traitements liés à la définition des variables (**harmonisation sémantique**), à l'échelon d'observation (**harmonisation spatiale**) et aux dates renseignées (**harmonisation temporelle**). De manière plus générale, ces questions se retrouvent

dans toute comparaison des territoires européens, à quelque échelle que ce soit. A ce titre, les réflexions menées par l'UMS RIATE et ses partenaires scientifiques sur la constitution d'une base de données européenne dans le cadre du programme ESPON (*ESPON Database 2013*), qu'elles portent sur l'adaptation de la structure des données et des métadonnées à l'interface de requête de la base, sur l'amélioration de la profondeur temporelle des données ou encore sur la comparaison d'objets géographiques complexes (ex : villes), pourraient être mobilisées pour un observatoire du transfrontalier. De plus, cette expertise méthodologique s'est accompagnée d'une réflexion de nature opérationnelle, qui a été portée par l'adaptation de l'interface d'interrogation aux attentes des acteurs de l'aménagement du territoire européen. Au final, s'il peut sembler important de rechercher des données originales pour analyser la spécificité des espaces transfrontaliers, il est essentiel d'assurer la diffusion d'indicateurs de base, disponibles à différents pas de temps et à différentes échelles géographiques (*ESPON M4D Core Database strategy, 2011*), ce qui n'a absolument rien d'évident. Une telle réalisation constituerait déjà une importante avancée, notamment si l'observation visée est exhaustive et qu'elle s'étend à l'ensemble des espaces transfrontaliers français.

Enfin, même si la question de l'harmonisation des données doit se poser à toutes les échelles spatiales, l'**échelon local** soulève des enjeux spécifiques qui n'ont été abordés qu'à la marge dans les études comparatives européennes. En dehors du problème d'accès à l'information, plus marqué pour les données locales que pour les données régionales (on pense en particulier aux données sur les flux transfrontaliers), les données locales soulèvent des enjeux particuliers dans l'analyse des variations temporelles (sensibilité plus forte aux variations extrêmes), ou encore d'interopérabilité spatiale entre les données diffusées au sein de maillages administratifs variés et des données ponctuelles fréquentes à l'échelon local (équipements, prix des logements à l'adresse, capteurs de pollution, etc.).

Face aux enjeux de pérennisation et d'harmonisation d'une base de données transfrontalière, nous proposons dans la partie suivante un bilan organisé des règles de constitution d'une base statistique et cartographique, en fonction des usages attendus de cette base.

2. Règles de bonnes pratiques pour la mise en place d'un observatoire, « volet données » : de la collecte à la diffusion des données

La mise en place d'une structure d'observation pérenne des territoires transfrontaliers passe par la constitution d'un socle solide sur lequel seront basées les études sur les territoires. Ce socle sera donc une base de données intégrée et complète portant sur ces espaces. Le terme « base de données » est ici entendu au sens large de *container* permettant de stocker et de retrouver un ensemble d'informations de nature hétérogène : données statistiques, méthodes, sources, fonds de carte.

La constitution d'une telle base repose sur un ensemble de choix et de principes qui relèvent aussi bien des sphères thématiques, méthodologiques, techniques qu'organisationnelles. Elle nécessite de prendre en considération, dans la mesure du possible, les points de vue de plusieurs acteurs : ceux qui assurent le pilotage stratégique de l'observatoire (qu'on désignera par la suite par le terme de « comité stratégique »), ceux qui sont chargés de son pilotage technique (« comité technique »), ainsi que les utilisateurs potentiels de l'interface ainsi créée.

Les fiches qui suivent présentent de manière synthétique les aspects qui nous semblent les plus importants pour guider la mise en place d'un tel système d'information. Pour en simplifier la présentation, nous avons identifié 6 étapes principales, étant entendu que de nombreux allers-et-retours sont en pratique nécessaires entre ces différentes étapes (Figure 3).

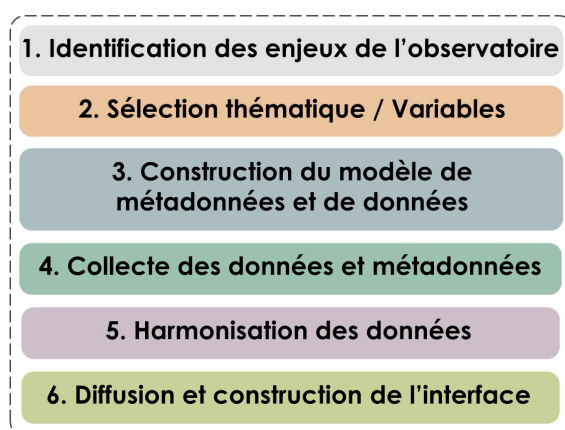


Figure 3 : Etapes de construction de l'observatoire

La plupart des exemples illustrant les étapes 3 et 4 s'inspirent d'un travail d'un mois confié à un étudiant de Master 1 (profil d'ingénieur d'études spécialisé dans la collecte et le traitement de l'information géographique). L'objectif de ce travail était de tester la faisabilité d'une collecte des données locales sur l'emploi, sur la portion nord-est de l'espace transfrontalier français, dans le cadre d'un scénario centralisé, une seule personne étant chargée de la collecte, et autonome, aucun intermédiaire institutionnel ou scientifique n'étant censé faciliter la collecte.

Fiche 1 - Identification des enjeux de l'observatoire, en fonction des usages et des moyens

1.1. Pour quoi ? Pour qui ?

- La création d'une base de données n'est pas une fin en soi, elle découle d'un **besoin d'information structurée autour d'une problématique donnée**. Si la problématique semble ici clairement définie (l'étude des territoires transfrontaliers), les **usages et les usagers de la base de données** et plus généralement de l'observatoire doivent être identifiés en amont de tout travail de collecte et même de conception/modélisation. De la nature de ces usages dépendront un certain nombre de choix conceptuels, méthodologiques et techniques, qui concernent aussi bien les phases de conception, de collecte, de maintenance que de valorisation et de mise à disposition des informations.
- Par exemple, si l'on accorde un poids important aux requêtes que les **acteurs politiques locaux** effectueront sur les caractéristiques précises de « leur » territoire, on évitera de choisir un fond de carte présentant un fort degré de généralisation des limites administratives ou un découpage géométrique de l'espace (carroyages). Cela pèsera aussi sur le choix de diffuser plutôt des données officielles, qui sont hétérogènes mais reflètent très précisément l'état du territoire local, ou bien des données corrigées par harmonisation (sémantique, temporelle et/ou spatiale), qui assurent une plus grande robustesse scientifique des comparaisons transfrontalières (*cf Fiche 5*).
- De manière plus générale, **plusieurs scénarios d'usages de la base de données** peuvent être proposés (*cf Fiche 6*), depuis la possibilité pour un analyste d'effectuer des requêtes, jusqu'à la mise en ligne de la base de données sur une application cartographique, en passant par l'adjonction de modules d'analyse adaptés aux problématiques transfrontalières (ex : module permettant de simuler une harmonisation de données qui compense les différences de définition d'une variable de part et d'autre de la frontière).

1.2. Avec quels moyens ? Selon quel mode de gestion ?

- La question de l'importance des **moyens humains et techniques** engagés intervient de manière évidente dans la **capacité de stocker (et de mettre à jour) une masse plus ou moins importante de données** ; les compétences requises ne seront pas non plus les mêmes selon que la restitution des données se fera sous la forme d'une application cartographique ou qu'elle ira jusqu'à l'adjonction de modules d'analyse spatiale permettant d'améliorer l'harmonisation des données (*cf Fiche 6*).
- Au-delà de ces aspects quantitatifs, c'est aussi la **question fondamentale du mode d'organisation** plus ou moins décentralisé du système d'information qui est en jeu, notamment pour toute la phase de collecte des données (*cf Fiche 6*) : dans le cadre d'une **organisation centralisée**, un opérateur pourra à lui seul récupérer toutes les données dans tous les pays en passant par les instituts statistiques. Suivant un **mode décentralisé**, la collecte des données des pays frontaliers peut être déléguée à des tierces personnes (*via* des conventions) dans ces mêmes pays, avec un coût supplémentaire d'apprentissage et de validation des opérations déléguées.

La définition de ces orientations thématiques et organisationnelles relève *a priori* du pilotage stratégique de l'observatoire, en interaction avec le comité technique.

Fiche 2 - Sélection des variables, des maillages pertinents d'observation, de l'échelon de collecte et du fond de carte

Sélection thématique / Variables	
Thème ?	Echelon géographique ?
Définition, limites, nb de variables associées, choix "emblématiques"...	Maillage pertinent d'observation

2.1. Sélection des variables

Les variables choisies dépendent bien entendu de la problématique abordée par l'observatoire. Mais il faut prendre en considération un certain nombre de **facteurs pratiques qui doivent intervenir dans le choix des variables** :

- la **disponibilité des données**, voire leur existence même (un PIB local pourrait être intéressant mais il n'existe tout simplement pas). De plus dans un cadre transfrontalier il faudra faire attention à la disponibilité des données dans les pays concernés par l'étude (MOT-FNAU, 2012)
- la **pérennité des données** invite à préférer les données qui sont mises à jour régulièrement. (indicateurs disponibles en séries temporelles sur les sites nationaux).
- le **nombre de variables** sélectionnées joue non seulement sur les coûts de constitution et de maintenance de la base, mais aussi sur la capacité à assurer un suivi temporel des variables (une base de données contenant 100 variables mises à jour annuellement sera beaucoup plus lourde à maintenir et à mettre à jour qu'une base n'en contenant qu'une vingtaine et mise à jour tous les 3-4 ans).
- la **nature statistique des variables collectées**. Il sera toujours préférable de **stocker des données brutes de stock** plutôt que leur transformation plus ou moins poussée en variables composites qui pourraient être recalculées dynamiquement par la suite (Figure 4). Par exemple, si l'on travaille sur l'emploi il sera plus intéressant de stocker les populations actives, employées et à la recherche d'un emploi plutôt que les taux de chômage et d'emploi qui peuvent être rapidement déduits des variables précitées. Cela permettra de conserver moins de variables et de faciliter leurs mises à jour voire leurs corrections.

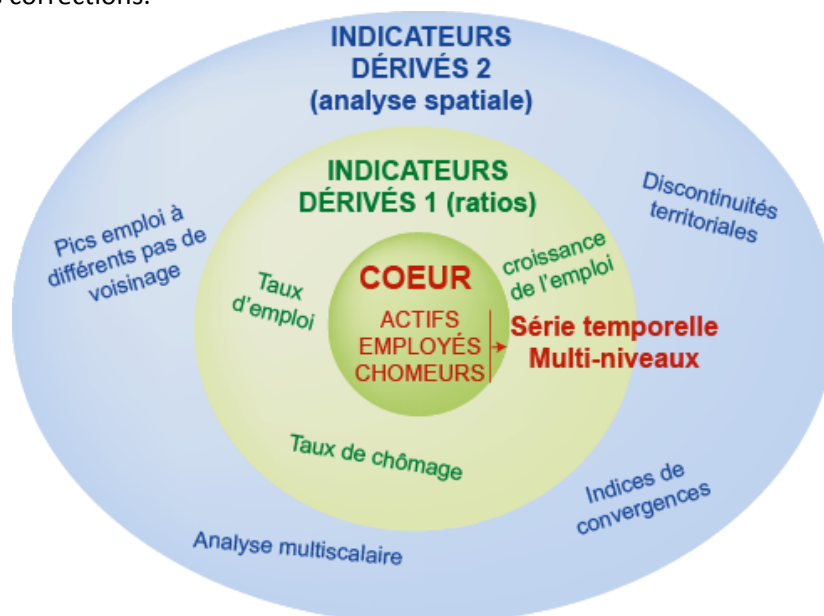


Figure 4 : Variables de base brutes, variables composites, variables élaborées intégrant des traitements spatiaux (d'après ESPON M4D, 2011)

Ces choix devront être réajustés en fonction des résultats de la collecte des métadonnées (les données sont-elles disponibles ? les définitions sont-elles comparables ?) et des données (données complètes ?) (Fiches 3 et 4).

2.2. Choix de la maille pertinente d'observation et de l'échelon de collecte

- Le **choix de la maille pertinente** pour l'étude relève aussi bien de l'échelon d'observation adapté à l'analyse territoriale que du type de maillage choisi (statistique, administratif, fonctionnel).
- Le **choix de l'échelon de collecte** des variables, pour des raisons un peu similaires à celles évoquées ci-dessus à propos des données de stock, doit être **celui du recueil des données**, sans agrégation, au niveau des « briques élémentaires » de l'information. Cela permet de **réaliser a posteriori les opérations d'agrégation spatiale** qui seront **pertinentes en fonction des échelles d'analyse choisies** (par exemple les zones d'emploi pour la thématique sur l'emploi, les agglomérations pour les études foncières...).
- Le choix du **périmètre d'analyse** (définition de la bande transfrontalière) permet de déterminer les États dont les données doivent être collectées et conditionne les traitements statistiques et cartographiques. Mais dans la phase de collecte, il est **préférable de tout stocker**, à l'échelon de l'ensemble des États ou des régions (dans le cas des États fédéraux).

2.3. Choix du fond de carte

- La question du fond de carte doit être traitée à partir du moment où les **niveaux de maillages de collecte** ont été choisis (choix des unités cartographiques de base associées aux données).
- Mais elle dépend aussi étroitement de la **question des usages (Fiche 1)** : selon que les analyses porteront sur un niveau Européen, Français, régional ou local, le choix des niveaux de généralisation du contour des unités territoriales ne sera pas le même.

Dans la perspective de maintenance et de suivi temporel de la base, il faudra prévoir un retour sur cette étape, en fonction de l'**évolution des mailles élémentaires** de diffusion des données (maillage évoluant dans le temps).

Fiche 3 – Modèle de métadonnées et modèle de données

Construction du modèle de métadonnées et de données	
Métadonnées /pays	Modèle de données
Description variables, géométries, couverture spatiale et temporelle, méthode de collecte, sources...	Types d'objets géographiques (découpages stats, rasters, POI...) et relations entre ces objets

La phase de conception du modèle de métadonnées et de données est une étape indispensable pour assurer la pérennité de la base, avant de s'engager dans la collecte.

3.1. Conception du modèle de métadonnées.

La pierre angulaire d'une gestion pérenne du flux d'information à l'échelle d'un observatoire repose sur la définition d'un modèle de données et de métadonnées adapté. Ce modèle sert tout d'abord à renseigner précisément les données qui collectées afin de les rendre intelligibles à un large public et aussi à réutiliser ces informations afin d'en organiser leur restitution. Une fois le modèle défini, la complétion des métadonnées implique méthode, méticulosité et rigueur. .

- La conception du modèle de métadonnées **s'inscrit dans un processus dont dépendent plusieurs paramètres fondamentaux** : identification des besoins, construction du modèle, collecte des informations, harmonisation des définitions, objectifs de dissémination, etc. Des standards adaptés à la description et la diffusion de données géographiques aident considérablement à la structuration de ce type d'information (directive INSPIRE, normes ISO-19115 et ISO-135). Notons également l'existence de groupes de travail proposant axes de réflexion permettant d'assurer la qualité des métadonnées. On peut citer par exemple le groupe de pilotage européen sur les métadonnées statistiques (METIS)¹, qui propose des méthodes de travail et un guide de bonnes pratiques pour sensibiliser à ces questions (Figure 5).

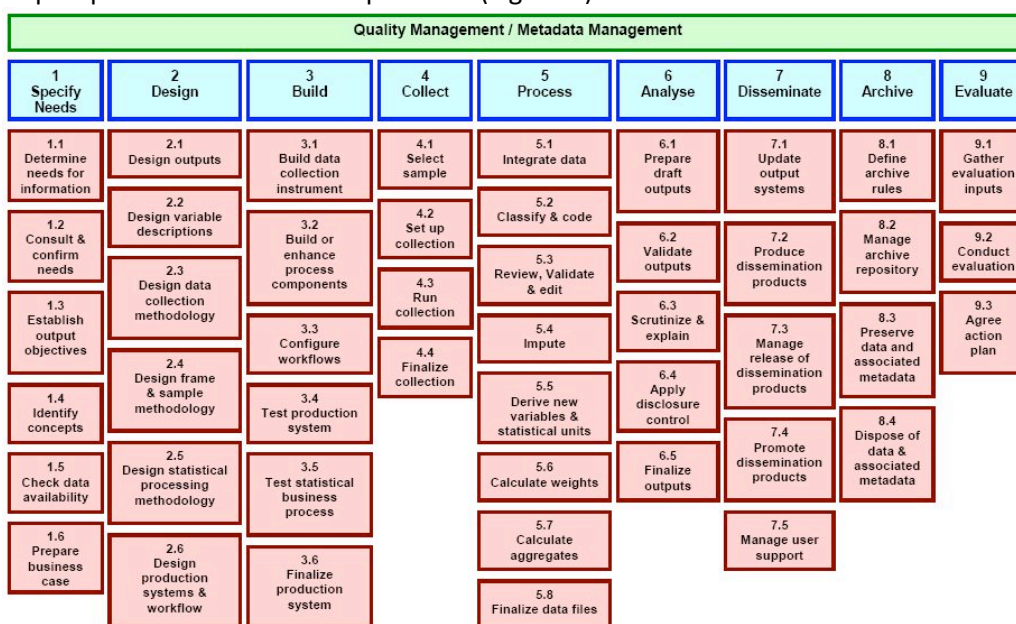


Figure 5 : Un modèle de gestion de métadonnées : le GSBPM (Generic Statistical Business Process Model)

- Néanmoins, il n'existe pas de « procédure type » concernant la mise en place d'un modèle de métadonnées. La conception du modèle de métadonnées est un compromis entre les standards internationaux en vigueur à respecter, les pratiques que l'on souhaite mettre en place dans le cadre de l'observatoire pour décrire des données hétérogènes (champs utilisés pour interroger la base de données *a posteriori* par exemple) et le degré de précision de description de l'information souhaité. Étant entendu que le nombre de champs intégrés dans un modèle de métadonnées est proportionnel au temps que cela nécessite pour les renseigner.

¹ <http://www1.unece.org/stat/platform/display/metis/METIS-wiki>

2. Règles de bonnes pratiques pour la mise en place d'un observatoire

- Du point de vue des ingénieurs en charge de la collecte des données et de leur diffusion, la conception des métadonnées est fondamentale pour **assurer la traçabilité des sources et des méthodes**. De plus, il s'agit d'une pièce centrale de la construction d'une interface de consultation/extraction interactive des données, car ce sont surtout les champs des métadonnées qui permettront **l'interrogation de la base de données** (choix de l'année ou de la période, choix du niveau géographique, choix de la variable en fonction de sa description précise...). Le nombre des champs renseignés, le choix des mots-clés associés aux données (mots-clés libres ou thesaurus), sont le résultat d'un compromis : des champs plus complets permettront d'améliorer la précision et la souplesse des requêtes, mais augmenteront d'autant plus le temps de collecte des métadonnées.
- Du point de vue des ingénieurs et des thématiciens chargés du traitement et de l'analyse des données, cette étape est nécessaire pour établir un **bilan de la comparabilité des données** et construire une documentation sur l'hétérogénéité sémantique et temporelle des variables, à l'issue de la collecte. Les métadonnées sont en ce sens une **pièce centrale de l'harmonisation de l'observation**, permettant d'utiliser des données hétérogènes en connaissance de cause.

Le modèle de métadonnées utilisé pour la base de données du programme ESPON est en ce sens un exemple intéressant d'adaptation des standards internationaux (le modèle est compatible INSPIRE) à un usage de la donnée qui est faite à l'échelle de ce programme de recherche appliquée². Dans ce cadre, un modèle au format tabulaire (Excel) a été mis en place. Il est constitué de trois modules : informations relatives au jeu de données, aux indicateurs et à l'origine des données. Dans le cadre de l'expérience menée pour cette étude, un modèle opérationnel a été adapté pour collecter les informations statistiques (Figure 6)³.

1	Indicator Identification								
2	Code	POP_POP							
3	Name	Population							
4	Abstract	Population							
5	Methodology	RP2009 exploitation principale							
6	Temporal Extent	2009							
7	Unit of measure	nombre d'habitants							
8	Difficulty	pas de difficultés particulières							
9	Indicator Identification								
10	Code	POP_POP							
11	Name	Population							
12	Abstract	Population							
13	Methodology	RP1999 exploitation principale							
14	Temporal Extent	1999							
15	Unit of measure	nombre d'habitants							
16	Difficulty	pas de difficultés particulières							
17	Indicator Identification								
18	Code	SUPERF							
19	Name	Superficie (en km²)							
20	Abstract	Superficie (en km²)							

1	Name	POP_POP	POP_POP	SUPERF	NA159909	DCE9909	POP_POP0204	POP_POP1529	POP_POP1529
2	L'habitat-dimentionnel	787	725	15,95	95	53	180,6	1	97,9
3	L'habitat-dimentionnel	207	167	9,15	36	23	38,9	1	31,0
4	Amboise-en-Bugey	13 350	11 432	24,6	1766	959	2734,4	1	2828,0
5	Amboise-en-Dombes	1 592	1 407	15,92	182	88	328,1	1	291,6
6	Amboise	120	86	5,88	8	10	12,9	1	19,3
7	Ambronnay	2 328	2 144	33,55	288	130	479,0	1	368,0
8	Ambrus	690	596	5,22	72	33	128,6	1	124,6
9	Anden-en-Cordon	336	275	6,94	37	33	60,9	1	43,6
10	Angleton	960	772	29,26	115	89	213,3	1	147,4
11	Apvemoort	352	328	14,57	47	15	72,9	1	55,0
12	Araso	297	275	21,65	31	33	36,3	1	37,4
13	Arandis	195	135	34,1	17	27	32,4	1	17,8
14	Arden	3 493	3 623	23,49	434	245	658,0	1	642,0
15	Ardegnay	471	431	13,07	41	45	84,6	1	48,8
16	Ardegnay	383	330	17,47	41	50	82,6	1	61,4
17	Arnis	421	389	7,84	50	48	73,0	1	63,0
18	Arns	22	15	6,82	0	3	3,3	1	0,0
19	Arre-sur-Fornans	1 528	1 100	5,5	135	53	305,7	1	229,3
20	Arre-sur-Fornans	1 112	970	3,75	98	109	197,9	1	146,5

Figure 6 : Modèle de métadonnées proposé = modèle simplifié d'ESPN DB

- Il existe plusieurs solutions pour implémenter et renseigner ce modèle : Un fichier auquel est associée la donnée (solution à moindre frais, simple d'utilisation). La création d'un éditeur de métadonnées est une autre possibilité, mais implique de plus en plus de coûts de développement plus importants.

² Ce modèle a été conçu par le projet ESPON M4D Multi Dimensional Database Design and Development. Un des enjeux de ce projet consiste à gérer la profusion de données hétérogènes, les normaliser et les harmoniser, évaluer leur qualité et les restituer pour un large public à partir d'une interface Web dédiée.

³ ESPON data and metadata specifications, disponible ici (dernière visite : 15 Mai 2014) : <http://database.espon.eu/db2/metaspécifs?idDoc=61>

3.2. Conception du système d'information

Si l'on souhaite mettre en place une interface de requête des informations contenues en base (qu'elle soit cartographique ou non), cela nécessite la mise en place d'un système d'information dédié. La conception de ce système d'information doit être réalisée par étapes et nécessite des compétences spécifiques (gestion de projet informatique, développement). Des méthodes de développement existent (MERISE par exemple) afin de guider le développement de tels projets.



Figure 7 : Étapes nécessaires à la mise en place d'un système d'information automatisé⁴

Pour résumer succinctement la méthode, la première étape consiste à définir les besoins du système d'information : quelles futures requêtes seront souhaitées, quelles sont les informations nécessaires pour réaliser ces traitements ? L'expression de ces besoins qui fait généralement l'objet de spécifications précises permet dans un second temps de définir un **MCC** (Modèle conceptuel de communication) définissant les flux d'informations à prendre en compte. L'étape suivante consiste à mettre en place un MCD (modèle conceptuel de données) et un **MCT** (modèle conceptuel des traitements) décrivant les règles à prendre en compte pour réaliser les traitements (relations, contraintes entre objets etc.) Le modèle logique représente un choix logiciel pour le système d'information. Le modèle physique reflète un choix matériel pour le système d'information.

- La conception d'un modèle de données repose à la fois sur l'identification des **types/classes d'objets géographiques** intégrés à la base (mailles administratives à différents échelons géographiques, mais aussi modes d'occupation du sol, infrastructures routières représentés par des lignes, équipements divers représentés par des points...) et sur la définition des **relations géométriques** entre ces différents types d'objets.
- Ce modèle conceptuel de données (MCD) est nécessaire pour anticiper sur les **requêtes spatiales des utilisateurs**, requêtes qui s'appuieront sur le **croisement entre ces objets** et sur leur **degré d'interopérabilité** géométrique et thématique. Le MCD repose donc sur un certain nombre de choix qui dépendent des usages attendus de la base : par exemple, si l'on souhaite travailler sur l'accessibilité aux emplois, aura-t-on besoin de stocker l'objet réseau routier (très compliqué), ou pourra-t-on se contenter d'associer une matrice de distances routières entre les unités administratives ? Si l'on s'intéresse à des données ponctuelles (par exemple les prix fonciers à l'adresse), doit-on les stocker telles quelles, ou bien peut-on les stocker sous la forme de carroyages ?
- S'agissant de l'organisation des unités territoriales de base, le modèle doit aussi prévoir l'**intégration de plusieurs versions de nomenclatures** spatiales. Les redécoupages territoriaux sont en effet fréquents à l'échelle locale (cf partie 2). Il est donc important de concevoir le système d'information à partir d'un MCD évolutif et ouvert à différents types d'objets géographiques, de manière prospective.

⁴ Plus d'informations sur la conception d'un système d'informations :

<http://www.commentcamarche.net/contents/655-merise-initiation-a-la-conception-de-systemes-d-information>

Fiche 4 – Collecte des métadonnées et des données (statistiques et cartographiques)



4.1. Modèle d'organisation de la collecte, qui fait quoi ?

De même que les étapes décrites dans les fiches 2 à 3 peuvent être conduites de manière collaborative (comité stratégique, comité technique, individus...) ou de façon très centralisée, la collecte des données, une fois mis au point un protocole et un cadre rigoureux, peut être faite selon ces deux modalités.

- Un **opérateur peut à lui seul** récupérer toutes les données dans tous les pays en passant par les sites web des instituts ou des relations qu'il aura pu créer avec les instituts. Cependant, la création d'un protocole de collecte contraignant et bien documenté permet une certaine **décentralisation de la collecte**. Ainsi une solution consistant à déléguer la collecte des données des pays frontaliers à des contacts (*via* des conventions) dans ces mêmes pays est tout à fait envisageable. Le format de collecte de données et de métadonnées est défini en amont ainsi que sa documentation détaillée. Ensuite la collecte elle-même peut être le fait de n'importe quel agent ayant été formé à la saisie de ce format particulier.
- L'implication des partenaires pourra aller du **partage de carnet d'adresses** à la **prise en charge de la collecte à proprement parler** sur les bases de la documentation réalisée.
- Dans le cadre de cet observatoire, un certain nombre de **contacts et de réseaux** existent déjà entre des acteurs français (MOT, DATAR, FNAU, INSEE...) et les acteurs équivalents dans les pays frontaliers. Le protocole de collecte des données devra a priori s'appuyer, à un degré qu'il reste à définir, sur ce réseau préexistant, qu'il faudrait renforcer. D'autres relais (réseaux politiques ou scientifiques) peuvent être activés pour avoir accès aux données.

Le modèle d'organisation de la collecte est à définir précisément au début de la mise en place de l'observatoire transfrontalier. Il est important que les partenaires sollicités, quel que soit le niveau de sollicitation, aient bien compris ce que représente la collaboration qui leur est demandée, en temps de travail et en qualification nécessaire.

4.2. Identification des sources et localisation des données

- Les difficultés rencontrées pour accéder aux données et à l'information sur les données (métadonnées) se multiplient proportionnellement à la **quantité des producteurs de l'information**, de part et d'autre de la frontière. Au-delà du nombre de systèmes statistiques ciblés, on se confronte également à différents **modèles de systèmes producteurs de données hétérogènes**, selon qu'il s'agit de systèmes centralisés (instituts statistiques nationaux) ou décentralisés au niveau régional : en Allemagne par exemple, la plupart des données locales doivent être recherchées depuis les instituts statistiques régionaux (Länder). D'autre part, ces données peuvent être diffusées par des **instituts généraux ou spécialisés** (ministères). Si en France, l'INSEE diffuse des données élémentaires pour l'ensemble des communes françaises, ce n'est pour l'ensemble de ses États frontaliers. En Belgique par exemple, les données détaillées sur les taux d'activité par sexe et par classe d'âge sont diffusées par le Centre de recherche sur le travail et l'économie sociale (Steunpunt WSE), dépendant du gouvernement flamand.
- Si des tentatives utiles existent pour référencer les informations rendues disponibles par les producteurs de données nationaux (rapport MOT-FNAU, 2012, ESPON « Data Navigator⁵ », 2013), l'expérience montre néanmoins que rien ne remplace un appui actif d'individus ou d'organisations nationales, qui disposent d'une connaissance experte portant sur les données disponibles, leur

⁵ <http://datanavigator.espon.eu/>

périodicité de mise à jour, leur qualité ou la méthodologie employées à un niveau national⁶. En effet, en dépit de l'existence de ces outils, la recherche des données appropriées reste une étape chronophage du fait de la complexité de l'organisation des sites internet (compliquée parfois par l'obstacle de la langue). La recherche de données sur la population active a pris environ 10 minutes pour la France, mais 1 heure pour chacun des autres Etats (Luxembourg, Belgique, Allemagne).

- Les appuis nationaux à la collecte d'information sont donc absolument nécessaires, et ce d'autant plus lorsqu'il s'agit d'informations locales qui ne font pas l'objet de contrainte de collecte ou d'harmonisation de la part d'Eurostat.

4.3. Collecte des données statistiques

- A l'issue de la collecte des métadonnées, la confrontation des différentes définitions nationales (qui suppose de maîtriser la langue...) peut amener à restreindre le choix des variables sélectionnées (*Fiche 2*) et donc de la collecte.
- Quelques **principes généraux guident la structure de la collecte**. D'une part, il est préférable d'effectuer une **collecte au niveau géographique le plus fin disponible**. D'autre part, cette collecte doit être effectuée **par pays** et d'associer les fichiers nationaux dans un second temps seulement : cette étape est importante pour récupérer les définitions parfois différentes des variables. La mise à jour en sera de plus facilitée. Enfin, quelles que soient les corrections effectuées sur les fichiers récupérés, il est important de garder une trace des fichiers transformés et d'**archiver l'ensemble des fichiers intermédiaires**, depuis la collecte primaire.
- Dans cas d'une organisation centralisée et autonome, un des principaux obstacles à la collecte des données statistiques est la **diversité des formats de diffusion des données**, qui se prêtent plus ou moins bien à la collecte (fichier à télécharger en format tableur ou PDF, page HTML interrogeable à partir de requêtes individuelles et d'un menu déroulant...). Des **solutions d'automatisation** peuvent être élaborées afin de récupérer ces données et de les réorganiser dans un format prédéfini. Dans le cadre d'une expérience pour cette étude, le développement d'un programme (en Python) nous a ainsi permis de lancer des requêtes à partir des codes des unités statistiques et de récupérer les données d'un site au sein d'un tableur (en l'occurrence, les données de population du site du Bade-Württemberg, statistik-bw.de).

4.4. Critique des données, pré-traitements, harmonisation statistique

- Il n'est pas rare que les fichiers issus de la collecte contiennent des données manquantes et/ou incohérentes. Il s'agit parfois de données exceptionnelles associées à des cas particulier (ex : le Liechtenstein par rapport aux autres régions européennes), et qui n'appellent pas de correction. Mais ces valeurs exceptionnelles peuvent aussi correspondre à des valeurs aberrantes, liées à des **incohérences de nature sémantique** (population de « -2100 » habitants), **temporelle** ou **spatiale** (déviation locale exceptionnelle sur une carte), du fait d'erreurs de saisie, de changement de définition d'un indicateur, ou encore d'estimations mal ajustées (entre deux recensements par exemple).
- L'identification des **incohérences** est une étape essentielle du prétraitement des données. Plusieurs démarches peuvent être mises en œuvre : de l'analyse élémentaire des **distributions statistiques**, notamment à partir des **valeurs extrêmes**, à l'application de méthodes plus sophistiquées basées sur des **modèles spatiaux ou temporels** (ESPON M4D, 2012). D'autre part, l'introduction de **variables de contrôle** permet de vérifier la **compatibilité entre niveaux hiérarchiques** (en comparant par exemple les résultats diffusés directement à l'échelon des NUTS3 avec les résultats

⁶ A ce titre, nous tenons spécialement à remercier l'Institut de Gestion de l'Environnement et de l'Aménagement du Territoire (IGEAT, Bruxelles) et particulièrement Xavier May et Pablo Medina Lockhart, dont les conseils concernant les spécificités des données belges se sont avérés précieux pour alimenter la troisième partie de ce rapport.

issus de l'agrégation des LAU2 appartenant au même NUTS3), ainsi que la **compatibilité entre catégories statistiques plus ou moins agrégées** (par exemple somme des classes d'âge par rapport à la population totale).

- Les données manquantes et incohérentes sont ensuite estimées à partir de modèles statistiques, spatiaux et/ou temporels. Dans tous les cas, il est important de **conserver deux types de fichiers** : les données officielles et les données modélisées.

4.5. La collecte des fonds de carte

Le choix du fond de carte dépend des moyens financiers et humains (temps de saisie) mis en œuvre, des conditions juridiques d'accès aux données ainsi que de la qualité de la visualisation. Dans le cas des espaces transfrontaliers, trois types de solutions peuvent être envisagés (Figure 8) :

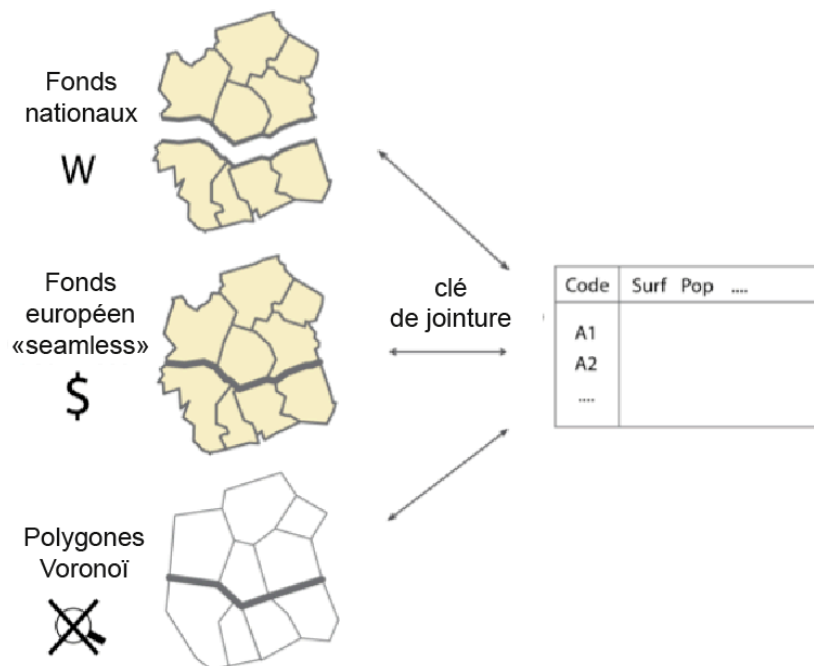


Figure 8 : Choix du fond de carte des unités spatiales locales et jointure avec les données statistiques

- **Instituts cartographiques officiels** : la solution d'une collecte effectuée auprès des instituts statistiques nationaux pose le problème de la concordance des géométries des unités spatiales aux frontières des Etats, et de la diversité de systèmes de projection nationaux. Le fond européen des communes fourni par Eurogeographics⁷ (EuroBoundaryMap) répond à ce problème en offrant une représentation « sans couture » (*seamless*) aux frontières, mais il s'agit d'un fond très onéreux, qui s'accompagne de droits restreints. Le niveau de généralisation très poussé de ce fond doit parfois être modifié, en fonction de l'échelle de représentation (Zanin *et al.*, 2013).
- **Données « libres » et harmonisées à l'échelle européenne** : s'il existe plusieurs ressources de fonds cartographiques libres, les découpages ne sont pas toujours renseignés à l'échelon local (GADM) ou ne sont pas toujours structurés de manière adéquate (dans *Open Street Map*, les différents niveaux hiérarchiques des découpages –LAU2, LAU1, NUTS3...- sont mélangés au sein d'un même fichier). Une solution particulièrement intéressante consiste à utiliser le **fond LAU2 libre de droit créé par J. Michelet (Université de Genève)**, qui est construit à partir des centroïdes des LAU2 et de la génération de polygones de Thiessen. Ce fond peut être obtenu par l'intermédiaire d'une convention avec l'Université de Genève. La jointure avec les données statistiques des instituts nationaux s'effectue via le code national des unités spatiales et la mise à jour de tels fonds est aisée. En revanche, la généralisation très poussée des contours des mailles administratives peut être jugée peu adaptée aux représentations à l'échelle locale.

⁷ Eurogeographics rassemble l'équivalent des IGN nationaux de l'ensemble des États Membres de l'Union Européenne ainsi que plusieurs pays du voisinage oriental européen.

- **Le fond cartographique peut être créé en interne**, à partir de la digitalisation des fonds de carte nationaux et européens collectés. C'est la solution qui a été adoptée par l'UMS RIATE, pour les découpages en NUTS 2 et en NUTS3 (Figure 9). Malgré le travail non négligeable que cela nécessite, cette solution présente des avantages évidents en termes de droits d'exploitation (aucune licence à acquitter, possibilités de diffusion multiples etc.).

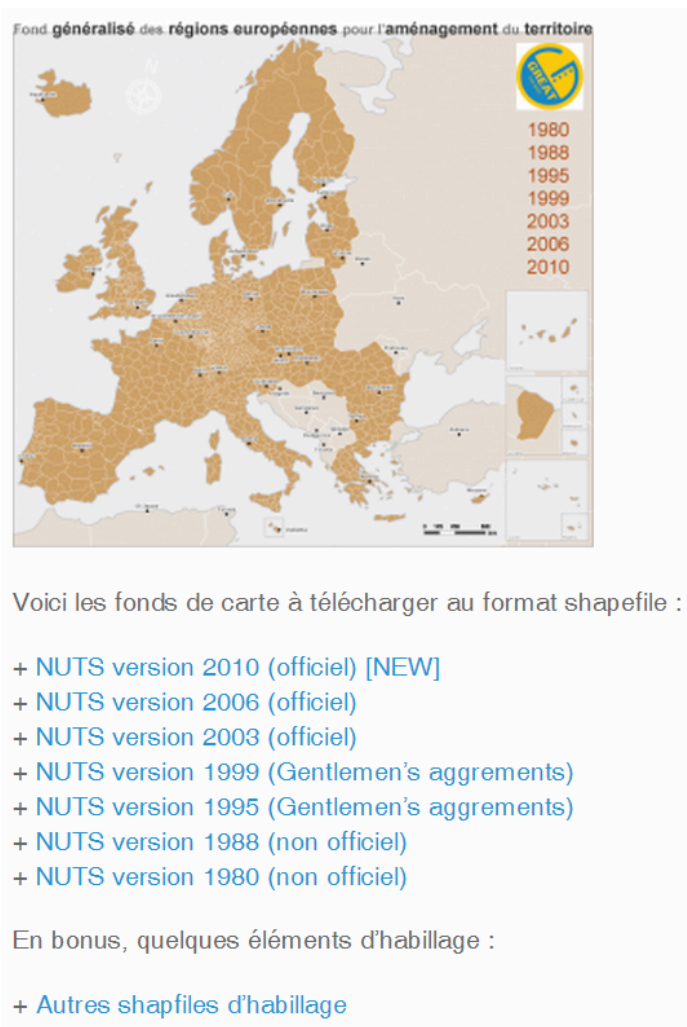


Figure 9 : Le fond généralisé des régions européennes pour l'aménagement du territoire. Des fonds de cartes géoréférencés en accès libre depuis le site de l'UMS RIATE (licence Creative Commons)

- Deux principaux **pré-traitements cartographiques** accompagnent la phase de collecte : d'une part, la vérification des liens entre les codes cartographiques et les codes statistiques des unités spatiales et la création de **dictionnaires de correspondance spatiale** (voire temporelle, entre différentes versions). D'autre part, l'ajout d'une **clé d'agrégation** permet d'associer des unités locales à des échelons géographiques supérieurs, comme le propose par exemple l'INSEE pour les zones d'emploi ou les bassins de vie⁸.

⁸ <http://www.insee.fr/fr/methodes/default.asp?page=zonages/liste-zonages.htm>

Fiche 5 – Harmonisation des données entre les pays

L'harmonisation des données entre les pays est une étape fondamentale de la mise en cohérence de données internationales hétérogènes. Ces traitements supposent a priori de bien maîtriser les outils statistiques et d'analyse spatiale.

5.1. Harmonisation sémantique

Si la définition des variables collectées par les instituts statistiques fait l'objet d'un effort d'harmonisation croissant à l'échelle européenne, il subsiste des **différences sémantiques plus ou moins grandes, révélées par l'étude des métadonnées** qui précisent les méthodologies employées. La population active, par exemple, ne sera pas définie de la même manière selon la catégorie de chômeurs prise en compte (de Ruffray *et al.*, 2011). Le dénombrement des emplois varie selon que certaines catégories d'emploi (emploi non salarié, emploi public) sont ou non prises en compte (MOT-FNAU, 2012)⁹. Dans d'autres cas, ce sont les bornes des catégories statistiques qui diffèrent, avec des classes d'âge de la population active qui varient en fonction de la législation sur le travail (en France, par exemple, l'école est obligatoire jusqu'à 16 ans, alors que l'âge limite est de 18 ans en Belgique). D'autres problèmes relèvent de l'application de méthodes différentes de recensement : en Allemagne, par exemple, les données locales sur les actifs sont à manipuler avec précaution puisque leur mesure s'effectue au moyen d'un micro-recensement représentatif seulement au niveau du Land (il porte sur 1% de la population de chaque Land) (de Ruffray *et al.*, 2011).

La gestion de cette hétérogénéité sémantique va **de l'affichage des différences de définitions, a minima** (cf. démarche suivie pour les cartes de la partie 4), jusqu'à **l'alignement des définitions sur des catégories plus homogènes**, lorsque cela est possible (par exemple, reconstitution de classes d'âge plus proches à l'aide de niveaux désagrégés de l'information).

5.2. Harmonisation spatiale

Même lorsque la variable collectée a exactement la même définition dans deux pays frontaliers, les phénomènes décrits peuvent ne pas être directement comparables, du fait de l'hétérogénéité des découpages géographiques. Les effets de cette hétérogénéité sur les phénomènes observés ont déjà été largement documentés à travers l'analyse du *MAUP (Modifiable Area Problem Unit)* (Openshaw, Taylor, 1979 ; ESPON 3.4.3. MAUP, 2006) et plusieurs solutions ont été proposées pour s'en affranchir. Rappelons seulement que les solutions proposées ne sont pas tout à fait du même registre selon qu'est considérée **l'hétérogénéité de la maille administrative** (par exemple, les LAU2) ou bien **l'hétérogénéité de zonages fonctionnels** plus complexes (par exemple des bassins d'emploi, des aires urbaines fonctionnelles, etc.).

Dans le premier cas, il faudra soit trouver un **niveau administratif de même gabarit** (par exemple les LAU2 Belges -communes- et les LAU1 Français -cantons-)¹⁰, soit **recréer un découpage homogène sur toute la zone** (carroyage) ou encore **une représentation lissée** (lissage cartographique par potentiels, selon un voisinage ; Grasland *et al.*, 2010). Dans le second cas, si on juxtapose les définitions nationales des bassins d'emploi ou des villes, il est essentiel de **déconstruire ces objets géographiques** qui sont des agrégats

⁹ « A titre d'exemple, pour l'indicateur relatif à la densité des emplois, malgré tout traité, l'obstacle majeur a été la comparaison de la donnée emploi, provenant côté français de l'UNEDIC et correspondant à l'emploi salarié privé, alors que dans certains pays voisins, cette distinction n'opérait pas et certains emplois publics faisaient partie du volume identifié dans ces zones frontalières étrangères, faisant apparaître une plus forte homogénéité dans la densité d'emploi entre les zones urbaines et les zones rurales frontalières de ces pays, qu'en France, où les zones rurales, dépourvues de leurs emplois publics, étaient particulièrement peu densément fournies en emploi. » (MOT-FNAU, 2012)

¹⁰ Cette solution ne permet pas pour autant de résoudre le problème de l'hétérogénéité spatiale interne aux Etats, qui est particulièrement forte lorsque les villes bénéficient, comme en Allemagne, d'un statut administratif à part (différence entre les superficies des *Landkreiss* et des *Stadt/Gemeinden*).

complexes de mailles administratives élémentaires, en **explicitant le modèle conceptuel et le modèle d'implémentation** sous-jacents, afin de prendre la mesure des différences de définitions nationales. Par exemple, pour les aires urbaines fonctionnelles, le modèle conceptuel, qui reflète la définition de la ville, peut renvoyer soit à la polarisation d'un bassin d'emploi par un centre, à partir de navettes domicile-travail, soit à la zone d'accessibilité à l'emploi dans des limites temporelles correspondant au budget-temps maximal des ménages. Le modèle d'implémentation correspond aux étapes nécessaires à la construction de ces objets géographiques, incluant le choix des paramètres (Bretagnolle *et al.*, 2014). D'autre part, la **création d'une délimitation harmonisée** de découpages fonctionnels, à partir d'un concept commun et de spécifications identiques (démarche descendante, *top-down*), a donné lieu à plusieurs bases de données récentes pour les villes (Bretagnolle *et al.*, 2014) et commence à être explorée pour les bassins d'emploi (Coombes *et al.*, 2012, cf. partie 3).

5.3. Harmonisation temporelle

La création de séries temporelles nécessite un travail d'harmonisation portant sur le suivi des variables descriptives comme sur les géométries des unités spatiales. L'harmonisation temporelle des données attributaires renvoie à des problèmes bien connus de **différences sensibles de dates, de répétitivité de la collecte** de données, ainsi que de données manquantes. Ces différences peuvent être redressées à l'aide d'**estimations temporelles** permettant de comparer correctement les pays, à partir de modélisations plus ou moins complexes des tendances observées (pour une synthèse récente et une expertise sur les possibilités d'automatisation de ces méthodes, voir Charlton *et al.*, 2014). D'autre part, la création de séries temporelles soulève la question de la gestion des **changements de maille administrative**. Ces changements sont fréquents à l'échelon local. A titre d'exemple, l'INSEE, qui recense les modifications opérées à l'échelon communal depuis 1943¹¹, a relevé près de 1500 modifications depuis 2000 (**Erreur ! Source du renvoi introuvable.**).

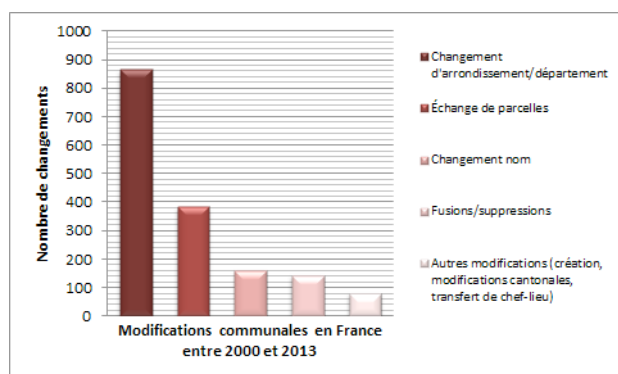


Figure 10 : Des modifications fréquentes de mailles à l'échelle locale (INSEE, 2014¹²)

Une solution peut consister à gérer ces changements de maille en proposant différentes versions de nomenclature spatiale comportant la liste des modifications introduites, à l'instar de ce qui est proposé pour les NUTS au niveau européen¹³. Pour aller plus loin, l'intégration de ces changements dans une base de données comportant des séries temporelles nécessite de formaliser, d'un point de vue conceptuel, les types de changements territoriaux pouvant intervenir et de créer un modèle de données spatio-temporel pour automatiser ces changements à partir d'une géométrie commune (Ben Rebah, 2008 ; Ben Rebah *et al.*, 2011 ; Plumejeaud, 2011).

Les différents aspects de l'harmonisation des données et des géométries évoqués ci-dessus impliquent que les données traitées dans les analyses futures ne soient plus les données officielles mais des données corrigées. Il faudra s'assurer de la clarté de ce contexte (harmonisation pour comparaison à des échelles transfrontalières), et éventuellement toujours proposer les données officielles lors de la mise à disposition des données.

¹¹ <http://www.insee.fr/fr/methodes/nomenclatures/cog/historique.asp>

¹² <http://www.insee.fr/fr/methodes/nomenclatures/cog/telechargement.asp>

¹³ http://epp.eurostat.ec.europa.eu/portal/page/portal/nuts_nomenclature/history_nuts

Fiche 6 – Diffusion, gestion courante et mise à jour de la base

Les questions de diffusion, de gestion courante et de mise à jour de la base dépendent de nombreux critères : moyens humains et financiers, complexité de la base de données, modes de collecte, solutions informatiques (de la plate-forme de fichiers à un système de gestion de base de données intégré), degré d'ouverture ou de verrouillage des données partagées, fréquence des mises à jour... Seul le **choix du mode d'organisation** de l'observatoire sera abordé dans cette fiche.

Différents modes d'organisation, plus ou moins centralisés, peuvent être envisagés de manière schématique (Figure 10) :

- Le **modèle le plus décentralisé** (en bas du schéma) s'appuie sur la délégation à des équipes nationales ou subnationales de la collecte des données, à partir d'un cadre commun définissant les normes de collecte et assurant le stockage comme la diffusion des données. Le projet *Correlates of war* (www.correlatesofwar.org) correspond à la mise en place d'un tel dépôt de fichiers, à partir d'un modèle de « décentralisation coordonnée », sans harmonisation ni validation *a posteriori*.

- Le **modèle le plus centralisé** (en haut du schéma) correspond à l'expérience menée dans le cadre de ce contrat, avec des variantes selon que la consultation/validation par des équipes nationales est associée au travail de l'ingénieur. C'est le modèle a priori le moins coûteux, mais qui nécessite une coordination et une participation active des équipes nationales.

- **Une organisation intermédiaire** peut s'appuyer sur la délégation des tâches de mise à disposition des données et par des représentants nationaux. L'observatoire ayant pour mission d'harmoniser, renseigner, intégrer et mettre à disposition ces informations de façon intelligible. Ce modèle se rapproche de ce qui est proposé par Eurostat.

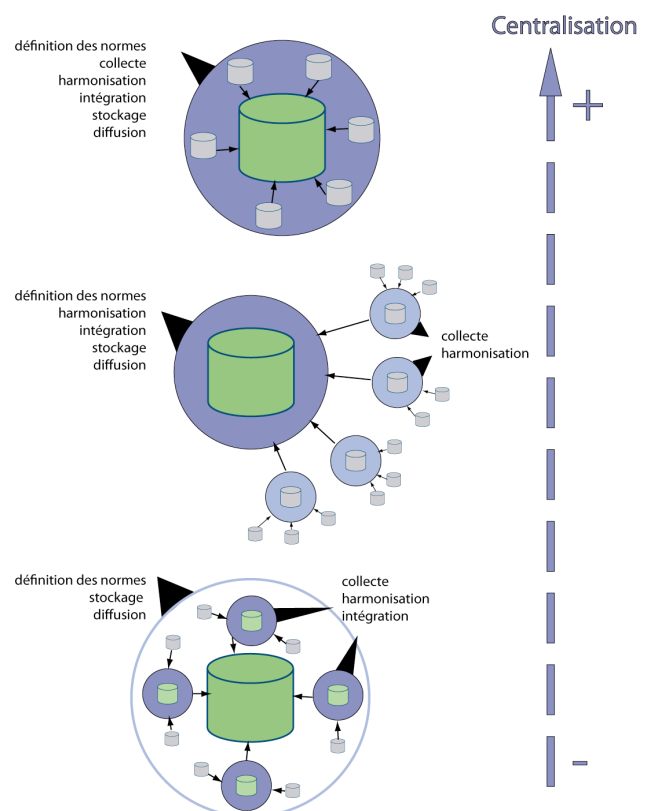


Figure 10 : Différents modèles de gestion pour une interface de diffusion des données

Le **choix de l'interface** et de ses fonctionnalités (depuis la plate-forme de mise en commun des données jusqu'à l'interface intégrant des outils cartographiques et d'analyse spatiale, avec différents niveaux d'interactivité) sera abordé en quatrième partie du rapport et intégré à la présentation des différents scénarios pour un observatoire transfrontalier (avec des éléments de coût et d'équivalent temps).

3. Etude de cas sur les enjeux de l'harmonisation spatiale dans l'espace transfrontalier franco-belge : des unités spatiales élémentaires aux bassins d'emploi

Dans cette partie, nous nous appuyons sur le **cas du tronçon transfrontalier franco-belge** pour illustrer certains enjeux de l'harmonisation spatiale des données sur l'emploi. A travers ces enjeux, c'est bien la question du **niveau pertinent d'analyse des phénomènes locaux** qui est soulevée, selon deux angles complémentaires :

-d'une part, il sera question du **découpage le plus fin à retenir** pour analyser et représenter l'information géographique de manière comparable de part et d'autre de la frontière. L'intérêt de choisir des mailles de taille équivalente ou des méthodes alternatives de carroyage et de lissage a déjà fait l'objet de plusieurs développements appliqués aux espaces transfrontaliers (voir par exemple de Ruffray *et al.*, 2008 et 2011). Nous prolongeons ici brièvement la réflexion en confrontant ces différentes solutions et en proposant une méthode générique d'évaluation des écarts de gabarit entre mailles, qui pourra accompagner le choix des découpages élémentaires au sein d'un observatoire de données.

-de plus, au-delà des maillages élémentaires de nature administrative, il est important d'explorer la question plus complexe de **l'harmonisation des maillages fonctionnels locaux**. Ces découpages, qui sont de taille intermédiaire entre les unités statistiques élémentaires des LAU2 et le niveau des NUTS3 ou des NUTS2, agrègent au sein de bassins de vie, d'emplois ou de services, des unités spatiales étroitement liées entre elles par des flux de différente nature et qui composent un ensemble territorial relativement autonome par rapport aux espaces environnants (Grasland, 2011). De tels périmètres, d'ores et déjà utilisés pour établir des diagnostics dans le cadre des politiques nationales, seraient également pertinents pour accompagner la mise en œuvre de certaines politiques européennes : Coombes *et al.* (2012) évoquent par exemple l'intérêt potentiel de bassins d'emploi harmonisés pour accompagner les politiques de cohésion territoriale et d'emploi (*Future Cohesion Policy & Territorial Agenda 2020, European Employment Policy & European Youth Policy*). Dans cette partie, nous repartirons de la méthode de construction de bassins d'emploi que l'Institut de Gestion de l'Environnement et d'Aménagement du Territoire (IGEAT) de Bruxelles a créée pour l'institut Wallon de l'Evaluation, de la Prospective et de la Statistique (IWEPS) (Marissal *et al.*, 2011) afin de transposer cette méthode côté français et de reconstruire un découpage harmonisé de ces zones de part et d'autre de la frontière. A titre plus exploratoire, nous proposerons également un découpage alternatif de bassins d'emploi transfrontaliers, fondés sur l'attraction théorique des pôles d'emploi modélisée à partir de la hiérarchie de ces pôles et d'hypothèses sur le coût des distances parcourues (distance kilométrique sur réseau).

3.1. Choix de l'espace d'étude et données

Les deux pistes d'harmonisation spatiale évoquées ci-dessus (harmonisation du découpage élémentaire, harmonisation de maillages fonctionnels locaux) sont illustrées à partir de données élémentaires sur l'emploi dans l'espace transfrontalier franco-belge, en 2000 et en 2010. Le choix du cas franco-belge tient en partie à l'inscription de ce secteur dans la portion frontalière la plus vive en France. Si la frontière franco-belge ne suscite pas autant de navettes transfrontalières que la portion franco-luxembourgeoise, qui enregistre deux fois plus d'actifs résidant en France (environ 73 000 contre 35 000 vers la Belgique), les flux échangés entre France et Belgique sont intéressants pour leur relative symétrie (flux belges vers la France seulement 3 fois plus faibles que les flux en sens inverse), en comparaison des écarts beaucoup plus marqués qui caractérisent les échanges avec l'Allemagne et le Luxembourg (flux vers la France 10 à 12 fois plus faibles) (INSEE Nord-Pas-de-Calais, 2014). D'autre part, l'existence de contacts scientifiques étroits entre l'UMS RIATE et l'IGEAT a été déterminante pour identifier plus rapidement les données belges et pour faciliter la transposition au cas français de la méthode de délimitation des zones d'emploi.

3. Etude de cas sur les enjeux de l'harmonisation spatiale dans l'espace transfrontalier franco-belge

L'Encadré 1 recense l'ensemble des données utilisées dans le cadre de cette étude¹⁴. Plusieurs variables classiques sur la population active et l'emploi sont renseignées à deux dates (environ 2000 et 2010) : population en âge de travailler, actifs et chômeurs au lieu de résidence, emplois, et actifs transfrontaliers. La connaissance des origines et des destinations des flux transfrontaliers reste un obstacle méthodologique majeur. Ces derniers peuvent être comptabilisés mais pas systématiquement localisés en fonction de leur origine et de leur destination. Néanmoins, certaines informations plus symétriques peuvent être exploitées : on connaît par exemple la commune de résidence des Français et des Belges sortants en 2000. On sait par ailleurs, en 2010, d'où partent les Français sortants et où travaillent les entrants en Belgique, sans pouvoir établir la correspondance entre les deux informations. Bien entendu, ces données doivent être manipulées avec précaution, compte tenu de l'hétérogénéité des définitions de la population active et du chômage dans les deux pays, ainsi que de la méthode employée pour collecter les flux¹⁵. Le choix a été fait de les utiliser malgré tout, tout en rappelant les divergences de définitions dans une note méthodologique qui accompagne chaque carte.

Encadré 1 : Disponibilité des données sur l'emploi en France et en Belgique

	France		Belgique	
	1999	2010	2000	2010
Population en âge de travailler	Oui (15 ans ou plus, 15-64 ans)	Oui (15 ans ou plus, 15-64 ans)	Non (a priori rien sur Statbel et Iweps)	Oui (15-64 ans)
Population active	Oui (15 ans ou plus)	Oui (15 ans ou plus, 15-64 ans)	Oui (15-64 ans Wallonie, 18-64 ans Flandre)	Oui (15-64 ans)
Population active occupée	Oui (15 ans ou plus)	Oui (15 ans ou plus, 15-64 ans)	Oui (15-64 ans Wallonie, 18-64 ans Flandre)	Oui (15-64 ans)
Emplois localisés	Oui (15 ans ou plus)	Oui	Non (manque les emplois occupés par les étrangers)	Oui
Chômeurs	Oui (15 ans ou plus)	Oui (15 ans ou plus, 15-64 ans)	Oui (15-64 ans Wallonie, 18-64 ans Flandre)	Oui (15-64 ans)
Travailleurs transfrontaliers	Oui	Oui (avec précision du pays de destination)	Oui	Non
Navetteurs	Oui (avec destination étranger)	Oui (avec destination étranger et précision des communes de destination)	Oui (avec destination étranger)	Oui (avec origine étranger)

Sources : INSEE (France), IWEPS (Wallonie), Seuntpunt WSE (Flandre)

D'un point de vue cartographique, c'est le fond de Jacques Michelet de l'université de Genève (cf. partie 2, fiche 4) qui a été utilisé. Il a été créé à partir des centroïdes des communes et d'une interpolation selon la méthode des polygones de Thiessen. Les contours reconstitués ne coïncident pas exactement avec les limites communales, ce qui peut gêner la lecture à un échelon local. Néanmoins, cette base de données spatiales présente l'avantage d'être quasi libre de droits (convention à établir) et ne pose pas de problème

¹⁴ Nous ne reprenons pas dans cette partie les étapes préalables de construction des métadonnées et de collecte des données, détaillées en partie 2 de l'étude.

¹⁵ Les derniers recensements de l'INSEE permettent d'obtenir des flux bi-localisés dans le sens France-Belgique, mais seulement à partir d'un échantillonnage qui est réalisé sur cinq années glissantes (INSEE Nord-Pas-de-Calais, 2014).

d'union de part et d'autre de la frontière. Enfin, nous avons défini une bande transfrontalière élargie (Encadré 2), qui couvre toute la Belgique et les arrondissements français situés à moins de 100 km de la frontière (Carte 1).

Encadré 2 : Profondeur de la bande transfrontalière

Pour définir le périmètre d'analyse de l'espace franco-belge, deux références ont été utilisées :

- la délimitation fixée par la convention fiscale entre la France et la Belgique en matière d'impôts sur le revenu, i.e. les communes à moins de 20 km de la frontière ;
- la distance retenue dans le précédent rapport pour la DATAR (de Ruffray *et al.*, 2011) : afin d'exprimer l'espace de vie quotidien des habitants, i.e. la distance kilométrique de 60 km de part et d'autre de la frontière, souvent significative de la distance parcourue en une heure à partir des pôles d'emplois principaux français ou étrangers.

Nous avons cherché des ruptures dans la variation de l'intensité du travail transfrontalier en fonction de la distance à la frontière, en Belgique et en France (Figure 11) : au-delà des différences dues à la plus forte intensité du travail transfrontalier en France, la première rupture correspond globalement aux communes et cantons limitrophes, tandis que la seconde coïncide plus ou moins avec la distance de la délimitation fiscale (Carte 1).

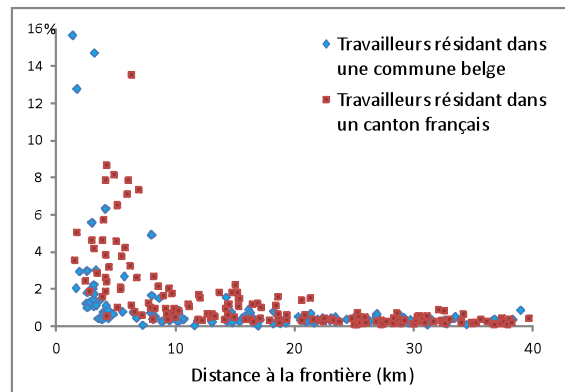
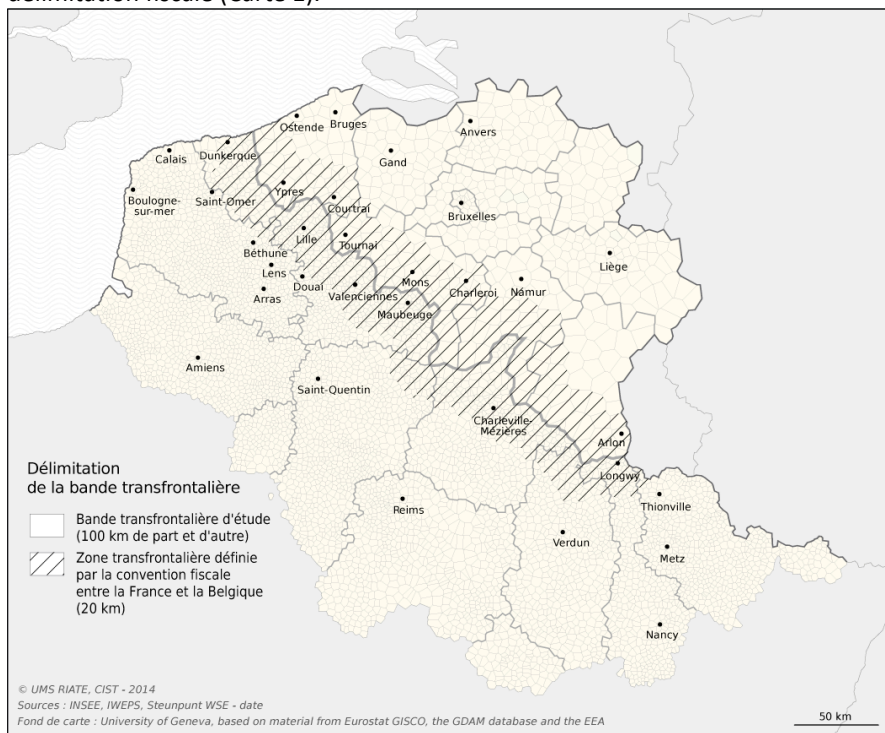


Figure 11 : Part des travailleurs frontaliers parmi les actifs occupés



Carte 1 : Choix du périmètre d'analyse pour l'étude

Dans le cadre de ce rapport, nous avons choisi de retenir, pour les représentations cartographiques (Carte 1), une distance de 100 km de chaque côté de la frontière, ce qui nous permet d'englober entièrement des cantons français et/ou départements (en particulier au moment de la définition des zones d'emploi). L'ensemble de la Belgique est alors pris en compte. Enfin, sur certaines cartes, nous avons distingué les unités territoriales à moins de 20 km par un léger décalage créant une impression de relief.

3.2. Harmonisation de découpages territoriaux élémentaires

3.2.1. Sélection de découpages élémentaires de taille équivalente

• Choix de mailles élémentaires de taille équivalente

Le choix de mailles élémentaires de taille équivalente de part et d'autre de la frontière présente un intérêt évident en termes d'harmonisation des représentations cartographiques et des analyses statistiques, même si ce choix peut sembler moins pertinent lorsqu'on s'intéresse aux niveaux politiques d'organisation des territoires locaux. Il conduit en pratique à combiner des mailles de niveau LAU2 et de niveau LAU1, en référence à la taille des unités les plus grandes. Dans le cas du tronçon transfrontalier franco-belge, les communes belges de niveau LAU2 sont alors associées aux cantons français de niveau LAU1.

Dans le prolongement du rapport coordonné par S. de Ruffray et G. Hamez (2008), nous examinons de manière plus systématique les conséquences de la sélection d'un niveau de découpage plutôt qu'un autre sur les analyses statistiques. En effet, afin de généraliser cette démarche à d'autres espaces transfrontaliers, une grille de comparaison des maillages peut être construite pour expliciter les critères retenus dans cette sélection, à partir de paramètres statistiques caractérisant la distribution des variables de surface et de population (Tableau 1).

Tableau 1 : Comparaison de maillages institutionnels en zone transfrontalière franco-belge, à partir des statistiques descriptives

		Moyenne		Médiane		Min		Max	
		Belgique	France	Belgique	France	Belgique	France	Belgique	France
Superficie (km ²)	Commune	51	9	43	8	3	2	167	52
	Commune/canton	51	123	43	124	3	5	167	445
	Arrondissement	703	1 216	583	1 188	64	14	1 927	2 894
Population active occupée	Commune	7 454	519	5 034	143	42	0	182 642	90 674
	Commune/canton	7 454	6 667	5 034	4 703	42	119	182 642	90 674
	Arrondissement	102 105	67 840	67 721	47 708	19 235	5 573	402 708	479 641
Taux de chômage*	Commune	9	10	8	9	2	-	34	42
	Commune/canton	9	13	8	12	2	4	34	29
	Arrondissement	11	14	10	13	4	9	23	20

Périmètre : 100 km côté français, Belgique entière. *chômeurs 15-64 ans / population active 15-64 ans

Sources : INSEE (France), IWEPS (Wallonie), Seuntpunt WSE (Flandre) ; données 2010

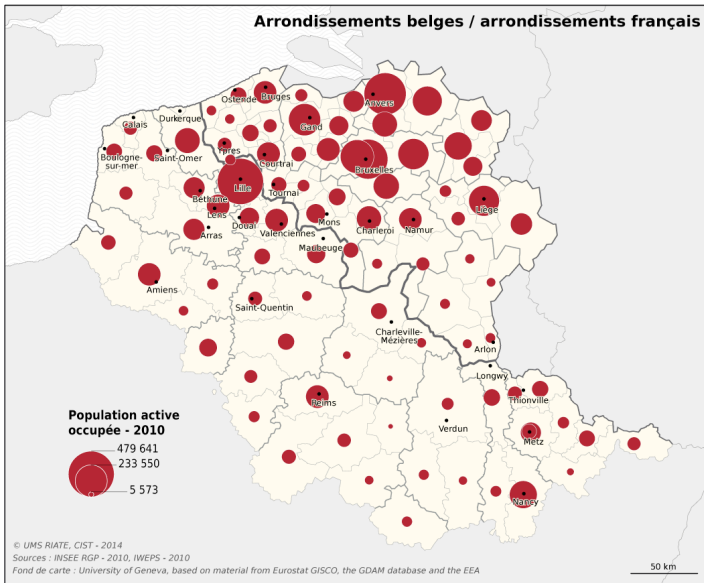
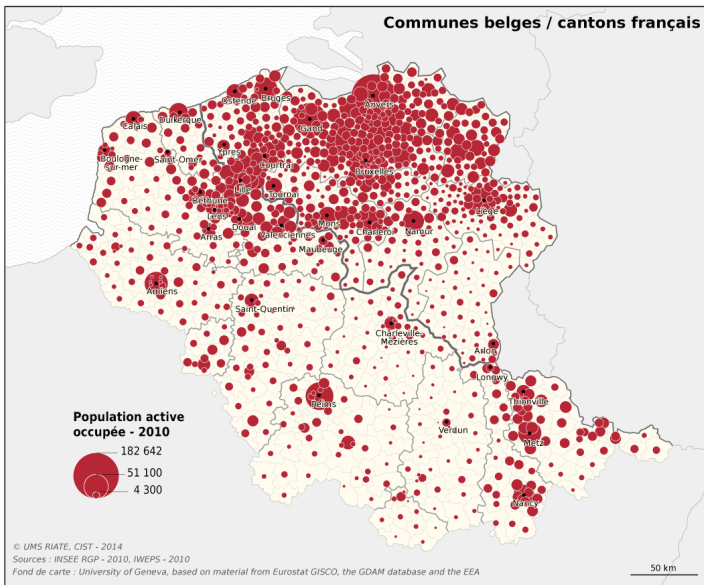
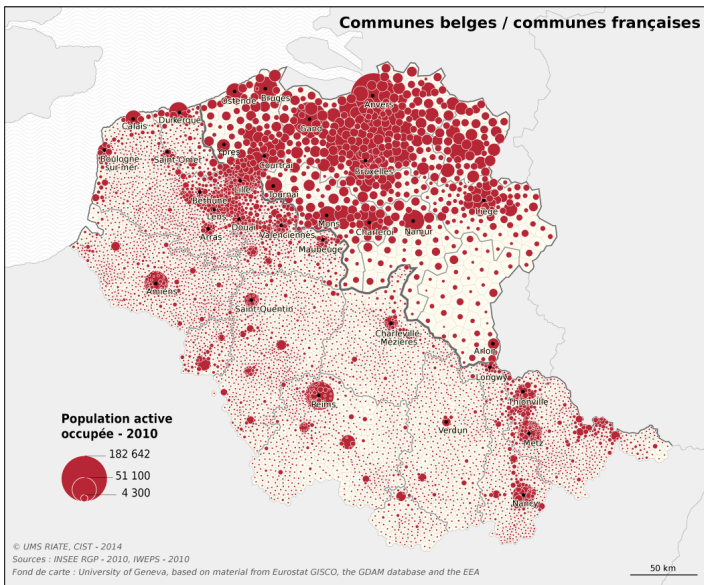
La comparaison des superficies des maillages institutionnels de part et d'autre de la frontière (plus précisément pour la France, en nous limitant à 100 km de la frontière) rappelle que les ordres de grandeurs diffèrent largement entre les pays. Les communes françaises sont environ cinq fois plus petites que les belges (Tableau 1 et Tableau 2), ce qui biaise l'analyse cartographique des répartitions de la population active occupée en 2010 sur la zone transfrontalière (Carte 2) : la population active occupée semble être surtout concentrée en Belgique. En prenant les maillages cantons pour la France / communes pour la Belgique, ou encore les arrondissements, le rapport est un peu plus équilibré pour les surfaces (mailles 2 à 3 fois plus vastes pour la France) et l'analyse comparative plus aisée. Dans le choix d'un maillage pertinent d'observation, la représentation graphique des distributions statistiques (Figure 12) des différents maillages peut compléter l'analyse et permet

d'estimer l'influence de l'agrégation spatiale, appelée *MAUP (Modifiable Areal Unit Problem)* ; Openshaw et Taylor, 1979 ; ESPON 3.4.3 MAUP).

Tableau 2 : Comparaison des superficies des unités territoriales entre la Belgique et la France

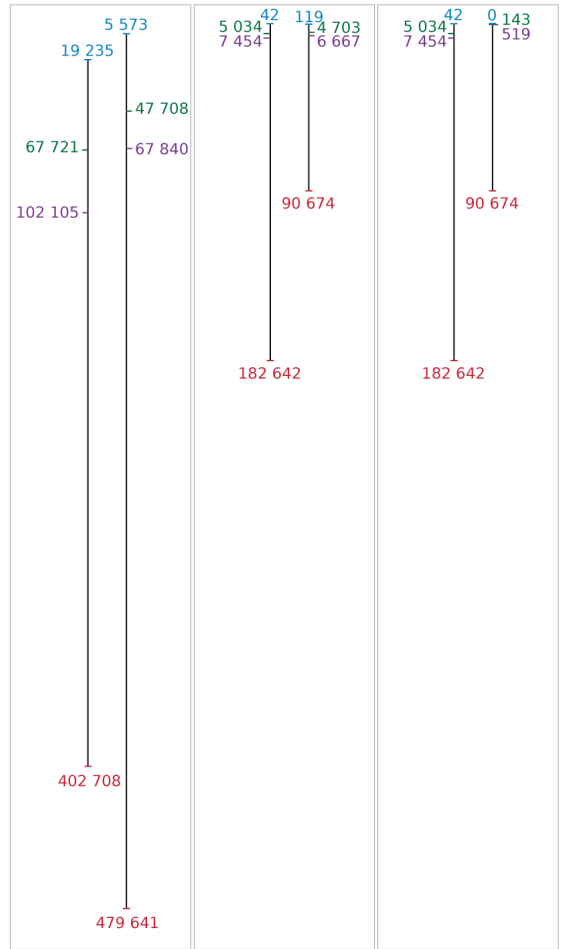
		France		
		Comm.	Cant.	Arr.
Belgique	Comm.	5,38	0,35	0,04
	Arr.	72,88	4,70	0,49

Comm. : communes, Cant. : cantons, Arr. : arrondissements
Calcul = médiane Belgique / médiane France
(ex : les communes belges sont 5,38 fois plus grandes que les communes françaises)



Note méthodologique
La population active occupée comprend les travailleurs salariés et non salariés de 15 à 64 ans recensés au lieu de résidence. En France, depuis 2004, le recensement permet de mieux prendre en compte les actifs ayant un emploi, même occasionnel ou de courte durée, et qui sont par ailleurs étudiants, retraités ou chômeurs.

Arrondissements		Cantons/Communes		Communes	
Belgique	France	Belgique	France	Belgique	France
43 arr.	47 arr.	589 com.	471 can.	589 com.	6055 com.



min
— médiane
— moyenne
max

Figure 12 : Représentation graphique des valeurs centrales et extrêmes de la population active occupée (2010) selon plusieurs maillages territoriaux

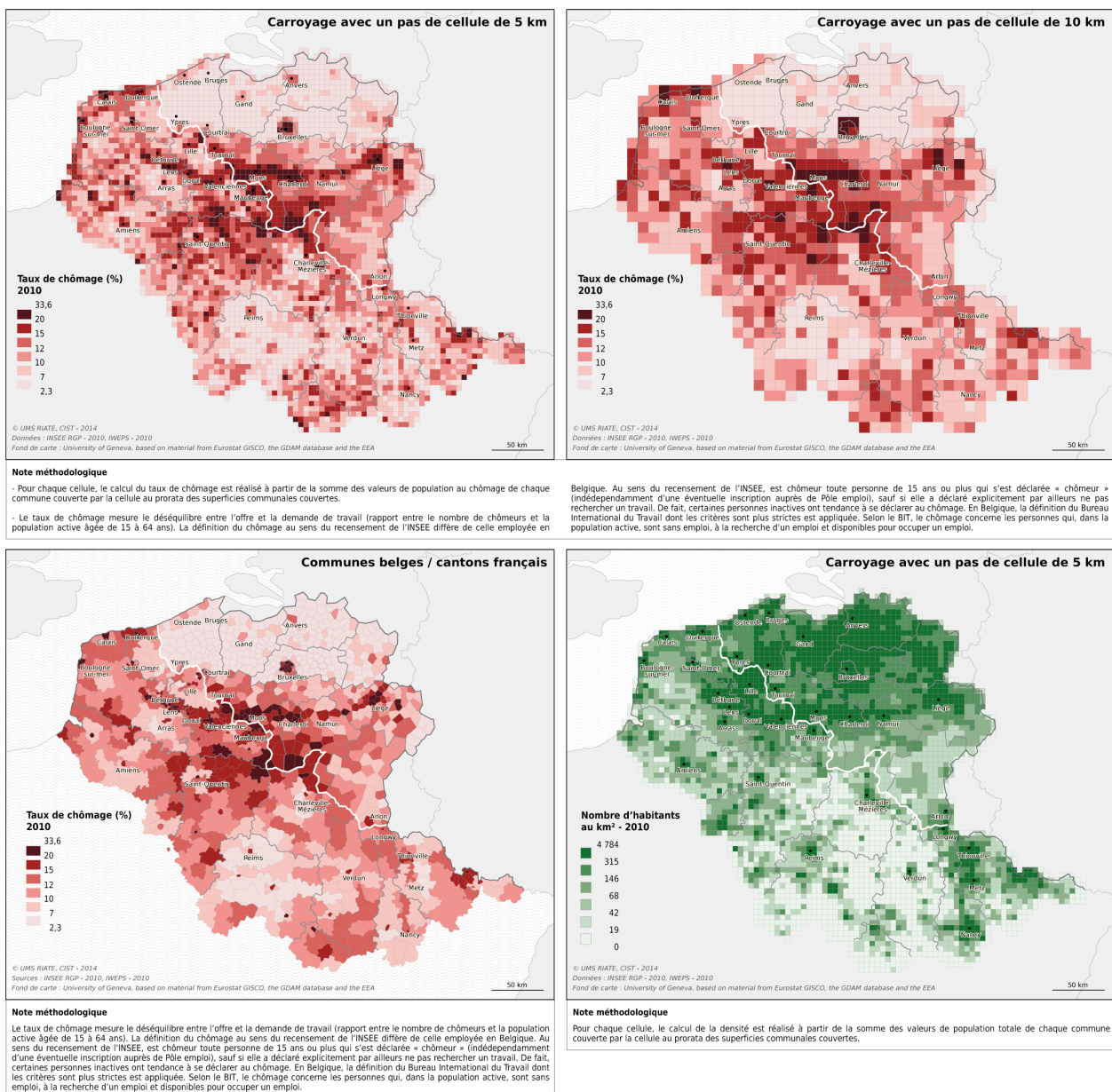
Carte 2 : Représentations cartographiques de la population active occupée (2010) selon plusieurs maillages territoriaux

3. Etude de cas sur les enjeux de l'harmonisation spatiale dans l'espace transfrontalier franco-belge

• Carroyages

Même si le choix de mailles administratives de taille équivalente permet d'améliorer l'analyse comparative, le rapport des superficies entre mailles belges et françaises reste encore légèrement déséquilibré (Tableau 1 et Tableau 2). Le carroyage des données est une méthode alternative qui repose sur la division de l'espace en carreaux de taille identique, auxquels on attribue les données des unités administratives au prorata de la surface occupée. La maille élémentaire devient ainsi le carreau, qui s'affranchit du découpage administratif et peut être à cheval sur une frontière. Cette méthode est plus satisfaisante du point de vue de l'harmonisation spatiale, peut également faciliter l'harmonisation temporelle et le croisement de données thématiques variées. C'est pourquoi les données carroyées sont un des enjeux actuels des statistiques européennes, à travers le projet GEOSTAT, lancé par Eurostat en 2010 et dont l'objectif est de favoriser l'intégration d'informations statistiques et géographiques au sein d'une infrastructure d'information européenne commune (ESSnet project GEOSTAT 1A, 2012).

Dans le cas du tronçon franco-belge, nous avons choisi de représenter en données carroyées le taux de chômage en 2010, avec des pas de cellule de 5km et de 10km (Carte 3). Nous ajoutons également à titre de comparaison la répartition selon le maillage cantons français / communes belges et, pour complément, la répartition de la densité de population (en vert).



Carte 3 : Représentations cartographiques du taux de chômage et de la densité de population par carroyage et, à titre de comparaison, selon des unités de taille équivalente

3.3. Harmonisation de découpages territoriaux fonctionnels (zones d'emploi)

3.3.1. Hétérogénéité des définitions nationales de zones d'emploi

Depuis une vingtaine d'années, un nombre croissant d'Etats européens ont défini des zones d'emploi et une majorité d'Etats d'Europe occidentale disposent à présent de ces découpages. Les méthodes et les données utilisées, tout comme les types d'acteurs produisant ces découpages (instituts statistiques nationaux ou autres), se caractérisent néanmoins par une très forte hétérogénéité (Figure 13). Parmi les définitions recensées, deux principaux types peuvent être distingués (Coombes *et al.*, 2012) :

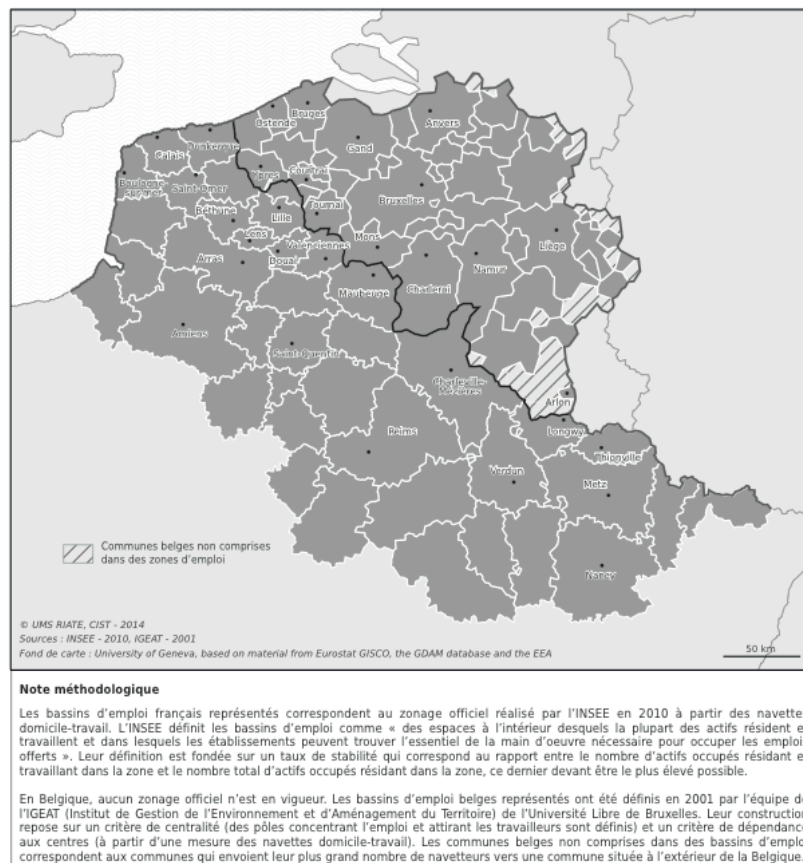
- Les zones d'emploi polarisées par un centre** (par exemple, en Suède) :
 Ces méthodes reposent (i) sur l'identification de centres, correspondant par exemple à des zones où la part des actifs ayant un emploi dans la zone d'emploi est supérieure à un certain seuil, (ii) puis sur l'agrégation des unités voisines en fonction de l'importance des flux vers ces centres. Relativement simples et transparentes, ces méthodes présentent plusieurs limites : pas de seuil minimal de population, ce qui peut poser des problèmes de transférabilité ; des agrégations de populations parfois hétérogènes (en contradiction avec un angle politiques publiques) ; une approche mono-centrée pas toujours pertinente dans des contextes polycentriques ; une forte sensibilité à l'hétérogénéité de la taille des mailles (MAUP).
- Les zones d'emploi non-centrées** (par exemple, les zones d'emploi INSEE en France ou les *Travel-to-Work-Areas* au Royaume-Uni) :
 Les zones sont définies à partir d'un critère d'autonomie relative des flux échangés (*self-containment*) : par exemple, la part des actifs résidant dans la ZE et qui ont un emploi dans la ZE, ramenée à la part des emplois de la ZE occupés par les actifs résidant dans la ZE. Un critère de contiguïté est en général imposé *a posteriori*. Ces méthodes s'avèrent moins sensible au MAUP et sont préférables pour des comparaisons européennes. Cependant, la construction des zones ainsi obtenues s'avère plus complexe et la documentation détaillée moins accessible.

Country	National denomination	MS:	BE ¹⁵	DE ¹⁶	EE	FI	FR	IT	SE	UK
		issue:								
		1 (non-core-based)	✓	?	✗	✗	✓	★	✗	✓
BE*	bassins d'emploi	2 (exhaustive & non-overlapping)	✓	✓	✓	✓	✓	✓	✓	✓
CZ	-	3 (explicitly defined labour markets)	✓	✓	✓	✓	✓	✓	✓	✓
DE*	Arbeitsmarktregionen	4a (consistency – method)	✓	✓	✓	✓	★	✓	✓	✓
EE*	Tööjõuareaal	4b (consistency – parameters)	✓	✓	✓	✓	✓	✓	✓	✓
EL	topikes agores ergasias	4c (consistency – analytical)	✗	✗	✓	✓	★	★	✓	★
FR*	zones d'emploi	5 (contiguous LMAs)	✓	?	✓	✗	✓	✓	✓	✓
IT*	Sistemi locali del lavoro	6 (aligned to adm. boundaries)	✗	★	✗	✗	★	✗	✗	✗
NL*	Regionale Platforms voor de Arbeidsmarkt	7a (clarity/simplicity)	★	?	✓	✓	★	★	✓	✓
PT	-	7b (transparency/reproducible)	✗	★	✓	✓	✓	★	✓	★
SK	zóny zamestnanosti	8 (aggregative)	✓	✓	✓	✓	✓	✓	✓	✓
FI*	Työssäkäyntialueet	9 (based on additional datasets)	✓	★	★	✓	✓	✓	✓	✓
SE*	Lokala arbetsmarknader	10a (minimum self-containment)	✗	✓	✗	✗	★	✓	✗	✓
UK*	Travel-To-Work-Areas	10b (self-containment guides the process)	✓	?	✗	✗	✓	✓	★	✓
		11a (minimum population)	✗	✓	✗	✗	?	✓	✗	✓
		11b (maximum area/distance/time)	✗	✓	✗	✗	✓	✗	✗	✗

Figure 13 : Comparaison des définitions nationales en Europe de zones d'emploi, selon plusieurs critères définis par l'étude de Coombes *et al.*, 2012

Les zones d'emploi définies en France et en Belgique sont une bonne illustration de cette diversité de définitions et de méthodes. Les zones d'emploi définies par l'IGEAT pour l'IWEPs correspondent à la première définition dite centrée. Les zones d'emploi de l'INSEE sont définies indépendamment d'un centre urbain et correspondent à la deuxième définition. Sur la zone transfrontalière précédemment définie (Encadré 2, p.27), ces deux découpages fonctionnels sont caractérisés (i) en Belgique, par 42 bassins d'emplois de tailles très variées, de 19 à 3222 km² et avec une superficie moyenne de 720 km² (médiane : 347) et (ii) en France, par 40 zones d'emplois en France d'environ 1480 km² (médiane : 1147 / min : 245 /

max : 3731) (Carte 4). Ces simples ordres de grandeur montrent bien les forts contrastes existant de part et d'autre de la frontière franco-belge.



Carte 4 : Découpages territoriaux fonctionnels : les bassins d'emploi en Belgique, définis par l'IGEAT (2001) et les zones d'emploi INSEE en France

3.3.2. Reconstitution de zones d'emploi à partir des flux domicile-travail

• Principes et méthode

Les travaux récents sur l'emploi dans l'espace transfrontalier franco-belge s'appuient fréquemment sur le choix de mailles de gabarit équivalent (cf. 3.3.1.), avec d'un côté, en France, le découpage en zones d'emploi de l'INSEE, et de l'autre, le découpage administratif des arrondissements belges. L'application qui suit vise plutôt, dans le même esprit que le travail effectué par Coombes *et al.* pour Eurostat (2012), à tester la pertinence d'une reconstruction des zones d'emploi harmonisées à partir d'une même définition, bâtie sur des critères fonctionnels et appliquée aux flux domicile-travail connus de part et d'autre de la frontière. Dans cette perspective exploratoire, la méthode qui a été choisie comme référence est celle qui apparaît la plus simple à mettre en œuvre, à savoir la méthode de l'IGEAT (Marissal *et al.*, 2006, voir Encadré 3 pour une présentation synthétique de la méthode). Elle a été transposée aux données françaises afin de définir de nouveaux bassins d'emploi sur la base des flux domicile-travail entre cantons. Il n'y a pas de bassins transfrontaliers, faute d'informations disponibles sur les flux transfrontaliers (de manière exhaustive et symétrique). L'objectif de cette démarche est double : d'une part prendre la mesure des changements induits par ce nouveau découpage, par comparaison aux zones d'emploi de l'INSEE. D'autre part, discuter de la pertinence d'une application aux données françaises de paramètres (seuils de densité et effectifs d'emploi) fixés en fonction du contexte belge de peuplement et d'emploi.

Encadré 3 : Méthode de construction de zones d'emploi de l'IGEAT

Méthode de définition des bassins d'emploi IGEAT

Un zonage en bassins d'emploi a été appliqué à la Belgique à partir de centres concentrant les emplois et de la polarisation des actifs par ces centres dans leur voisinage. La définition de ces bassins d'emploi a été réalisée en 2 étapes :

1. Définition des centres d'emploi

Les centres d'emploi sont constitués à partir d'un critère d'**attractivité** et d'un critère de **masse** (seuils définis de façon arbitraire) :

- *Critère d'attractivité* : Une commune est considérée comme un centre d'emploi si elle concentre plus d'emplois que de travailleurs résidents => **1,25 fois plus d'emplois** sur le territoire que de travailleurs résidents.
- *Ou critère de masse* : Une commune est considérée comme un centre d'emploi si elle concentre un volume d'emploi supérieur à **15 000 emplois**.

Certaines communes adjacentes des grands centres d'emploi (Anvers et Bruxelles) ne remplissant pas les critères précédents mais ayant une forte concentration d'emplois leur sont agrégées selon un critère de densité fixé à **600 emplois au km²**. Les centres d'emplois définitifs sont alors formés de ces **agrégations**.

2. Définition des zones d'influence

La méthode employée est la méthode des **flux majeurs**. Chaque commune est polarisée par le centre d'emploi vers lequel elle envoie le plus de navetteurs.

Cas particulier n°1 : si deux flux majeurs orientés vers deux centres d'emploi existent, la commune est considérée comme étant polarisée par les deux centres d'emplois (un seul cas lors de la définition des bassins d'emploi de 2001).

Cas particulier n°2 : si le flux majeur n'est pas orienté vers un centre d'emploi, la commune est considérée comme faisant partie du centre d'emploi polarisant la commune vers laquelle elle envoie son flux majeur (processus itératif).

Cas particulier n°3 : si la commune envoie son flux majeur vers l'étranger, elle est considérée comme faisant partie de la zone d'influence de centres extérieurs.

NB1 : La méthode a été implémentée à l'aide d'un programme sous le logiciel R. Aucun réajustement manuel n'a été effectué à la suite de ces deux étapes

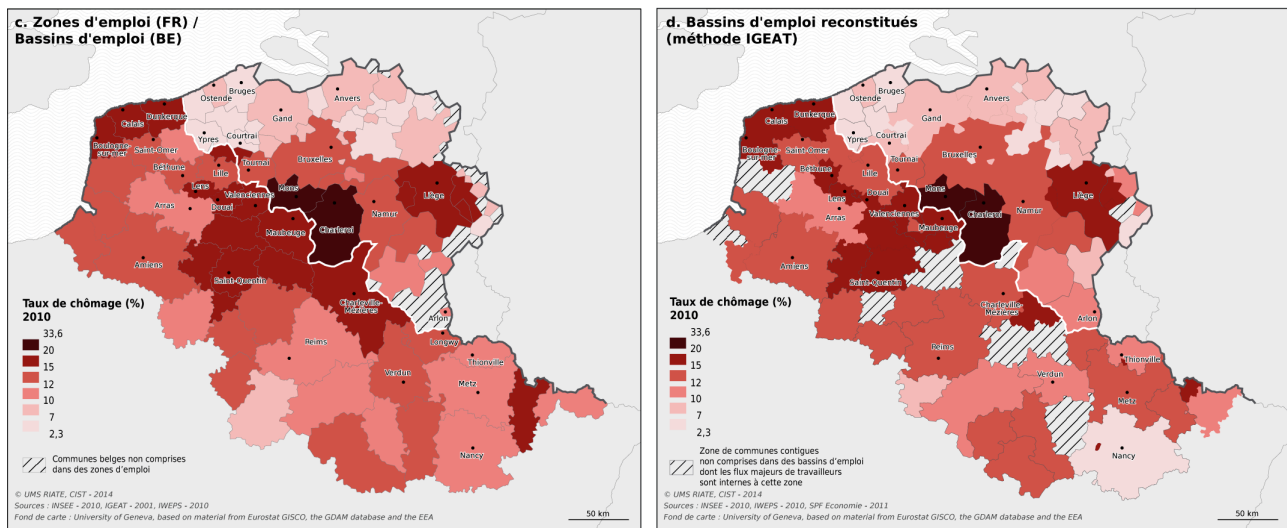
NB2 : Une autre méthode de définition des zones d'influence tenant compte de la part des actifs résidents se rendant vers les centres d'emploi, avec un seuil fixé à 10% du total des actifs résidents, a été appliquée aux 12 premiers centres d'emploi concentrant plus de 30 000 emplois

Voir : Marissal P., Medina Lockhart P., Vandermotten C., Van Hamme G., 2006, *Les structures socio-économiques de l'espace belge - Une exploitation des données de l'emploi de l'enquête socio-économique de 2001*, Monographies de l'enquête socio-économique 2001, Bruxelles, INS, 133 p.

• Résultats et discussion

La comparaison des découpages fonctionnels pour la portion française (40 zones d'emploi INSEE vs 42 zones reconstituées selon la méthode IGEAT), montre des **configurations dans l'ensemble assez proches**, notamment autour des principaux centres d'emploi (Carte 5). Néanmoins, plusieurs cantons français (37 sur 471) ne font pas partie de bassins d'emploi reconstitués à partir du transfert de la méthode de l'IGEAT et représentent une population active de 156 200 personnes sur 6950 km². L'application de la méthode des bassins d'emploi polarisés n'aboutit pas, comme en Belgique, à un découpage offrant une partition complète de l'espace d'étude. Dans la plupart des cas, les **zones interstitielles composées de cantons non polarisés** correspondent à des périmètres assez proches de certaines zones d'emploi de l'INSEE : c'est le cas par exemple autour de Berck-sur-mer et du Touquet plage, en Thiérache (autour de Vervins) à l'est de Saint-Quentin, ou encore dans la Woëvre entre Nancy et Verdun. Dans d'autres cas, elles agrègent les marges de plusieurs zones d'emploi INSEE, par exemple dans les Ardennes, entre Charleville-Mézières et Verdun. En dehors de l'existence de ces marges peu polarisées, les différences les plus remarquables tiennent à la **fragmentation de certaines zones d'emploi INSEE en plusieurs bassins d'emploi** polarisés autour de centres qui s'individualisent davantage : ainsi, la zone d'emploi correspondant à la vallée de la Sarre est partagée en plusieurs bassins de plus petite taille, centrés notamment autour de Forbach et Saint-Avold. Ces découpages plus fins ont d'autant plus de conséquences sur la représentation des taux de chômage ou de leur variation qu'ils individualisent des aires concentrant de fortes difficultés, par exemple à Forbach.

3. Etude de cas sur les enjeux de l'harmonisation spatiale dans l'espace transfrontalier franco-belge



Note méthodologique

Les bassins d'emploi français représentés correspondent au zonage officiel réalisé par l'INSEE en 2010 à partir des navettes domicile-travail. L'INSEE définit les bassins d'emploi comme « des espaces à l'intérieur desquels la plupart des actifs résident et travaillent et dans lesquels les établissements peuvent trouver l'essentiel de la main d'œuvre nécessaire pour occuper les emplois offerts ». Leur définition est fondée sur un taux de stabilité qui correspond au rapport entre le nombre d'actifs occupés résidant et travaillant dans la zone et le nombre total d'actifs occupés résidant dans la zone, ce dernier devant être le plus élevé possible. En Belgique, aucun zonage officiel n'est en vigueur. Les bassins d'emploi belges représentés ont été définis en 2001 par l'équipe de l'IGEAT (Institut de Gestion de l'Environnement et d'Aménagement du Territoire) de l'Université Libre de Bruxelles. Leur construction repose sur un critère de centralité (des pôles concentrant l'emploi et attirant les travailleurs sont définis) et un critère de dépendance aux centres (à partir d'une mesure des navettes domicile-travail). Les communes belges non comprises dans des bassins d'emploi correspondent aux communes qui envoient leur plus grand nombre de navetteurs vers une commune située à l'extérieur de la Belgique.

Note méthodologique

Les bassins d'emploi ont été définis selon la méthode élaborée par l'équipe de l'IGEAT (Institut de Gestion de l'Environnement et d'Aménagement du Territoire) de l'Université Libre de Bruxelles. Ils sont construits à partir des communes belges et des cantons français. Une commune (ou un canton) est centre d'emploi si elle concentre au moins 15 000 emplois ou/et si elle concentre 1,25 fois plus d'emplois sur son territoire que de travailleurs résidents. Chaque commune (ou canton) est polarisée(e) par le centre d'emploi vers lequel elle/il envoie son plus grand flux de travailleurs.

Carte 5 : Représentations cartographiques du taux de chômage selon deux types de découpages fonctionnels

Ces comparaisons et le traitement des zones non polarisées suscitent plusieurs questions d'ordre général sur la démarche d'harmonisation. D'une part, on peut s'interroger sur l'adaptation au contexte français des seuils définis pour la Belgique. On pourrait en effet choisir d'**ajuster à la baisse les seuils d'effectifs d'emploi ou d'attractivité**, de manière à adapter ces seuils aux plus faibles concentrations d'emploi des régions françaises et à capter un plus grand nombre de centres principaux. Toutefois, les essais effectués avec des seuils différents (Tableau 3) montrent que les périmètres définis sont globalement stables et qu'il faudrait modifier considérablement les ordres de grandeur de ces seuils (par exemple 5000 emplois au lieu de 15000) pour capter le plus grand nombre de centres. L'effort d'homogénéisation statistique (unités spatiales élémentaires de même gabarit, même méthode et mêmes paramètres pour les seuils de densité et les seuils de dépendance à l'emploi) n'apparaît pas suffisant pour assurer la comparabilité des découpages fonctionnels du fait de forts contrastes de peuplement internes à ces régions. Une autre piste de réflexion consisterait à **définir des seuils de centralités non pas absolus mais relatifs au voisinage de ces centres**, en s'inspirant de la démarche utilisée par Berroir *et al.* (2008) pour identifier des pôles d'emploi franciliens. Enfin, on pourrait plus simplement déterminer les bassins d'emploi auxquels se raccordent les cantons non polarisés en s'appuyant sur le **critère du plus grand flux envoyé vers l'un des centres** (et non pas sur le critère des flux majeurs). Cela dit, l'identification de cantons non polarisés représente aussi une **information pertinente sur l'existence d'espace de marge**, et il peut être justifié de les laisser tels quels, en les identifiant comme une catégorie spécifique de bassins d'emploi.

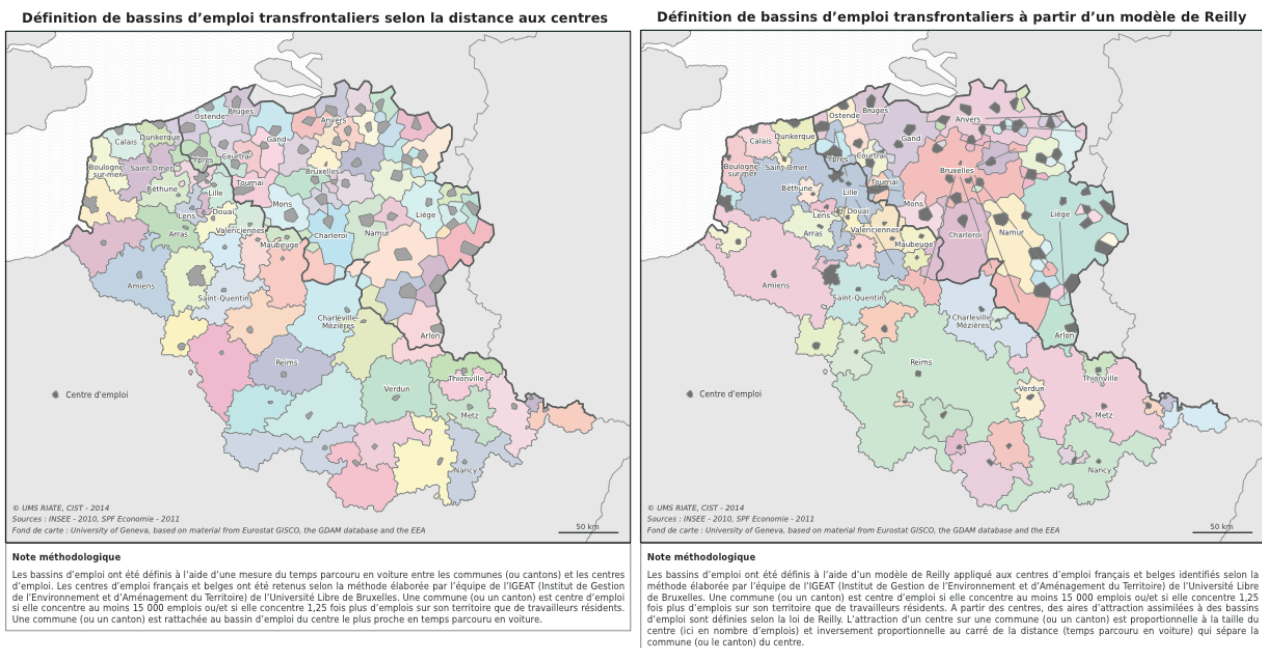
Tableau 3 : Effets de la variation des seuils d'emplois et d'attractivité

Seuils		Nombre de centres principaux	
Emplois localisés	Attractivité (emplois localisés / travailleurs résidents)	Belgique	France - zone frontalière (100 km)
10 000	1,25	49	41
15 000	1,25	43	39
15 000	1,15	45	44
15 000	1,50	50	45

3.3.3. Reconstitution de zones d'emploi transfrontalières théoriques

Les zones d'emploi définies à partir des navettes domicile-travail ne permettent pas de créer des zones transfrontalières, étant donné les données disponibles (Encadré 1, p. 26). Aussi avons-nous cherché à les reconstituer théoriquement, à partir de deux modèles : celui reposant sur la distance minimale aux pôles d'emploi et le modèle de Reilly (attraction proportionnelle à la masse d'emplois et inversement à la distance au carré). Les pôles d'emploi ont été identifiés à l'aide de la méthode IGEAT (Encadré 3, p.33)¹⁶. La distance entre les communes belges / cantons français et ces pôles¹⁷ est ensuite calculée à partir du réseau routier d'*OpenStreetMap* (pour plus d'informations sur la procédure, élaborée par Timothée Giraud : <http://rgeomatic.hypotheses.org/134>).

Dans le premier modèle qui tient compte de l'accessibilité aux pôles d'emploi, chaque canton / commune est associé(e) à son pôle d'emploi le plus proche, même s'il est situé de l'autre côté de la frontière. La Carte 6 (à gauche) montre ainsi que la majorité des zones d'emploi (à l'exception de Lille, Tournai, Mons) sont situées à cheval sur la frontière. Le périmètre de ces zones dépend essentiellement de la forme du semis des pôles d'emploi, i.e. de leurs espacements respectifs. Lorsque les bassins d'emploi transfrontaliers sont définis à partir d'un modèle de Reilly (Carte 6, à droite), la prise en compte de la hiérarchie des pôles d'emploi introduit des écarts beaucoup plus importants entre des zones d'attraction très étendues (pour certaines, transfrontalières) et des pôles de portée locale. Néanmoins, ces écarts dépendent fortement du choix du frein de la distance dans le modèle (ici égal à 2) et il serait important d'affiner le choix de ce seuil, en fonction des flux observés. D'autre part, l'application du modèle attire l'attention sur le traitement des zones d'influence des très grands pôles (Bruxelles, Lille), qui ressortent ici comme discontinues et drainent des actifs parfois bien au-delà de leur bassin de recrutement immédiat, de l'autre côté de la frontière (cantons d'Hirson et d'Aubenton, au sud de Maubeuge).



Carte 6 : Zones d'emploi franco-belges, reconstruites selon la distance aux centres (gauche) et à partir d'un modèle de Reilly (droite)

¹⁶ Précisons que nous avons choisi d'ajouter Berck (6269), situé dans une zone hors zone d'emploi, comme centre : bien que le canton ne remplisse par les critères pour devenir un centre selon la méthode IGEAT, le nombre d'emplois localisés y est supérieur à 10 000 et son taux d'attractivité est proche de 1,2. On peut donc imaginer que ce canton exerce une influence significative sur les espaces environnants.

¹⁷ Si le centre est composé de plusieurs communes ou cantons, la commune ou le canton retenu pour déterminer les coordonnées du centre est celle/celui dont le nombre d'emplois localisés est le plus important.

L'harmonisation des mailles d'observation est, nous l'avons vu, un enjeu crucial de toute étude transfrontalière. Les solutions envisageables pour rendre ces découpages comparables sont multiples. Dans le cadre d'un observatoire, ces solutions d'harmonisation dépendent non seulement de critères scientifiques mais aussi de l'adéquation aux usages potentiels de la base. La question de la pertinence des choix scientifiques se pose notamment lorsque ces choix s'affranchissent des limites administratives. Sans démultiplier les représentations, la possibilité d'avoir accès à des représentations alternatives d'un même phénomène enrichit sa compréhension et permet de prendre en compte la diversité des usages de la base de données.

4. Propositions pour la maintenance et la gestion d'un observatoire

Les trois premières parties du rapport ont mis en lumière les éléments incontournables qui doivent guider la réflexion pour la mise en place d'un observatoire pérenne du transfrontalier. Comme cela a été décrit, quelles que soient les étapes de constitution de cette base de données (définition du modèle, de l'espace d'étude, collecte, harmonisation, diffusion etc.), il n'existe jamais une solution unique mais toute une gamme d'alternatives qui peuvent être explorées.

Néanmoins, le fait de décomposer le montage de cet observatoire en grands « **working packages** » (données statistiques, données spatiales, infrastructure spatiale de données, dissémination) permet d'envisager la complexité des problèmes soulevés et des choix à effectuer sous un autre angle (Figure 14). En effet, **l'ambition** des chercheurs, ingénieurs et/ou managers qui seront mobilisés sur ce projet sera nécessairement **proportionnelle aux moyens financiers et temporels** mis à disposition pour mener à bien leurs activités.

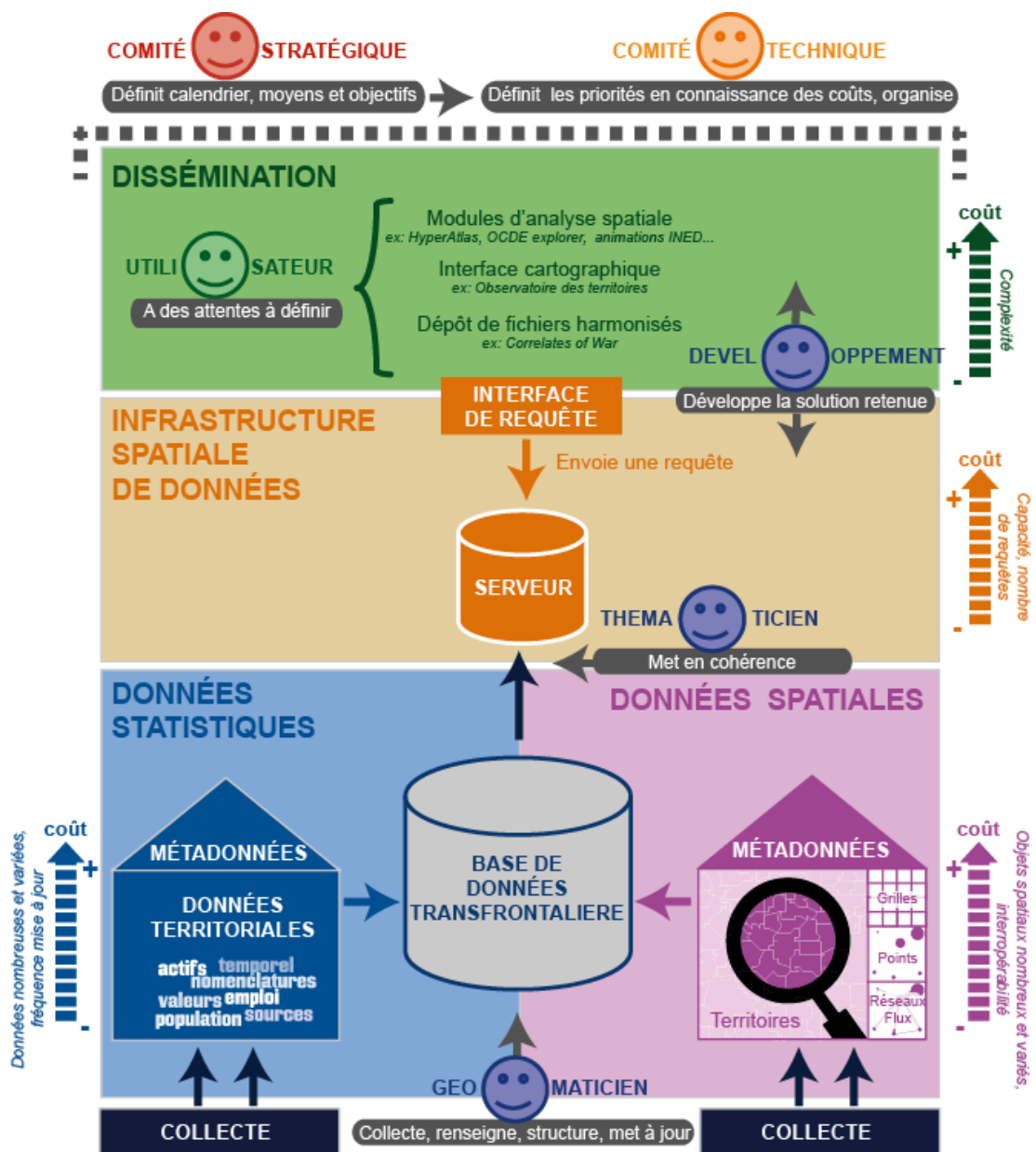


Figure 14 : Éléments de réflexion sur les enjeux, usages et moyens questionnés pour la mise en place d'un observatoire pérenne du transfrontalier

S'agissant du volet **statistique**, la question des moyens financiers mis en œuvre guidera les choix portant sur la quantité des données à collecter, leur profondeur temporelle (dernière donnée disponible ou plusieurs recensements ?), le nombre de fournisseurs de données qu'il faudra mobiliser pour cette collecte, la complexité du modèle de métadonnée à réaliser ou la fréquence de mise à jour de la base de données réalisée.

La prise en compte des coûts aura également son importance concernant **la quantité et la complexité des objets géographiques** qui seront collectés et mis en cohérence (mailles territoriales, réseaux, grilles etc.). De plus, ces éléments budgétaires guideront le choix du fond de carte de référence qui pourra être utilisé pour représenter ces résultats (création par l'observatoire ou non) ainsi que sa périodicité de mise à jour.

Par ailleurs, la complexité de **l'infrastructure spatiale de données** qui sera mise en œuvre dépendra étroitement des choix des **objectifs de diffusion** qui seront confiés à cet observatoire du transfrontalier. Le fait de définir un dépôt de fichiers harmonisés n'impliquera pas nécessairement la mise en place d'un système d'information reposant sur un modèle conceptuel de données. En revanche, il est évident que les utilisations potentielles des données disponibles seront selon cette solution beaucoup plus limitées. Inversement, la mise en place d'une interface cartographique d'interrogation des données – qui pourra toucher un public beaucoup plus large et créer une réelle synergie sur les questions transfrontalières - impliquera un investissement bien plus important : définition de spécifications guidant le développement informatique, création d'un système d'information approprié, développement, création d'une interface de requête, prise en compte d'aspects d'ergonomie et de design, etc.

Pour mener à bien la solution retenue, la mise en place d'un **cahier des charges** réalisé conjointement par les comités stratégique et le groupe technique, définissant clairement les **objectifs atteignables** pour toutes les personnes impliquées dans ce projet, constituera un préalable indispensable. Néanmoins et quelque soit le degré d'exactitude de ce cahier des charges, **les imprévus** seront d'autant plus conséquents que la solution retenue sera complexe à mettre en œuvre : des adaptations seront à prévoir et auront des implications sur l'ensemble de la chaîne de traitement de l'information (modifications du modèle de métadonnées, organisation de la collecte etc.).

L'engagement à temps-plein de personnels aux compétences variées sur le long terme constitue également un facteur de réussite supplémentaire pour la réalisation d'un observatoire pérenne du transfrontalier. Plusieurs profils métiers sont ainsi indispensables à associer à cette entreprise (figure 16) :

- **Le « géomaticien »** connaît les techniques d'acquisition, d'intégration et de diffusion des données géographiques. Ce professionnel participe à la création de base de données associant des cartes, images aériennes ou satellites, du texte et des statistiques. Il constitue ainsi la personne ressource essentielle pour collecter et/ou créer et/ou mettre à jour les données statistiques et spatiales nécessaires à l'existence de l'observatoire.

- **Le thématicien** est au fait des problèmes provoqués par l'existence de fournisseurs de données reposant sur des systèmes statistiques hétérogènes (MAUP, problèmes de définition d'indicateurs, aspects temporels etc.). Il est capable de proposer et d'implémenter des solutions permettant de surmonter ces difficultés afin de rendre l'information géographique comparable et intelligible pour le plus grand nombre.

- **L'ingénieur développeur** est la personne appropriée pour mettre en œuvre le système d'information au regard de spécifications préétablies (modèle conceptuel de données etc.). Une fois le système d'information mis en place, il est capable d'identifier le langage de programmation approprié à la réalisation d'une interface de requête et de son déploiement sur le Web. Selon le public visé, il pourra être associé à un graphiste pour rendre l'interface plus agréable et plus intuitive à consulter.

- Il est évident que ces profils métiers seront amenés à collaborer ensemble régulièrement. Aussi, la présence d'un **chef de projet** facilitera grandement la mise en musique des grandes étapes de la collecte, de l'harmonisation et du développement.

Il n'existe donc pas une solution unique pour constituer cet observatoire du transfrontalier. Afin de faciliter la reprise en main de ce travail ultérieurement, nous proposons pour terminer trois scénarios plausibles de réalisation. Ces scénarios sont présentés sous forme de fiches organisées de façon systématique. Nous avons fait varier deux paramètres essentiels : le temps du projet et le nombre de personnes impliquées dans le projet. Avec ces deux variables d'ajustement essentielles, il est possible de préfigurer l'ambition de la réalisation, ses objectifs, les principales caractéristiques du travail à réaliser et le flux de travail étalé dans le temps.

Les coûts estimés sont présents ici à titre indicatif, comme ordre de grandeur. Ils sont à relativiser au regard de l'organisme employeur, l'expérience des personnes engagées sur le projet et la période d'implémentation du projet. Par ailleurs, et quel que soit le scénario retenu, l'implication des Instituts Statistiques Nationaux nous semble, après expérience, un élément incontournable pour lancer ce projet sur de bonnes bases. Les coûts potentiels d'implication de ces organisations n'ont pas été évalués dans les scénarios proposés plus bas.

Les scénarios et estimations de coût reposent tous sur l'idée d'une collecte systématique d'informations statistiques et cartographiques pour la France métropolitaine et ses États frontaliers. Ces scénarios visent à la création d'un observatoire structuré de façon plutôt centralisée. Quel que soit le scénario retenu, la collecte des indicateurs devra être réalisée au niveau géographique le plus fin généralement disponible (équivalent LAU2). Trois scénarios sont ici proposés :

- **scénario « bas », réalisable en 6 mois avec un salarié temps-plein** (Figure 147);
- **scénario « intermédiaire », réalisable en 18 mois avec deux salariés temps plein** (Figure 148) ;
- **scénario « haut », réalisable en 36 mois avec trois salariés temps-plein** (Figure 149).

Ces scénarios n'incluent pas les activités de mise à jour nécessaires qui seront à évaluer une fois le projet terminé. Néanmoins, les coûts de mise à jour seront nettement allégés une fois la base de données mise en place.

Soulignons pour terminer que la réalisation de ces scénarios a fait l'objet d'un arbitrage difficile à réaliser, et que les scénarios doivent faire l'objet de mise à jour en fonction des orientations définies par le comité stratégique.

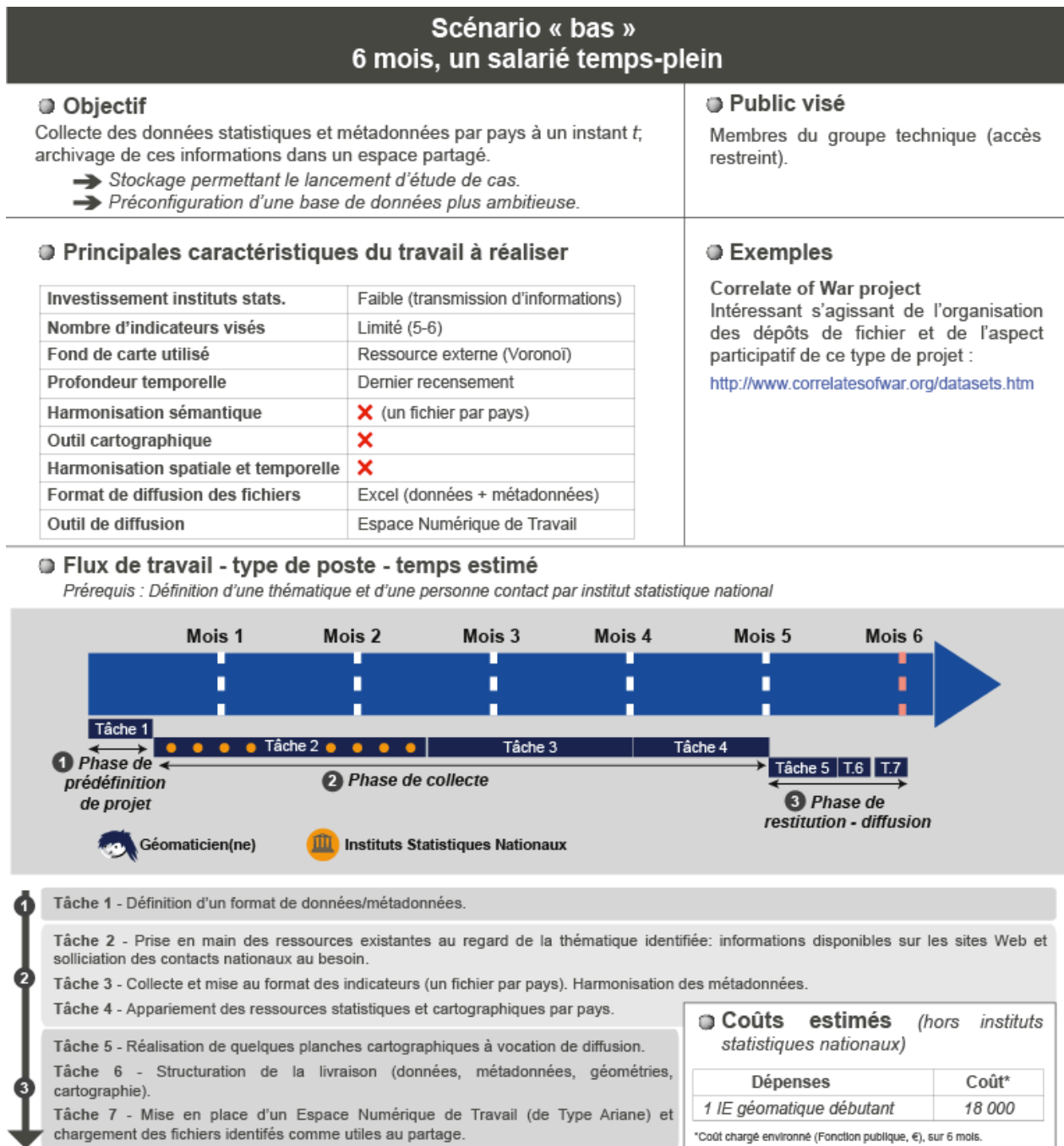


Figure 15 : Scénario « bas » pour la gestion et la maintenance d'un observatoire des espaces transfrontaliers

Scénario « intermédiaire » 18 mois, 2 salariés temps-plein

● Objectif

Collecte des données statistiques et géométries à plusieurs pas de temps et restitution sous forme d'une base de données transfrontalière requêteable.

→ Représentation comparable des territoires transfrontaliers.

● Public visé

Analystes et géomaticiens s'intéressant aux questions transfrontalières.

● Principales caractéristiques du travail à réaliser

Investissement instituts stats.	Importante (Partage de jeux de données)
Nombre d'indicateurs visés	10 (extensible)
Fond de carte utilisé	à créer (généralisé et harmonisé)
Profondeur temporelle	Intermédiaire (2 recensements)
Harmonisation sémantique	— Partielle
Outil cartographique	✗
Harmonisation spatiale et temporelle	— Partielle
Format de diffusion des fichiers	BD transfrontalière requêteable
Outil de diffusion	Site Web

● Exemples

ESPON Database

Intéressant s'agissant du modèle de métadonnées et les possibilités de requête (filtres implémentés):

<http://database.espon.eu/db2/search>

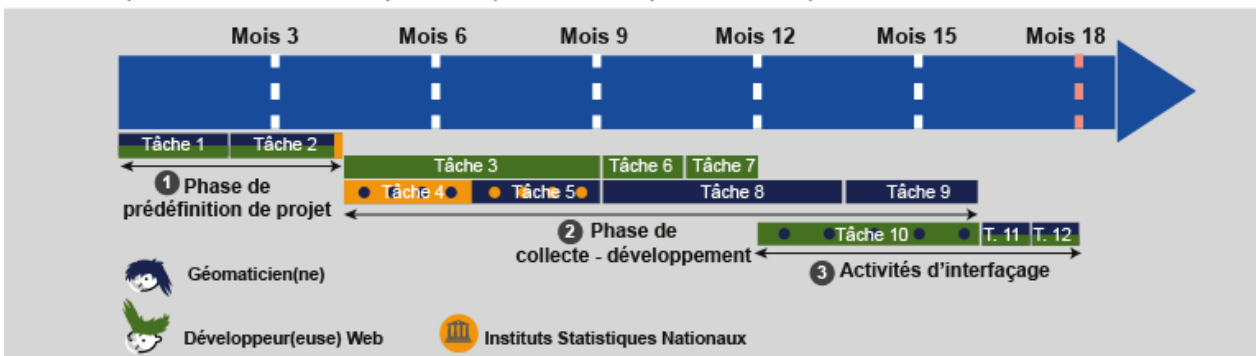
UNEP Geodata Portal

(sans les fonctionnalités cartographiques):

<http://geodata.grid.unep.ch/>

● Flux de travail - type de poste - temps estimé

Prérequis : Définition d'une thématique et d'une personne contact par institut statistique national



- 1 Tâche 1 - Création de spécifications : définition des informations (e.g. données et métadonnées) que l'on souhaite collecter et partager via le portail Web.
- 1 Tâche 2 - Définition du format de données - métadonnées et du schéma relationnel de la base de données. En fin de tâche 2, diffusion aux instituts statistiques nationaux des fichiers « modèle » à délivrer à l'observatoire du transfrontalier.
- 2 Tâche 3 - Implémentation du système de gestion de base de données.
- 2 Tâche 4 - Collecte et mise au format (selon le modèle défini en tâche 2) des indicateurs ciblés par chaque institut statistique national. Le géomaticien assure un support technique et veille à la conformité des fichiers délivrés.
- 2 Tâche 5 - Harmonisation sémantique: mise en cohérence de l'hétérogénéité des données et métadonnées fournies par les instituts statistiques nationaux. Les personnes contacts des instituts statistiques nationaux apportent des précisions si nécessaire.
- 2 Tâche 6 - Phase de tests sur quelques jeux de données et ajustements.
- 2 Tâche 7 - Intégration des jeux de données.
- 2 Tâche 8 - Création du fond de carte harmonisé et généralisé.
- 2 Tâche 9 - Adaptation du fond de carte aux recensements précédents (gestion des unités territoriales agrégées, scindées et changements de codes). L'idée consistant à disposer d'un fond de carte adapté à chaque recensement.
- 3 Tâche 10 - Développement de l'interface de requête. Réunions régulières avec le géomaticien à prévoir.
- 3 Tâche 11 - Création et structuration du site Web.
- 3 Tâche 12 - Création d'un guide utilisateur et manuel administrateur permettant d'alimenter/gérer la base de données réalisée a posteriori.

● Coûts estimés (hors instituts statistiques nationaux)

Dépenses	Coût*
1 développeur Web confirmé	75 600
1 IE géomatique confirmé	63 000

*Coût chargé environné (Fonction publique, €), sur 18 mois.

Figure 16 : Scénario « intermédiaire » pour la gestion et la maintenance d'un observatoire des espaces transfrontaliers

4. Propositions pour la maintenance et la gestion d'un observatoire

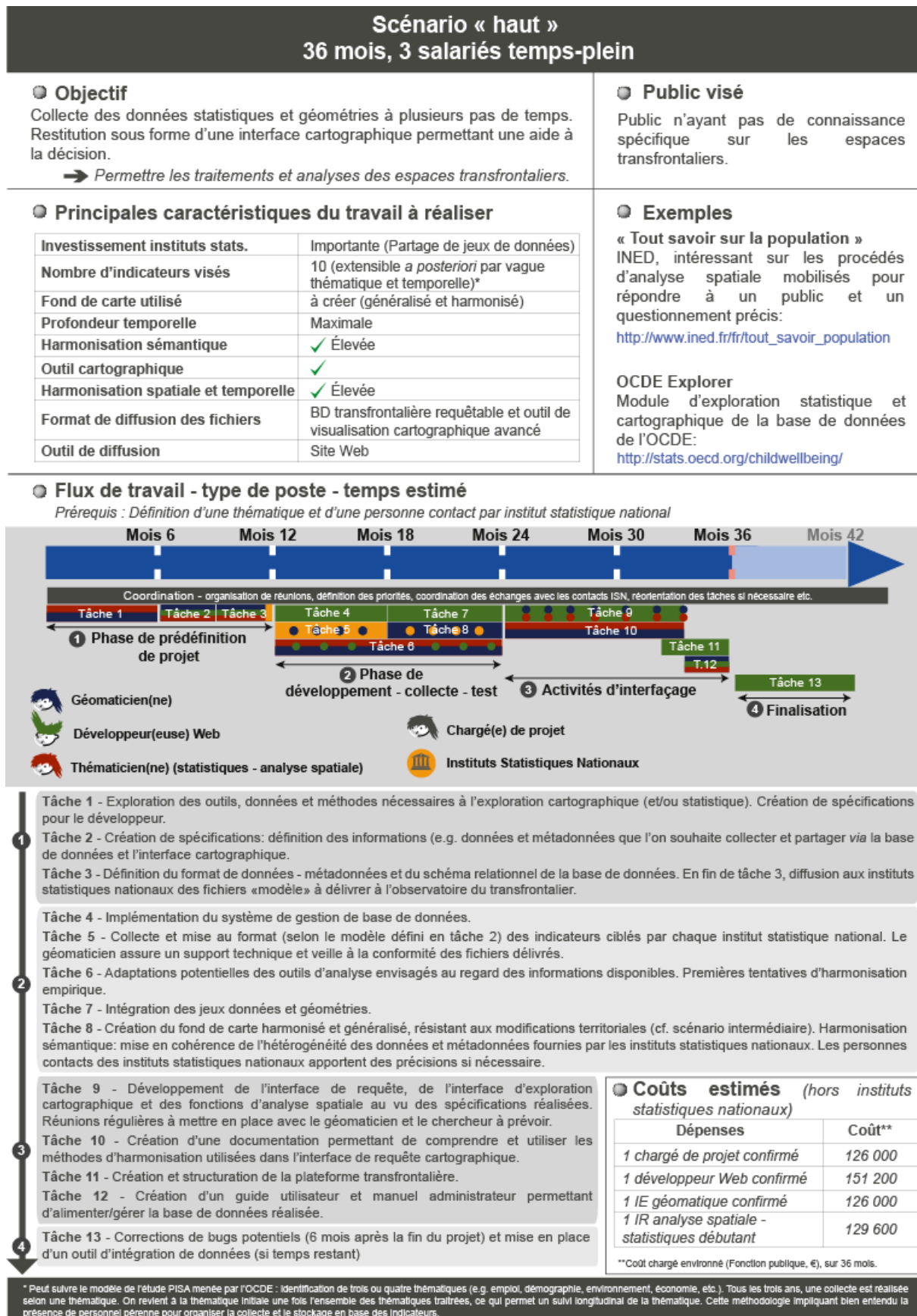


Figure 17 : Scénario « haut » pour la gestion et la maintenance d'un observatoire des espaces transfrontaliers

Conclusion

Cette étude de faisabilité d'une observation pérenne des espaces transfrontaliers réalisée par l'UMS RIATE et le GIS CIST pour le compte du CGET, a abordé les étapes nécessaires à la conception et à l'élaboration d'une base de données transfrontalière, d'un point de vue méthodologique. Elle a permis d'insister sur les enjeux de l'harmonisation de l'information géographique (statistique et cartographique) et, au-delà, sur l'intérêt de multiplier les représentations d'un même phénomène pour mieux le comprendre et pour permettre une diversité des usages de la base. Dans une optique plus opérationnelle, plusieurs scénarios de mise en place d'un observatoire ont été proposés, en fonction des attentes des utilisateurs, du degré de complexité de l'interface d'interrogation des données et du mode de gouvernance privilégié. L'étude visait à montrer en quoi le choix d'un scénario plutôt qu'un autre influence toute la chaîne de construction de la base de données, de la collecte à la valorisation. Enfin, plus généralement, elle a permis de poser clairement les conditions de la réalisation d'une base de données transfrontalière généralisée sur toutes les frontières, notamment l'existence d'une gouvernance et d'une coopération internationale qui est encore à parfaire et sans laquelle l'observation transfrontalière généralisée ne peut exister.

Références bibliographiques

- Ben Rebah M., 2008, *Cartographie dynamique et investigation territoriale : le cas de l'évolution du découpage administratif tunisien*, Thèse de doctorat de géographie, Université Paris Diderot.
- Ben Rebah M., Peeters D., Plumejeaud C., Ysebaert R., 2011, *Modeling territorial changes and time series database process, empirical approach and applications*, ESPON, Technical Report, 32 p.
- Berroir S., Mathian H., Saint-Julien T., Sanders L., 2008, La mobilité dans la construction du polycentrisme métropolitain", in : Thériault M., Des Rosiers F. (dir.), *Information géographique et dynamiques urbaines*, Paris, Hermès, 31-57.
- Bretagnolle A., Guérois M., Mathian H., Pavard A., 2014, *Functional Urban Areas in Europe. A feasibility study through the comparison of two models: commuting flows and accessibility isochrones*, Technical report, ESPON Database 2013.
- Charlton M. et al., 2014, *A strategy for elaboration and update of coherent time series of hierarchical territorial units*, ESPON M4D, Technical Report, June 2014.
- Coombes M., Casado-Díaz J.M., Martínez-Bernabeu L., Carausu F., 2012, *Study on comparable Labour Market Areas*, Research report for Eurostat, 68 p.
- De Ruffray S., Hamez G., Grasland C., Lambert N., Hamm A., Gallet-Moron E., 2011, *Enjeux des territoires frontaliers à l'échelle nationale – vers la détermination et la délimitation de pôles transfrontaliers*, DATAR, 33 p.
- De Ruffray S., Hamez G., Meddahi D., Moron E., Smits F., 2008, *La comparaison des territoires transfrontaliers à l'échelle nationale – Vers une typologie fondée sur les logiques d'intégration*, DIACT, 29 p.
- Decroly J. M., Grasland C., 1992, Frontières, systèmes politiques et fécondité en Europe. *Espace, populations, sociétés*, 10(2), 135-152.
- ESPON M4D, 2011, *The Core Database Strategy, a new paradigm for data collection at regional level*, Technical Report, ESPON, 107 p.
- ESPON M4D, 2014, *ESPON data and metadata specifications*, ESPON, 123 p.
- ESPON 3.4.3. MAUP, 2006, *The modifiable areas unit problem*, ESPON, 254 p.
- ESSnet project GEOSTAT, 2012, *GEOSTAT 1A – Representing Census data in a European population grid*, European Forum for GeoStatistics, 82 p.
- Grasland C., Mathian H., Vincent J.M., 2000, Multiscalar Analysis and map generalisation of discrete social phenomena: Statistical problems and political consequences, *Statistical Journal of the United Nations ECE*, 17, IOS Press, 1-32.
- Grasland C., 2011, *Towards an ESPON hierarchy of functional regions?*, Technical report, ESPON, 14 p.
- INSEE Nord-Pas-de-Calais, 2014, Les systèmes territoriaux en Nord-Pas-de-Calais : une région façonnée par les déplacements, *Pages de Profils*, n°115, avril 2014, 10 p.
- Marissal P., Medina Lockhart P., Vanderhoff C., Van Hamme G., 2011, *Les structures socio-économiques de l'espace belge - Une exploitation des données de l'emploi de l'enquête socio-économique de 2001*, Monographies de l'enquête socio-économique 2001, Bruxelles, INS, 133 p.
http://economie.fgov.be/fr/binaries/mono_200106_fr%5B1%5D_tcm326-36628.pdf
- MOT (Mission Opérationnelle Transfrontalière), 2014, L'observation des territoires transfrontaliers, *Cahiers de la MOT*, n°9, 15 mars 2014.
- MOT-FNAU, 2012, *Étude comparée et observation statistique des agglomérations transfrontalières, Tome 1, Méthodologie*, DATAR, 125 p.
- Openshaw S., Taylor J., 1979, A million or so correlation coefficients: three experiments on the modifiable areal unit problem. In: *Statistical applications in the Spatial Sciences*, Wrigley N. (ed.), Pion, Londres: 127-144.
- Plumejeaud C., 2011, *Modèles et méthodes pour l'information spatio-temporelle évolutive*, Thèse de doctorat en informatique, Université de Grenoble.
- Zanin C., Lambert N., Ysebaert R., 2013, *Mapping guide - Cartography in ESPON*, ESPON, 24 p.

Table des figures

Figure 1 : Exemples d'analyse comparative des discontinuités démographiques (a) à la frontière Autriche - Tchécoslovaquie - Hongrie et (b) à la frontière France - Belgique (Decroly et Grasland, 1992)	6
Figure 2 : Exemple de démarche institutionnelle de partenariat entre instituts statistiques dans le massif des Pyrénées	7
Figure 3 : Etapes de construction de l'observatoire.....	11
Figure 4 : Variables de base brutes, variables composites, variables élaborées intégrant des traitements spatiaux (d'après ESPON M4D, 2011)	13
Figure 5 : Un modèle de gestion de métadonnées : le GSBPM (Generic Statistical Business Process Model).....	15
Figure 6 : Modèle de métadonnées proposé = modèle simplifié d'ESPON DB	16
Figure 7 : Étapes nécessaires à la mise en place d'un système d'information automatisé.....	17
Figure 8 : Choix du fond de carte des unités spatiales locales et jointure avec les données statistiques	20
Figure 9 : Le fond généralisé des régions européennes pour l'aménagement du territoire. Des fonds de cartes géoréférencés en accès libre depuis le site de l'UMS RIATE (licence Creative Commons)	21
Figure 10 : Différents modèles de gestion pour une interface de diffusion des données	24
Figure 11 : Part des travailleurs frontaliers parmi les actifs occupés.....	27
Figure 12 : Représentation graphique des valeurs centrales et extrêmes de la population active occupée (2010) selon plusieurs maillages territoriaux.....	29
Figure 13 : Comparaison des définitions nationales en Europe de zones d'emploi, selon plusieurs critères définis par l'étude de Coombes <i>et al.</i> , 2012.....	31
Figure 14 : Éléments de réflexion sur les enjeux, usages et moyens questionnés pour la mise en place d'un observatoire pérenne du transfrontalier	37
Figure 15 : Scénario « bas » pour la gestion et la maintenance d'un observatoire des espaces transfrontaliers.....	40
Figure 16 : Scénario « intermédiaire » pour la gestion et la maintenance d'un observatoire des espaces transfrontaliers.....	41
Figure 17 : Scénario « haut » pour la gestion et la maintenance d'un observatoire des espaces transfrontaliers.....	42

Table des cartes

Carte 1 : Choix du périmètre d'analyse pour l'étude	27
Carte 2 : Représentations cartographiques de la population active occupée (2010) selon plusieurs maillages territoriaux	29
Carte 3 : Représentations cartographiques du taux de chômage et de la densité de population par carroyage et, à titre de comparaison, selon des unités de taille équivalente.....	30
Carte 4 : Découpages territoriaux fonctionnels : les bassins d'emploi en Belgique, définis par l'IGEAT (2001) et les zones d'emploi INSEE en France.....	32
Carte 5 : Représentations cartographiques du taux de chômage selon deux types de découpages fonctionnels.....	34
Carte 6 : Zones d'emploi franco-belges, reconstruites selon la distance aux centres (gauche) et à partir d'un modèle de Reilly (droite).....	35

Table des tableaux

Tableau 1 : Comparaison de maillages institutionnels en zone transfrontalière franco-belge, à partir des statistiques descriptives.....	28
Tableau 2 : Comparaison des superficies des unités territoriales entre la Belgique et la France	28
Tableau 3 : Effets de la variation des seuils d'emplois et d'attractivité	34

Table des encadrés

Encadré 1 : Disponibilité des données sur l'emploi en France et en Belgique.....	26
Encadré 2 : Profondeur de la bande transfrontalière	27
Encadré 3 : Méthode de construction de zones d'emploi de l'IGEAT	33