HAL
open science

Recueil, traçabilité et restitution des données territoriales du programme ESPON

Ronan Ysebaert, Isabelle Salmon, Benoit Le Rubrus, Camille Bernard

To cite this version:

Ronan Ysebaert, Isabelle Salmon, Benoit Le Rubrus, Camille Bernard. Recueil, traçabilité et restitution des données territoriales du programme ESPON. CIST2014 - Fronts et frontières des sciences du territoire, Collège international des sciences du territoire (CIST), Mar 2014, Paris, France. pp.387-393. hal-01353418

HAL Id: hal-01353418

https://hal.science/hal-01353418v1

Submitted on 11 Aug 2016

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.

Recueil, traçabilité et restitution des données territoriales du programme ESPON

AUTEURS [1]
Ronan YSEBAERT, RIATE (France)
Isabelle SALMON, RIATE (France)
Benoit LE RUBRUS, LIG–STEAMER (France)
Camille BERNARD, LIG–STEAMER (France)

RÉSUMÉ
Le projet ESPON M4D *Multi Dimensional Database Design and Development* consiste à intégrer, vérifier et restituer les données territoriales produites par les projets de recherche appliquée du programme européen ESPON. Les principaux enjeux de ce projet consistent à gérer la profusion de données hétérogènes, les normaliser et les harmoniser, évaluer leur qualité, assurer leur traçabilité, créer des outils de suivi et de restitution de ces données. La mise en place de tels méthodes et outils adaptés apportent quelques solutions et pistes de réflexion. Force est de constater que la centralisation de la gestion et du traitement de ces données aide à leur restitution à un très large public.

MOTS CLÉS
Base de données, ESPON, M4D, INSPIRE, traçabilité, outils, diffusion

ABSTRACT
The ESPON M4D Multi-Dimensional Database Design and Development Project entails integrating, verifying and presenting the territorial data produced by ESPON Applied Research Projects. The main challenges that the project faced were managing, standardising and coordinating a wealth of diverse data, ensuring data quality, traceability and creating display tools for territorial data. Implementing specialised methods and tools has provided solutions and opportunities for further analysis. It is clear that centralising the data management and processing helps a wider audience to access this data.

KEYWORDS
Database, ESPON, M4D, INSPIRE, tracking, tools, dissemination

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 ESPON [2].

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

2 European Observation Network for Territorial Development and Cohesion, programme financé par la Commission européenne (Fonds de Développement Régional – Objectif 3 Coopération Territoriale) : www.espon.eu.

Sur la période 2007-2013, ce programme européen gère de nombreux consortiums de recherche européens (ESPON *Transnational Project Group* - TPG) 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 d'importance : l'aire d'étude des données produites porte sur les territoires européens (NUTS et objets urbains) [3] ; 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 en charge 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 interface web dédiée [4].

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*.

Le travail de normalisation réalisé grâce à la mise en place d'un modèle de données et métadonnées dédié est le premier point exposé ici. 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'interface vise à restituer intelligiblement et efficacement les données produites pour un large public.

1. 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 aussi 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 aident considérablement à la structuration de ce type d'information (directive INSPIRE, normes ISO-19115 et ISO-19135). Ces standards sont cependant davantage préfigurés pour décrire des données géographiques environnementales de type matriciel, ou *raster* [5], et peu adaptés à des données reposant sur des nomenclatures d'unités territoriales comme les NUTS.

3 80 % des cartes produites dans les rapports finaux du programme ESPON reposent sur la Nomenclature des Unités Territoriales Statistiques (NUTS) (ESPON Cartographic Language project, 2013).

4 Portail de la base de données ESPON : database.espon.eu.

5 Données présentées sous forme de pixels géoréférencés généralement issues de la télédétection, utilisés par exemple par l'Agence européenne de l'Environnment pour décrire la couverture du sol en Europe (LandUse).

Adaptation de ces standards aux spécificités du programme ESPON
La première étape pour le projet M4D fut de réaliser un modèle de données/métadonnées respectant les standards internationaux de la diffusion de métadonnées et adapté aux besoins du programme (Plumejeaud, 2010). Afin de rendre plus aisée la complétion des métadonnées par les TPG, la directive INSPIRE a été adaptée :
– les champs optionnels non utiles aux données territoriales ont été retirés du modèle ;
– certains champs obligatoires mais redondants sont pré-remplis ;
– les champs obligatoires sont clairement identifiables ;
– 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.

Au final, un modèle au format tabulaire a été fourni aux ESPON TPG. Ce document est constitué de quatre onglets [6] : informations relatives au jeu de données (*dataset*), aux indicateurs (*indicators*), à l'origine des données (*source*) et aux valeurs des indicateurs (*data*).

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. Le projet M4D opte ainsi pour une stratégie de diffusion plurielle, pédagogique et simplifiée tant sur la forme que sur le contenu : réalisation de documents au titre évocateur (*How to deliver my data?*), présentations à chaque séminaire, création de lettres d'information semestrielles et mise à disposition de fichiers « modèle » proposant un panel de bonnes pratiques.

2. 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 intégrés en base, un outil de suivi a été développé dans le cadre du projet M4D : le *tracking tool*. 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 en temps réel l'état d'avancement du processus d'intégration.

Ce dernier est disponible en accès restreint depuis le site web dédié. La figure 1 illustre le processus d'intégration sous-jacent au *tracking tool*, constitué de cinq étapes :
1. Le fournisseur (TPG) dépose ses jeux de données sur un espace de téléchargement. Une procédure automatique est lancée pour vérifier notamment que tous les champs obligatoires sont correctement renseignés (*syntactic check*).
2. Une notification par courriel invite un thématicien du projet M4D à procéder à la vérification de la clarté et de la cohérence des métadonnées délivrées (*semantic check*), en particulier des champs de description textuels. Cette expertise donne lieu à un rapport que

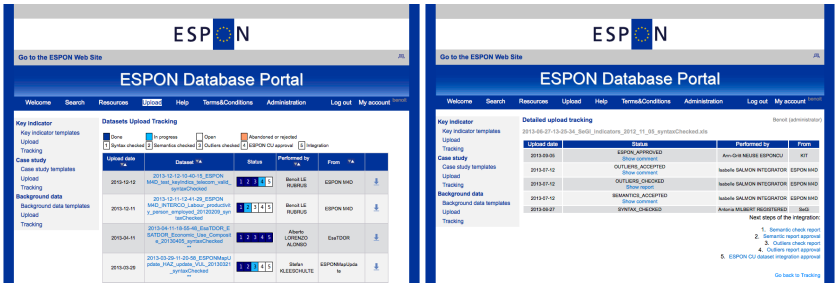
6 Portail de la base de données ESPON, section *Help* : database.espon.eu/db2/metaspecifs.

le TPG est invité à consulter. 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.
3. Une fois le niveau de qualité sémantique assuré, une notification est envoyée aux statisticiens du projet M4D pour effectuer la recherche de valeurs exceptionnelles (Charlton, 2012). Un rapport est généré et déposé sur 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).
4. L'unité de coordination ESPON consulte les rapports et décide en conséquence de l'intégration des indicateurs dans la base de données ou non.
5. Les jeux de données sont intégrés en base. L'interface de requête permet leur recherche et leur consultation.

Figure 1. Légende du *tracking tool* : étapes et états

Le tracking tool automatise la gestion de ce flux d'information complexe (figure 2). 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 2. 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)



3. INTERFACE DE RESTITUTION DE DONNÉES

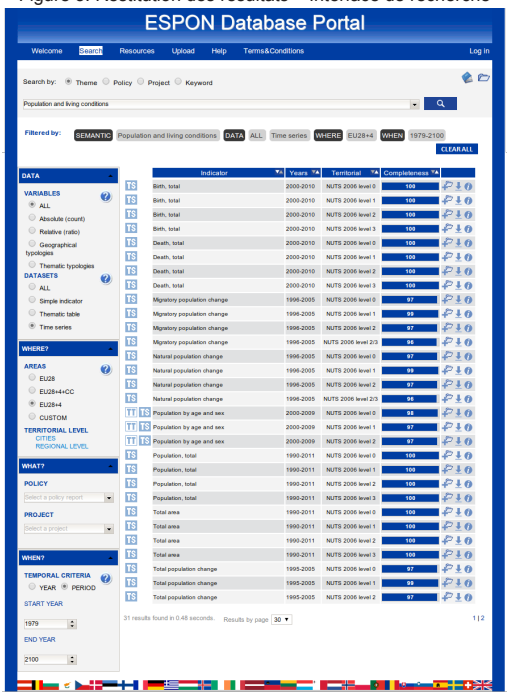
La collecte et l'harmonisation des données constituent une des phases du projet M4D. Mais l'objectif principal reste la restitution, à un large public, des indicateurs et de leurs valeurs associées : praticiens, décideurs politiques, chercheurs, etc. L'interface de requête en ligne permet à l'utilisateur 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 nécessairement 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. Celle-ci est modulaire, 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 3) 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.

Figure 3. Restitution des résultats – Interface de recherche



Les modules *data*, *what*, *where*, *when* offrent la possibilité de filtrer les résultats retournés lors d'une recherche sémantique. 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*.

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.

L'analyse de l'activité des utilisateurs finaux, dans une dernière phase du projet, permettra de mettre en évidence des comportements d'utilisation et d'évaluer les représentations visuelles du modèle de données, mises en place au sein de l'interface.

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'interface de requête, 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) permettront 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 l'hétérogénéité du contenu de la base rend bien compte de la richesse du programme ESPON, elle souligne aussi la difficulté à interpréter et synthétiser ce contenu pour des utilisateurs peu familiers des productions du programme. Cela souligne la nécessité de disposer au préalable d'une réelle vision stratégique à long terme sur le contenu et l'organisation de la base.

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. 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.

RÉFÉRENCES

Charlton M., Harris P., Caimo A., 2012, "Detecting and handling anomalous data in M4D", ESPON M4D Project.

European Commission, INSPIRE Directive, 2007, "Infrastructure for Spatial Information in the European Community", March 14th, 2007.

European Commission, Commission Regulation (EC), 2008, n° 1205/2008, December 3rd, 2008,

implementing Directive 2007/2/EC of the European Parliament and of the Council as regards metadata.
International Organization for Standardization (ISO), 2013, Geographic information – Metadata, ISO 19115.
Plumejeaud C., Gensel J., Villanova-Oliver M., 2010, « Opérationnalisation d'un profil ISO 19115 pour des métadonnées socio-économiques », actes du 28e congrès INFORSID, Marseille.
Telechev A., Le Rubrus B., 2013, "ESPON Data and Metadata Specification", ESPON M4D Project.
Ysebaert R., Le Rubrus B., 2012, "How to deliver my data?", ESPON M4D Project.

LES AUTEURS

Ronan Ysebaert
RIATE
ronan.ysebaert@ums-riate.fr

Isabelle Salmon
RIATE
isabelle.salmon@ums-riate.fr

Benoit Le Rubrus
LIG STEAMER
benoit.le-rubrus@imag.fr

Camille Bernard
LIG STEAMER
camille.bernard@imag.fr