



Numéro [numéro 120](#) · [Articles](#) · Mise en ligne : 04/2017

Recueil, traçabilité et restitution des données territoriales du programme ESPON

par : [Ronan Ysebaert](#), [Isabelle Salmon](#), [Benoit Le Rubrus](#) et [Camille Bernard](#)

Mots-clés : [base de données](#), [diffusion](#), [ESPON](#), [INSPIRE](#), [M4D](#), [outils](#), [traçabilité](#)

Le projet ESPON M4D consiste à collecter, vérifier, intégrer et restituer les données territoriales produites par les projets du programme de recherche appliquée ESPON. Les principaux enjeux de ce projet consistent en premier lieu à gérer la profusion de données hétérogènes, les normaliser et les harmoniser, évaluer leur qualité et assurer leur suivi. Ensuite, de créer des outils de suivi et de restitution de ces données. La mise en place de ces méthodes et outils adaptés apporte quelques solutions et pistes de réflexion pour le lecteur s'intéressant à la reproduction de cette expérience de projet dans un environnement similaire (montage d'observatoire, coordination d'acteurs producteurs de données multiples, etc.).

Gathering, verifying, and presenting territorial data from the ESPON Program

Abstract : The ESPON M4D Project entails gathering, verifying, integrating and presenting the territorial data produced by ESPON Applied Research Projects. The main challenges that the project faced were firstly managing, standardising and coordinating a wealth of diverse data, ensuring data quality and traceability. The second hurdle was the creation of tracking and display tools for territorial data. Implementing specialised methods and tools has provided solutions and opportunities for further analysis in the same field of applied research (data center creation, coordination of several data providers).

Keywords: database, ESPON, M4D, INSPIRE, tracking, tools, dissemination

Recogida, trazabilidad y restitución de datos territoriales del programa ESPON

Con el proyecto ESPON M4D se ha recogido, verificado, integrado y restituido la información territorial de los proyectos de investigación aplicada de ESPON. Sus principales retos consisten en poder gestionar datos muy heterogéneos, normalizarlos, armonizarlos, evaluar su calidad y asegurar su continuidad. Este método de trabajo y sus herramientas de trabajo aportan innovaciones y procedimientos de trabajo que pueden servir al lector para replicarlos en entornos similares (observatorios territoriales, coordinación de proveedores de información, etc.).

Palabras clave: Bases de datos, ESPON, M4D, INSPIRE, trazabilidad, herramientas, difusión

Contributeurs (conception et développement) : Camille Bernard, Jérôme Gensel, Timothée Giraud, Claude Grasland, Nicolas Lambert, Benoit Le Rubrus, Clément Louis, Christine Plumejeaud, Laurent Poulénard, Isabelle Salmon, Anton Telechev, Cassia Trojahn Dos Santos, Marlène Villanova-Oliver, Ronan Ysebaert.

Introduction

Le partage et la diffusion des données sont des sujets sensibles dans de nombreux projets de recherche en



réseau : restituer les données produites par de multiples acteurs de façon interopérable, compréhensible et normalisée constitue un des besoins fondamentaux du programme ESPON¹² (*European Observation Network for Territorial Development and Cohesion*).

Sur la période 2007-2014, ce programme européen gère de nombreux consortiums de recherche européens (TPG, *Transnational Project Group*) d'origines disciplinaires variées. Les thématiques de ces projets portent aussi bien sur des questions démographiques, environnementales ou socio-économiques. Les données produites dans ce cadre ont vocation à être diffusées largement à tous les acteurs du développement territorial (politiques, chercheurs, étudiants). En dépit de cette diversité, les TPG partagent des points communs : l'aire d'étude des données produites porte sur les territoires européens ; la Nomenclature des Unités Territoriales Statistiques (NUTS) est la maille territoriale sur laquelle portent la plupart des analyses — mais pas exclusivement (ESPON ECL, 2013) ; les données collectées doivent être délivrées au projet en charge de la base de données (ESPON M4D — *Multi Dimensional Database Design and Development*) au terme de l'activité du TPG ; les TPG ont généralement peu d'experts en métadonnées. Le projet ESPON M4D est chargé durant quatre ans (2011-2014) de la gestion et du développement de la base de données du programme ESPON. Une de ses principales activités consiste à collecter puis restituer les données et indicateurs produits par les TPG du programme à l'aide d'une application Web dédiée : *ESPON Database Portal*³, module logiciel de l'Infrastructure de Données Spatiales (IDS) *STeDI* développée dans le cadre du projet ESPON M4D (Bernard *et al.*, 2017). Les TPG ont le choix de fournir trois types de jeux de données, notamment en fonction du niveau de détails des métadonnées décrivant leur production: les *key indicators*, les *case studies*, et les *background data*. Dans cet article, nous concentrons notre attention sur les jeux de données du type le plus riche en termes d'information, les *key indicators*. Les *key indicators* contenus dans ces jeux de données constituent aussi les réalisations les plus emblématiques des TPG et sont les informations statistiques les plus diffusées hors ESPON (*Maps of the Month*⁴, *Posters*⁵, *Synthesis Reports*⁶, *Présentations*⁷).

Par le prisme de l'expérience acquise dans le cadre du projet M4D, l'ambition de cette communication consiste à sensibiliser le lecteur à certains éléments à considérer en priorité lors de la mise en place d'une infrastructure de données dont les caractéristiques principales sont les suivantes :

- de multiples contributeurs ;
- l'impossibilité de prévoir la nature des données livrées au projet M4D ;
- la nécessité de respecter les standards européens en vigueur (INSPIRE) ;
- des pratiques diverses en matière de saisie de données et métadonnées dans les différents TPG ;
- du côté de l'interface de restitution, un public potentiel large aux attentes multiples.

Après une première partie dédiée à la description des spécificités de la production statistique du programme ESPON, nous discutons du travail de normalisation réalisé par le biais de la mise en place d'un modèle de données et métadonnées dédié, diffusé à l'ensemble des TPG impliqués dans le programme. Puis, nous expliquons la procédure définie pour assurer la qualité et la cohérence des jeux de données intégrés en base. La dernière partie décrit comment l'application vise à restituer intelligemment et efficacement les données produites pour un large public.

Particularités de la base de données ESPON

Dans un cadre européen, la référence en matière de production statistique reste Eurostat. Sa mission principale consiste à exploiter et publier des informations statistiques comparables au niveau européen. À l'échelle régionale des NUTS, Eurostat propose ainsi au téléchargement un grand nombre d'indicateurs, organisés en 16 thématiques et disponibles sur une profondeur temporelle large (20 années pour les indicateurs les plus usités). Par cette masse d'information bien organisée et régulièrement mise à jour, Eurostat répond aux attentes d'un grand nombre d'utilisateurs : du chercheur qui souhaite bénéficier de



données de base permettant de mener des analyses statistiques ou économétriques poussées au décideur politique cherchant à situer le niveau de sa région ou de son pays au regard des objectifs politiques définis par l'Union européenne (Stratégie *Europe 2020*⁸).

En comparaison des principaux fournisseurs de données européens, ESPON se situe à un autre niveau : il s'agit d'un programme de recherche appliquée en aménagement du territoire européen financé par la Commission européenne et les États membres. Son objectif est d'étudier les principales tendances d'évolution du territoire européen au regard des objectifs fixés par la politique de cohésion de l'Union européenne. Son rôle est de créer des connaissances et des outils permettant de développer des politiques et des stratégies efficaces en matière de développement territorial. À travers la production d'analyses et de scénarios, ESPON permet d'échanger des informations, des résultats scientifiques et des bonnes pratiques utiles tant à la communauté scientifique qu'aux praticiens. Le but est aussi de proposer un outil d'aide à la décision politique qui propose des comparaisons à l'échelle de l'Union européenne. Ainsi, par l'intermédiaire du réseau de chercheurs que le programme mobilise, ESPON produit des ressources statistiques complémentaires de celles des autres fournisseurs de données, comme Eurostat ou l'Agence européenne pour l'environnement. Nous décrivons ici trois spécificités fondamentales pour situer ESPON dans le paysage européen de la production statistique.

Mobilisation d'objets géographiques et de sources hétérogènes

Les planches cartographiques multi-niveaux constituent une des spécificités du programme ESPON. Un même phénomène socio-économique tel que l'évolution démographique 1990-2010 (**figure 1**) est décliné sur plusieurs espaces d'étude (ici, l'Europe, l'Europe et ses voisinages, le monde). Les données servant à construire ces cartes reposent sur des mailles géographiques hétérogènes (Unités administratives locales, régions/NUTS2, États). Les sources d'information utilisées pour recueillir ces données sont multiples (Nations unies, Eurostat, Instituts statistiques nationaux). L'appariement de contextes d'analyse, d'objets géographiques et de sources d'information hétérogènes constitue une des singularités proposées par les équipes de recherche impliquées dans le programme ESPON. Cependant, la variabilité des maillages utilisés peut devenir un obstacle à la comparabilité des territoires inégalement subdivisés et causer des biais interprétatifs importants, phénomène formalisé par Stan Openshaw sous l'appellation *Modifiable Area Unit Problem* (MAUP). Depuis son lancement, le programme ESPON propose des pistes de réflexion pour y répondre. La combinaison des représentations multi-échelles et des techniques de lissage par potentiel produisent des visuels (**figure 1**) qui permettent de mieux appréhender les principales tendances spatiales pour un indicateur donné (ESPON 3.4.3, 2006).



Figure 1. Approche multi-niveau de l'évolution de population 1990-2010 développée dans le cadre du programme ESPON. Source: Lambert, Ysebaert, 2014 (in ESPON Synthesis Report, 2014). Traduit en français pour l'article.

La base de données du programme soutient donc la mise à disposition d'objets géographiques hétérogènes pour mener de telles analyses, comme les objets urbains. Actuellement, l'interface de restitution des données ESPON (voir [dernière partie](#)) permet le téléchargement de données fondées sur des nomenclatures d'objets urbains (**figure 2**). Ces nomenclatures sont composées selon différentes méthodes d'agrégation spatiale basées sur la maille des *Local Administrative Units*, équivalent européen des communes françaises (LAU2) (Guérois *et al.*, 2014): les UMZ (*Urban Morphological Zones*) et MUA (*Morphological Urban Areas*) sont ainsi construites selon des critères morphologiques (continuité du bâti, densité minimale, etc.). Alors que les FUA (*Functional Urban Areas*) sont définies par des critères fonctionnels (flux minimum de navetteurs vers des pôles d'emploi identifiés). À l'avenir, la base de données du programme ESPON restituera également des indicateurs reposant sur les LUZ (*Large Urban Zone*), autre objet urbain défini par des critères d'agrégation



fonctionnels.



Figure 2. Objets urbains mobilisés dans ESPON, l'exemple de Vienne (Autriche). Source : Pavard, 2012 (in ESPON M4D Newsletter, 2013). Traduit en français et modifié graphiquement pour l'article.

Ces confrontations d'approches et de contextes d'analyse proposés dans le programme ESPON sont le fruit de collaborations de chercheurs et d'institutions productrices de données de toute l'Europe. Elles permettent aussi de proposer une compréhension plurielle des dynamiques territoriales observées en Europe. Aussi, la base de données ESPON doit-elle rendre compte de l'hétérogénéité de cette production statistique et cartographique.

Des séries temporelles complètes à l'échelle des régions européennes

Au contraire d'Eurostat, le programme ESPON n'est pas tenu d'utiliser les données statistiques officielles pour mener ses analyses. Un travail de fond est réalisé dans le cadre du projet ESPON M4D pour proposer des séries temporelles longues à l'échelle des régions européennes pour quelques indicateurs de base (Charlton *et al.*, 2014 ; Grasland *et al.*, 2013). Cette activité mobilise le savoir-faire d'experts statisticiens en matière d'estimation de données manquantes. Ce travail d'estimation, dont l'intérêt est croissant, a d'ailleurs fait l'objet d'un appel d'offres spécifique et devrait être poursuivi et étendu à l'avenir. La question de la disponibilité des données est en effet souvent cruciale pour le lancement d'analyses quantitatives portant sur les dynamiques socio-économiques des régions européennes. Pour ne prendre qu'un exemple, la base de données ESPON propose une série complète de données de population, de 1990 à 2011 et à tous les niveaux de la nomenclature NUTS (0, 1, 2 et 3). Pour la même série statistique, le taux de complétude de la donnée Eurostat ne s'élève qu'à 64%. Cela s'explique par le fait que la collecte et l'estimation de données ne font pas partie des prérogatives d'Eurostat. « Cette tâche incombe aux instituts de statistique des États membres, qui vérifient et analysent les données nationales avant de les transmettre à Eurostat » (Eurostat, 2014). Sa mission consiste principalement à compiler et publier des informations statistiques comparables (produites par les Instituts statistiques nationaux) au niveau européen. Du fait des réformes territoriales successives (la nomenclature NUTS est en moyenne révisée tous les trois ans), les coupures dans les séries statistiques sont nombreuses. Il est dès lors complexe pour l'analyste de mener à bien des recherches à l'échelle régionale en Europe. Plusieurs projets du programme ESPON proposent des solutions pour remédier à ces problèmes de disponibilité de données.

Des indicateurs composites pour l'aide à la décision

Une autre caractéristique du programme ESPON réside dans la production d'indicateurs composites innovants. Cette production statistique est riche et multithématique : 25 projets de recherche appliquée ont été engagés pour la période de programmation 2007-2013. Ils donnent lieu à la création d'indicateurs synthétiques permettant d'éclairer le débat public européen. Les sources primaires des données utilisées sont nombreuses et de qualité (Eurostat⁹, Agence Européenne de l'Environnement¹⁰, Eurogeographics¹¹, United Nations Population Prospect¹², United Nations Environment Programme¹³, OCDE¹⁴, Instituts statistiques nationaux, etc.). Ces indicateurs sont construits selon des méthodes de travail issues de plusieurs champs disciplinaires : analyse spatiale (analyse multiscalaire, distance au réseau de transport, etc.), économétrie (modélisation, scénarios) ou géomatique (agrégation de grilles). Les **figures 3 et 4**, diffusées dans les séminaires et rapports de synthèse du programme ESPON, proposent deux représentations cartographiques qui illustrent la variété des données utilisées et produites. L'indice d'accessibilité multimodale produit dans le cadre du projet ESPON TRACC et par l'agence Spiekermann & Wegener (**figure 3**) propose par exemple une mesure synthétique de performance du réseau de transport européen. Il prend en compte un grand nombre de paramètres : réseaux



routier, ferroviaire et aérien et la population qu'elle permet d'atteindre par ces infrastructures. Sur une autre thématique, l'indice d'évolution d'utilisation du sol (**figure 4**), réalisé dans le cadre du projet ESPON EU-LUPA, utilise les grilles kilométriques de couverture du sol de l'Agence européenne de l'environnement disponible à différents pas de temps pour proposer une analyse évolutive à un niveau agrégé, le NUTS3. Ces deux indicateurs composites ne constituent qu'un extrait d'une production bien plus large.



Figure 3. Accessibilité multimodale potentielle, 2011. Source : ESPON Synthesis Report, projet TRACC, 2014. Traduit en français pour l'article.



Figure 4. Changement d'utilisation du sol, 1990-2006. Source : ESPON Synthesis Report, projet EU-LUPA 2014. Traduit en français pour l'article.

Ces caractéristiques soulignent la singularité d'un programme tel qu'ESPON en matière de production d'information à caractère géographique au niveau européen. Elles nécessitent la mobilisation d'un savoir-faire méthodologique qui implique souvent conjointement estimation de données, appariement de sources hétérogènes et création d'indicateurs complexes.

Le projet M4D dans l'environnement ESPON

Dans ce contexte de production de données multidimensionnelles à l'échelle d'un programme de recherche appliquée, le rôle du projet ESPON M4D (2011-2014) consiste à intégrer, vérifier et restituer les données territoriales produites par les projets du programme ESPON. Les principaux enjeux de ce projet consistent à gérer la profusion de données statistiques hétérogènes, les normaliser et les harmoniser, évaluer leur qualité, assurer leur traçabilité, créer des outils de suivi et de restitution de ces données. Le projet M4D a également mené une réflexion sur de la mise en cohérence des nomenclatures de données géographiques au niveau européen (objets urbains, création de nomenclatures comparables à la NUTS pour les voisinages européens, etc.). Plusieurs rapports techniques, jeux de données et outils ont ainsi été produits pour faire évoluer les connaissances sur la création de données innovantes et harmonisées dans un contexte européen. Les parties suivantes de cet article présentent plusieurs réalisations du projet M4D qui ont permis de répondre aux enjeux soulevés par ce contexte de production de données.

L'enjeu des métadonnées

Des standards à respecter

La pierre angulaire d'une gestion pérenne du flux d'information dans le programme ESPON repose sur la définition d'un modèle de données et de métadonnées adapté. Ce modèle sert tout d'abord à renseigner précisément les jeux de données qui sont délivrés pour les rendre intelligibles à un large public et à réutiliser ces informations afin d'en organiser la restitution. Dans le programme ESPON, la définition de ce modèle permet également de normaliser et d'harmoniser la mise en forme des informations attendues de la part des TPG. Des standards adaptés à la description et la diffusion de données géographiques (directive INSPIRE, normes ISO-19115 et ISO-19135) aident considérablement à la structuration de ce type d'information.



Adaptation de ces standards aux spécificités du programme ESPON

Ces standards sont cependant davantage utilisés pour décrire des données géographiques environnementales et peu adaptés aux spécificités de l'information statistique territoriale mobilisée dans le programme ESPON. En effet, les jeux de données créés présentent de manière schématique trois niveaux de description différents : le jeu de données (responsables, personnes contacts, modalités de diffusion), l'indicateur (code, nom, méthodes de mesure, unités de mesure) et les valeurs des indicateurs (sources, méthodes d'estimation/harmonisation sémantique de données, etc.) (Plumejeaud, 2010). Les champs obligatoires proposés par la norme ISO-19115 ne permettent pas de décrire efficacement l'information statistique à ce niveau de granularité. La norme ISO-19115 prévoit en revanche la possibilité d'être étendue à volonté dans un profil. C'est ce qui a été réalisé dans le projet ESPON M4D (et son prédécesseur *ESPON Database*, 2008-2011) avec l'implémentation du profil *esponMD*, qui définit l'extension de la norme ISO-19115 pour l'information socio-économique (Plumejeaud, 2010). Afin de rendre opérationnel ce profil à l'échelle du programme, un modèle tabulaire (fichier Excel) a été mis en œuvre pour faciliter sa prise en main par les TPG.

Ce modèle tabulaire est constitué de quatre onglets (**figure 5**) : informations relatives au jeu de données (*dataset*), aux indicateurs (*indicators*), à l'origine des données (*source*) et aux valeurs des indicateurs (*data*). Ce modèle propose une solution partagée à l'échelle du programme pour le référencement des données et métadonnées.

Afin de rendre plus aisée la complétion des métadonnées par les TPG dans le modèle tabulaire,

- certains champs optionnels non utiles à la description de l'information statistique ont été retirés du modèle ;
- certains champs obligatoires sont pré-remplis ;
- des champs sont ajoutés pour compléter l'information thématique : en particulier la description des méthodes de calcul des indicateurs ou le référencement des sources pour chaque combinaison d'indicateur à une date donnée pour chaque unité territoriale.

Par ailleurs, le référencement de données et métadonnées suivant le profil *esponMD* est suffisamment générique pour être mobilisé dans d'autres champs d'application que ceux du cadre européen. Une première tentative concluante a été entreprise dans le cadre d'une étude visant à évaluer la faisabilité du montage d'un observatoire des espaces transfrontaliers (UMS RIATE, CIST, CGET, 2014). Une telle entreprise suppose en effet la collecte, l'harmonisation et le référencement d'information statistique sémantiquement hétérogène. La solution proposée dans le cadre du programme ESPON a en ce sens été d'un apport notable.

Importance de documents explicatifs synthétiques et variés

Si la mise en place de standards et de procédures de contrôle des jeux de données a facilité le processus d'intégration, l'appropriation du modèle de métadonnées par les TPG et l'explication du processus d'intégration des données a nécessité du temps. Les occasions de rencontrer les TPG sont rares et veiller à la mise en application du modèle par les producteurs de données constitue un défi en soi. Le projet M4D opte ainsi pour une stratégie de diffusion plurielle, pédagogique et accessible tant sur la forme que sur le contenu : réalisation de documents au titre évocateur (*How to deliver my data?*¹⁵), présentations à chaque séminaire¹⁶, [création de lettres d'information semestrielles](#) et mise à disposition de [fichiers « exemples »](#) qui proposent un panel de bonnes pratiques.



Figure 5. Le modèle de métadonnées ESPON (format tabulaire), composé de quatre onglets.



Processus d'intégration: Le *tracking tool*

Gestion du flux d'information: le *tracking tool*

Afin d'assurer la conformité et la qualité des jeux de données délivrés par les TPG, un outil de suivi a été développé dans le cadre du projet M4D : le *tracking tool*, module de l'IDS *STeDI* (Bernard *et al.*, 2017). La vocation de cet outil est de répondre à plusieurs objectifs :

- contrôler la conformité des jeux de données au format attendu ;
- assurer la mise en cohérence de la description des indicateurs et la qualité des valeurs ;
- impliquer les différentes parties prenantes dans le processus d'intégration de façon transparente : les fournisseurs de données (TPG), l'administrateur de la base de données (M4D) et le commanditaire du programme (Unité de coordination ESPON) ;
- visualiser l'état d'avancement du processus d'intégration pour chaque jeu de données.

Le *tracking tool* est disponible en accès restreint depuis l'application Web (il faut faire partie d'un projet ESPON pour y accéder). Le processus d'intégration sous-jacent au *tracking tool* est constitué de cinq étapes dont trois phases de contrôle (**figure 6**) :



Figure 6. Le processus d'intégration des données des TPG dans ESPON.

- Le TPG dépose ses jeux de données sur le serveur. Une procédure automatique est lancée pour vérifier notamment que tous les champs obligatoires sont correctement renseignés (*syntactic check*, **figure 7**). Des messages d'erreurs informent le fournisseur des correctifs à apporter à son jeu de données.



Figure 7. Dépôt de fichier de données et vérification syntaxique sur l'interface de téléversement du portail ESPON Database.

- Une notification par courriel invite un expert thématique du projet M4D à procéder à l'analyse de la clarté et de la cohérence des métadonnées délivrées (*semantic check*), en particulier des champs de description textuels qu'il n'est pas possible de vérifier de façon automatique. Cette expertise donne lieu à un rapport que le TPG est invité à consulter (**figure 8**). Il décide alors de la suite à donner à son jeu de données : soumettre une nouvelle version pour améliorer sa qualité ou poursuivre l'intégration.



Figure 8. Extrait d'un rapport de vérification sémantique où un TPG est invité à apporter des correctifs à son jeu de données.

- Une fois le niveau de qualité sémantique assuré, une notification est envoyée aux statisticiens du projet M4D pour effectuer la recherche de valeurs exceptionnelles des indicateurs contenus dans le jeu de données (Charlton, 2014). [Un rapport](#) est généré suivant un processus semi-automatique combinant procédure programmée dans le langage R et expertise statistique des résultats obtenus. Ce rapport est ensuite soumis en ligne et disponible depuis le *tracking tool*. Ici encore, le TPG est invité à le consulter et à décider de la suite à donner aux opérations (abandon ou poursuite).
- L'unité de coordination ESPON consulte les rapports et décide en conséquence de l'intégration des



indicateurs dans la base de données ou non.

- Les jeux de données sont intégrés en base. L'interface de requête permet leur recherche et leur consultation.

Le *tracking tool* automatise la gestion de ce flux d'information complexe (**figure 9**). L'outil distribue les rôles aux acteurs concernés à chaque étape de l'intégration et permet une coordination plus efficace. L'ensemble des rapports est centralisé. Ainsi, le TPG se concentre sur la qualité et la cohérence de son jeu de données, la gestion du projet est facilitée et le commanditaire a une vision synthétique de l'ensemble du processus.



Figure 9. Le tracking tool : vue générale des jeux de données en cours d'intégration (gauche) ; vue détaillée d'un jeu de données et accès aux rapports (droite).

Interface de restitution de données

La collecte et l'harmonisation des données constituent une des tâches du projet M4D. Mais l'objectif principal reste la restitution à un large public des indicateurs et de leurs valeurs associées : praticiens, décideurs politiques, chercheurs, etc. L'interface de requête en ligne permet à l'utilisateur de rechercher des données. Si la richesse du contenu des métadonnées rend possible l'interrogation de la base de données selon de multiples critères, une approche consensuelle a été retenue pour être accessible au plus grand nombre.

L'interface est épurée afin de favoriser l'utilisation intuitive de l'application par des utilisateurs aux profils hétérogènes, ne partageant pas les mêmes pratiques de recherche ou n'ayant pas les mêmes objectifs quant à l'analyse du territoire.

La nature hétérogène du public visé a déterminé l'interface utilisateur. L'interface garantit la flexibilité des requêtes (recherche par défaut ou recherche avancée) et met en évidence les différentes dimensions du modèle de données. Un filtre sémantique permet la recherche d'indicateurs (**figure 10**, disponible à l'adresse suivante : <http://database.espon.eu/db2/search>) par :

- thème (démographie, emploi, etc.) ;
- documents de référence de la politique de cohésion de l'Union européenne (Agenda 2020, rapports sur la Cohésion, Agenda Territorial, etc.) ;
- projet fournisseur de données (ESPON TPG) ;
- mot-clé (nom d'indicateur, mots-clés dérivés du thesaurus GEMET¹⁷).

Les filtres *data*, *what*, *where*, *when* offrent la possibilité de restreindre les résultats retournés lors d'une recherche sémantique :

- Le filtre *data* permet de requêter les indicateurs selon leur propriété statistique (données quantitatives de stock, ratios, typologies, etc.) ou selon le type de jeu de données auxquels ils appartiennent: indicateur simple, série temporelle ou tables de contingences (pyramides des âges ou structure de l'emploi par branche d'activité, par exemple).
- Le filtre *where* peut être utilisé pour sélectionner des aires d'études spécifiques (Union européenne des 28, espace ESPON¹⁸ ou un État membre en particulier). Le taux de complétude (*completeness*) est calculé en fonction de l'aire d'étude sélectionnée et permet à l'utilisateur d'évaluer si les données recherchées sont globalement disponibles ou non. C'est aussi par le filtre *where* qu'il est possible de définir le type d'objet géographique recherché: donnée régionale (NUTS 0, 1, 2 ou 3) ou urbaine (UMZ, FUA ou MUA).



- Le filtre *what* permet d'ajouter à la requête initiale un critère sémantique supplémentaire.
- Le filtre *when* permet de restreindre la requête à une période de temps ou une année particulière.

Libre alors à l'utilisateur, selon l'analyse qu'il souhaite effectuer, de découvrir des indicateurs portant sur des objets urbains, en paramétrant le filtre *where*, d'observer des séries longues à partir du filtre *data*, ou d'affiner sa recherche temporelle en activant le filtre *when*.



Figure 10. Restitution des résultats – Interface de recherche.

L'attribution consciencieuse de métadonnées aux jeux de données et aux indicateurs les composant, forme un modèle de données riche, accessible selon trois dimensions : le temps, l'espace et la thématique. La souplesse du système est censée répondre aux attentes de différents profils utilisateurs : politique, scientifique ou praticien. Nous proposons ici trois scénarios d'utilisation appelant différentes fonctionnalités de l'interface de restitution. Pour alimenter la démonstration, quelques post-traitements cartographiques et statistiques de base sont mobilisés. Ces scénarios illustrent de façon concrète trois comportements utilisateurs auxquels les concepteurs ont dû répondre pour penser et structurer cette interface de restitution. Les valeurs et indicateurs présentés ci-dessous ont été établis à partir de l'interface de restitution des données¹⁹ consultée en septembre 2014. Le lecteur qui souhaite reproduire l'expérience obtiendra certainement d'autres résultats puisque la base de données est constamment alimentée par les données des TPG.

Scénario 1 — Un chercheur souhaite disposer de données de base sur le temps long afin d'établir des scénarios prospectifs de population. Il peut dans un premier temps rechercher par thème en sélectionnant l'entrée « *population and living conditions* ». Il constate alors que le programme ESPON propose un grand nombre d'indicateurs sur la thématique puisque 231 résultats sont retournés par la requête. Dès lors, il a la possibilité d'activer le filtre *data* et l'option *time series*. L'activation du filtre lui permet de constater que la base de données contient une pyramide des âges pour la période 2000-2009 dont le taux de complétude avoisine les 100%. Il peut également bénéficier des données de naissances et décès pour la période 2000-2010 et celle de population sur la période 1990-2011. Une fois les données téléchargées il remarque, après avoir utilisé un fond de carte adapté²⁰ et après avoir créé un indice d'évolution, que certaines valeurs paraissent surprenantes. L'analyse des sources du jeu de données lui permet alors de comprendre la méthode d'estimation ou la source utilisée pour collecter la donnée. Le cas échéant, il peut se reporter aux documents méthodologiques associés aux sources. En cas de doute, il peut aussi contacter le producteur du jeu de données afin de procéder à son analyse en connaissance de cause.

<https://mappemonde.mgm.fr/wp-content/uploads/temp/scenario1.mp4>

Scénario 2 — Un représentant français spécialiste des questions d'aménagement du territoire européen s'intéresse au positionnement des régions françaises au regard des objectifs définis par l'Union européenne. Il trouve des éléments de réponse en activant la recherche par mot clé politique. Il peut par exemple sélectionner un des piliers de la Stratégie « Europe 2020²¹ », *inclusive growth* et consulter l'ensemble des indicateurs produits par le programme ESPON sur cette thématique. Il a alors la possibilité de télécharger l'indicateur *Europe 2020 index*, qui synthétise le positionnement de chacune des régions au regard des objectifs de croissance définis par l'Union européenne. Après un succinct traitement statistique, il peut constater qu'au regard des 315 NUTS2 européens, la situation des régions françaises est assez disparate en 2010 : la région la moins bien positionnée est la Corse (219^e) ; la région la mieux classée est Midi-Pyrénées (16^e place). En analysant les dynamiques observées sur cinq ans, il remarque que la région Franche-Comté a



fortement progressé (gain de 55 places dans la hiérarchie des régions européennes). Dans le même temps, la région Picardie a significativement régressé (perte de 52 places). Cette rapide analyse alimente son rapport sur la dynamique des territoires qu'il remettra au Commissariat général pour l'égalité des territoires (CGET).

<https://mappemonde.mgm.fr/wp-content/uploads/temp/scenario2.mp4>

Scénario 3 — Un membre de la Commission européenne, de la Direction générale mobilité et transports, peut par l'intermédiaire de la recherche par thème sélectionner l'entrée « *transport and accessibility* » et disposer de l'ensemble des indicateurs produits par les TPG sur cette thématique à une granularité géographique fine (NUTS3). Il retrouve ainsi aisément les valeurs d'accessibilité multimodale présentées [plus haut](#) qu'il peut remobiliser dans le cadre de ses travaux. L'utilisation du filtre *when* lui permet de constater que le programme ESPON propose ces indicateurs à plusieurs pas de temps : 1997, 2001, 2006 et 2011. À la lecture des métadonnées, il observe que les principaux projets contributeurs de cette thématique sont TRACC (*TRansport Accessibility at regional/local scale and patterns in Europe*), TIPTAP (*Territorial Impact Assessment*) et SeGi (*Services of General Interest*). Il peut dès lors se reporter aux rapports scientifiques produits par ces projets, disponibles sur le site du programme ESPON et découvrir leur production cartographique, sur laquelle il pourra éventuellement s'appuyer pour guider les décisions politiques prises au niveau de la Commission européenne.

<https://mappemonde.mgm.fr/wp-content/uploads/temp/scenario3.mp4>

Les possibilités offertes par cette interface ont théoriquement été pensées pour convenir aux habitudes d'un large panel d'utilisateurs, qui correspond au public ciblé par le programme ESPON. Néanmoins, idéalement, l'analyse de l'activité des utilisateurs finaux permettrait de mettre en évidence des comportements d'utilisation types et d'adapter l'interface de recherche en conséquence.

Conclusion

Dans le cadre du programme ESPON et du projet M4D, le modèle de métadonnées et le processus de mise à jour de la base de données ont été mis en place pour rendre intelligible et accessible une information complexe. En ce sens, cette base de données est un outil qui contribue au débat social et citoyen en donnant un accès facile à des indicateurs portant sur la politique de cohésion de l'Union européenne. Ces indicateurs sont précisément documentés afin que leur méthodologie puisse être reproduite et mise à jour, voire discutée ou remise en question.

La base de données ESPON devient une pièce stratégique du programme, non seulement interrogeable depuis l'application *ESPON Database Portal*, mais aussi par d'autres outils d'analyse (*HyperAtlas*) et de visualisation (*ESPON Online Mapping Tool*). À terme, des services Web d'accès aux métadonnées (*Catalog Service for the Web - CSW*) et aux données (*Web Feature Service - WFS*) permettraient la diffusion du contenu de la base ESPON à d'autres infrastructures de données spatiales, à des applications tiers et autres initiatives d'exploitation et de représentation des indicateurs et de leurs métadonnées.

Si le contenu de la base de données rend bien compte de la richesse du programme ESPON, il combine aussi des indicateurs aux caractéristiques hétéroclites : résultats des recherches menées par les TPG (typologies, indices composites) qui constituent la mémoire du programme ESPON ; indicateurs de long terme (séries temporelles multi-objets) qui ont vocation à être maintenus et mis à jour ultérieurement. Dès lors, pour l'utilisateur extérieur peu familier des productions statistiques du programme, cette variété rend parfois complexes l'interprétation et la synthèse du contenu de la base de données. Pour garantir une meilleure



visibilité de cette information statistique aux contributeurs multiples, la mise en place d'une réflexion de fond pour orchestrer la diffusion de ces données à l'échelle du programme nous semble particulièrement stratégique. Des propositions ont d'ailleurs été réalisées en ce sens par le projet ESPON M4D dans son rapport final (ESPON M4D, 2014).

Par ailleurs, l'expérience de la base de données M4D a permis de nombreuses avancées en matière d'organisation, de partage et de diffusion de données complexes. Pour que ces avancées ne restent pas lettre morte, la question de la pérennité et la reprise de la base de données à l'échelle de la coordination du programme est maintenant une question qui doit se poser, les activités de développement informatique et de maintenance du projet M4D se sont terminées en décembre 2014. L'absence de structure centrale pérenne a un impact à court terme sur le suivi de la qualité des données. Actuellement, ce processus est semi-automatique, puisque la plupart des vérifications qualitatives sont réalisées par des opérateurs humains. Que se passera-t-il à la fin du projet sans personnes ressources ? Une solution optimale consisterait à faire reposer cette plateforme technique sur une structure centrale forte (de type *data center*) à l'échelle d'un programme. Aussi serait-il nécessaire de doter une telle structure technique de ressources humaines capables d'administrer le contenu de ce type de base de données et d'en définir les enjeux et objectifs à long terme. Il faut néanmoins être réaliste : comme dans toute recherche financée sur programme, les infrastructures de données développées ont généralement une existence limitée dans le temps, ce qui est parfois préjudiciable et contre-productif au regard des moyens humains et financiers investis dans ce type d'initiative.

Au-delà des questions soulevées par la pérennité de cette infrastructure de données en particulier, l'expérience du projet M4D a permis d'identifier des pistes de réflexion qui conduisent à améliorer l'organisation, la gestion et la diffusion des informations statistiques produites à l'échelle d'un programme de recherche en réseau tel que le programme ESPON. Quelles que soient les orientations futures du programme, nous espérons que les nouveaux projets qui seront lancés pour gérer et diffuser les données des prochains TPG (2015-2020) bénéficieront de l'expérience accumulée et des solutions apportées par le projet M4D. En plus d'une solution logicielle opérationnelle (du recueil à la restitution des données), M4D a aussi proposé des méthodes qui permettent de répondre aux spécificités de l'information statistique et à une organisation qui repose sur des contributeurs multiples.

Bibliographie

BERNARD C., VILLANOVA-OLIVER, M. GENSEL, J., LE RUBRUS B. (2017). « Spatio-Temporal evolutive Data Infrastructure : a Spatial Data Infrastructure for managing data flows of Territorial Statistical Information ». *International Journal of Digital Earth*, n°10, p. 257-283.

CHARLTON M., BRUNSDON C., CAHALANE C., PFORTE L. (2014). « Time-series data in the ESPON Database ». *ESPON M4D Technical Report*, 122 p. En ligne : <http://database.espon.eu/db2/resource?idCat=31>

CHARLTON M., HARRIS P., CAIMO A., CAHALANE C. (2014). « Data quality check : Methods and procedures ». *ESPON M4D Technical Report*, 62 p. En ligne : <http://database.espon.eu/db2/resource?idCat=31>

ESPON (2014). « Third ESPON 2013 Synthesis Report 'Territories finding a New Momentum : Evidence for Policy Development, Growth and Investment' ». ESPON, Luxembourg, 68 p. En ligne : http://www.espon.eu/main/Menu_Publications/Menu_SynthesisReports/

ESPON 3.2. (2006). « Data Navigator 2, part 1 - Handbook for data collection ». *Final Report*, ESPON, Luxembourg, 153 p. En ligne : http://www.espon.eu/export/sites/default/Documents/Projects/ESPON2006Projects/CoordinatingCrossThematicProjects/Scenarios/fr-3.2-DN2_Final_Jan2007.pdf



ESPON 3.4.3 (2006). « The Modifiable Area Unit Problem ». *Final Report*, ESPON, Luxembourg, 254 p. En ligne : http://www.espon.eu/export/sites/default/Documents/Projects/ESPON2006Projects/StudiesScientificSupportProjects/MAUP/espon343_maup_final_version2_nov_2006.pdf

ESPON CARTOGRAPHIC LANGUAGE (2013). « Review report ». ESPON, Luxembourg, 130 p. En ligne : http://www.ums-riate.fr/Webriate/wp-content/uploads/2014/04/ECLReport_May2013_Task123_ok.pdf

ESPON DATABASE (2011). « ESPON 2013 Database, quality rather than quantity ». *Final Report*. ESPON, Luxembourg, 74 p. En ligne : https://www.espon.eu/export/sites/default/Documents/Projects/ScientificPlatform/ESPONDatabase2013/FR_for_the_website/M4D_Final_Report_March2011.pdf

ESPON M4D (2013). *Newsletter n°3*. ESPON, 6 p.

ESPON M4D (2014). *ESPON Database 2013 Phase II (2011-2014), Multi Dimensional Database Design and Development, Draft Final Report*. ESPON, Luxembourg, 80 p. En ligne : http://www.espon.eu/main/Menu_Projects/Menu_ESPON2013Projects/Menu_ScientificPlatform/espondatabase2013phaseII.html

ESPON M4D (2014). « ESPON 2013 Database Dictionary of Spatial Units ». 40 p. En ligne : <http://database.espon.eu/db2/dico?idDoc=66>

EUROPEAN COMMISSION, INSPIRE DIRECTIVE (2007). « Infrastructure for Spatial Information in the European Community ». 14 mars 2007.

EUROPEAN COMMISSION, COMMISSION REGULATION (EC) (2008). *N° 1205/2008 of 3 December 2008 implementing Directive 2007/2/EC of the European Parliament and of the Council as regards metadata*.

GRASLAND C., YSEBAERT R., GIRAUD T., CHARLTON M., CAIMO A., MATHIAN H. (2013). « A strategy for elaboration and update of coherent time series of hierarchical territorial units ». *ESPON M4D Technical Report*, 45 p. En ligne : <http://database.espon.eu/db2/resource?idCat=32>

GUÉROIS M., BRETAGNOLLE A., MATHIAN H., PAVARD A. (2014). « Functional Urban Areas (FUA) and European harmonization. A feasibility study from the comparison of two approaches : commuting flows and accessibility isochrones ». *ESPON M4D Technical Report*, 35 p. En ligne : https://www.espon.eu/export/sites/default/Documents/Projects/ScientificPlatform/ESPONDatabaseII/M4D-DFR_TR-FUA-construction_20140630.pdf

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION (ISO) (2014). *Geographic information - Metadata. ISO19115*.

PLUMEJEAUD C., GENSEL J., VILLANOVA-OLIVER M. (2010). « Opérationnalisation d'un profil ISO 19115 pour des métadonnées socio-économiques ». Actes du 28^e Congrès INFORSID, Marseille, 25-28 mai 2010. En ligne : https://liris.cnrs.fr/inforsid/sites/default/files/2010_plumejeaud.pdf

TELECHEV A., LE RUBRUS B. (2014). « ESPON Data and Metadata Specification ». ESPON M4D Project, 123 p. En ligne : <http://database.espon.eu/db2/metaspécifs?idDoc=61>

UMS RIATE, CIST (2014). « Étude de faisabilité d'une observation pérenne des espaces transfrontaliers ». CEGET, 56 p. En ligne : <http://www.ums-riate.fr/Webriate/?portfolio=etude-de-faisabilite-pour-la-creation-dun-observatoire-du-transfrontalier>



YSEBAERT R., LE RUBRUS B. (2012). « How to deliver my data ? ». ESPON M4D Project, 52 p. En ligne : <http://database.espon.eu/db2/htdmd?idDoc=65>

Notes

1. ↑ <http://www.espon.eu/>
2. ↑ Les liens hypertexte présents dans le corps du texte ont été vérifiés le 22 février 2017.
3. ↑ <http://database.espon.eu/>
4. ↑ http://www.espon.eu/main/Menu_Publications/Menu_MapsOfTheMonth/
5. ↑ http://www.espon.eu/main/Menu_Publications/Menu_Posters/
6. ↑ http://www.espon.eu/main/Menu_Publications/Menu_SynthesisReports/
7. ↑ https://www.espon.eu/main/Menu_Events/
- 8, 21. ↑ http://ec.europa.eu/europe2020/index_fr.htm
9. ↑ <http://epp.eurostat.ec.europa.eu/>
10. ↑ <http://www.eea.europa.eu/fr>
11. ↑ <http://www.eurogeographics.org/>
12. ↑ <http://esa.un.org/wpp>
13. ↑ <http://geodata.grid.unep.ch/>
14. ↑ <http://stats.oecd.org/>
15. ↑ http://database.espon.eu/db2/jsf/HowToDeliverMyData/HowToDeliverMyData_html/index.html
16. ↑ http://www.espon.eu/export/sites/default/Documents/Events/InternalSeminars/PaphosDecember2012/09.m4d_tracking_tool.pdf
17. ↑ <https://www.eionet.europa.eu/gemet/>
18. ↑ Espace incluant les États membres de l'Union européenne, la Suisse, le Liechtenstein, la Norvège et l'Islande.
19. ↑ <http://database.espon.eu/db2/search>
20. ↑ http://www.ums-riate.fr/Webriate/?page_id=153