

APPROCHE INTEGREE ET MULTICRITERES DE LA MODELISATION TERRITORIALE ELEMENTS DE CADRAGE POUR LA GESTION DES DONNEES TERRITORIALES

Rapport d'étude

Janvier 2016
Mise à jour juin 2016

N° de contrat : 1417C0006

Étude réalisée pour le compte de l'ADEME par : I Care & Consult

Coordination technique ADEME : MARQUET Sarah – Direction\Service : VALBONNE DVTD SOU



RAPPORT D'ETUDE

En partenariat avec :



icare
& consult

ThéMA

UNIVERSITÉ
LUMIÈRE
LYON 2
UNIVERSITÉ DE LYON

exosys



e-business agency

Euréval

REMERCIEMENTS

Membres du Comité de pilotage

- Yves MOCH (ADEME SB)
- Solène MARRY (ADEME SOU)
- Séverine BOULARD (ADEME STM)
- Nathalie MARTINEZ (ADEME STM)
- Sophie DEBERGUE (ADEME SOU)
- Amandine CRAMBES (ADEME SOU)
- Gabriel PLASSAT (ADEME STM)

Membres du groupement

- Boris BAILLY (I Care Environnement)
- Charlotte SUAUD (I Care Environnement)
- Léone-Alix MAZAUD (I Care Environnement)
- Marie-Hélène DE SEDE MARCEAU (Laboratoire ThéMa)
- Matthieu NOUCHER (Laboratoire PASSAGES)
- Hubert FAUCHEUX (Exosys)
- Eric Monnier (Euréval)

Membres du Comité scientifique

- Anne CHENE (ADEME SEP)
- Olivier DENNERY (Efficacy)
- Fabien LEURENT (GIS-MU)
- Olivier TOURNAIRE (CSTB)
- Julie RIEG (CHRONOS)
- Mathieu SAUJOT (IDDRI)
- Marc BOITEL (ARENE IDF)
- Emmanuel RIVIÈRE (ATMO Alsace)
- Pierrick YALAMAS (RAEE)

CITATION DE CE RAPPORT

ADEME. Boris BAILLY, Hadrien HAINAUT, Léone-Alix MAZAUD, Charlotte SUAUD. 2016. Approche intégrée et multicritère de la modélisation territoriale – Eléments de cadrage pour la gestion des données territoriales – *Rapport d'étude. 82 pages.*

Cet ouvrage est disponible en ligne www.ademe.fr, rubrique Médiathèque (<http://www.ademe.fr/mediatheque>)

Toute représentation ou reproduction intégrale ou partielle faite sans le consentement de l'auteur ou de ses ayants droit ou ayants cause est illicite selon le Code de la propriété intellectuelle (art. L 122-4) et constitue une contrefaçon réprimée par le Code pénal. Seules sont autorisées (art. 122-5) les copies ou reproductions strictement réservées à l'usage privé de copiste et non destinées à une utilisation collective, ainsi que les analyses et courtes citations justifiées par le caractère critique, pédagogique ou d'information de l'œuvre à laquelle elles sont incorporées, sous réserve, toutefois, du respect des dispositions des articles L 122-10 à L 122-12 du même Code, relatives à la reproduction par reprographie.

TABLE DES MATIERES

Résumé	5
Introduction	6
1. Panorama des outils de modélisation et d'observation territoriale	8
1.1. Un recensement large des outils existants.....	8
1.1.1. Qu'est-ce qu'un outil ?	8
1.1.2. Le recensement : plus de 170 outils	8
1.1.3. Le constat d'une vaste diversité d'outils	8
1.1.4. Premiers critères de sélection des outils	9
1.2. Organisation et typologie des outils.....	10
1.2.1. Le caractère intégré des outils.....	10
1.2.2. Le caractère multicritères des outils	10
1.2.3. Perspectives complémentaires	11
1.2.4. Vingt-huit outils présélectionnés	12
2. Etudes de cas approfondies de 11 outils de modélisation et d'observation territoriale	13
2.1. Sélection d'un échantillon d'outils pour analyse approfondie.....	13
2.2. Méthodologie d'analyse	16
2.3. Le tableau synoptique des outils	17
3. Les données territoriales de l'environnement : définitions, acteurs et législation	18
3.1. Que sont les données territoriales ?.....	19
3.2. Les données territoriales : une « matière première » mobilisée par de nombreux acteurs	23
3.2.1. Produire et collecter la donnée : une distinction de plus en plus ténue	23
3.2.2. La consommation des données : qui, quelles données, pour quoi faire ?	24
3.2.3. Qui sont les transformateurs de données ?.....	24
3.2.4. Quels sont les services liés à la donnée territoriale ?	24
3.3. Le cadre juridique de la collecte et de l'utilisation de la donnée territoriale	28
3.3.1. Un premier cadre : la loi sur le secret statistique.....	28
3.3.2. Les perspectives d'évolution de la juridiction française	28
3.3.3. La législation européenne.....	30
4. Le cycle de vie des données territoriales	32
4.1. La collecte primaire ou captage, obtention initiale des données.....	32
4.1.1. En quoi ça consiste ?	32
4.1.2. Quelles sont les méthodes de captage ?.....	32
4.1.3. Enjeux techniques, économiques et juridiques associés à la collecte primaire des données	35
4.2. La collecte secondaire : réutilisation de données existantes	38

4.2.1.	En quoi ça consiste ?	38
4.2.2.	Enjeux techniques, économiques et juridiques associés à la collecte secondaire des données	38
4.3.	Le traitement : enrichissement et intégration des données	42
4.3.1.	En quoi ça consiste ?	42
4.3.2.	Enjeux techniques, économiques et juridiques associés au traitement des données	43
4.4.	La valorisation : compréhension, diffusion et modèles économiques	45
4.4.1.	En quoi ça consiste ?	45
4.4.2.	Enjeux techniques, économiques et juridiques associés à la valorisation des données	46
5.	Pour une utilisation renforcée des données : une gouvernance et une organisation à mettre en œuvre	48
5.1.	Principes généraux des structures de gestion des données existantes	48
5.1.1.	La Régie de Gestion des Données des Pays de Savoie	48
5.1.2.	Les IDG : structures de gestion des données de l'information géographique	50
5.1.3.	ForCity : une version privée de structure de gestion des données	51
5.1.4.	La régie de donnée selon le programme expérimental Datact	52
5.2.	Missions et points de vigilance formulés à partir de l'étude de l'existant	54
5.2.1.	Missions associées aux structures de gestion des données	54
5.2.2.	Points de vigilance associés aux structures de gestion de données	55
5.3.	Quels modèles économiques pour les structures de gestion des données?	57
5.3.1.	La valeur attribuée à la donnée : un fondement du modèle économique	57
5.3.2.	Analogie avec des modèles économiques existants	58
5.4.	Organisation des collectivités et des acteurs territoriaux pour la gestion des données	62
5.4.1.	Evolution de la place du numérique et de la gestion des données dans l'organisation des collectivités	62
5.4.2.	Les acteurs et rôles dans la gestion des données	63
6.	Pistes de réflexion pour la gestion des données territoriales	66
7.	Annexes	68
7.1.	Tableau synoptique des 11 outils analysés	68
7.2.	Grille d'analyse pour les outils de modélisation	73
7.2.1.	Introduction	73
7.2.2.	Cinq facteurs de succès et recommandations associées pour concevoir un outil de modélisation territoriale	74
	Sigles et acronymes	80

Résumé

La compréhension et le traitement des enjeux environnementaux, sociaux et économiques à l'échelle territoriale nécessitent des approches transversales, et pour cela des outils d'aide à la décision et à l'action publique, en faveur de politiques publiques plus intégrées, notamment pour la planification territoriale et urbaine. En parallèle, on assiste à un élargissement exponentiel des possibilités offertes par le numérique ainsi qu'à de nouvelles opportunités en matière de données mobilisables et de moyens de les traiter et de les valoriser via des outils pour la création de connaissance dans une logique d'aide à la décision.

L'étude menée s'appuie sur l'analyse approfondie des outils d'observation et de modélisation territoriale, ainsi que sur l'analyse du cycle de vie des données et de leur environnement technique, juridique et économique, pour proposer une grille et des facteurs de succès pour la conception d'outils, et des pistes de réflexion pour la gestion et l'exploitation des données territoriales.

Abstract

Understanding and addressing jointly environmental, social and economic issues at the territorial level implies cross-cutting approaches, allowed by tools in support for decision-making, for more horizontal and integrated public policies, especially on urban and territorial planning. In addition, digital tools offer expanding possibilities and new opportunities are emerging within available data and new means for processing and using this data in order to create knowledge in support for decision-making.

The study is based on the analysis of territorial observation and modelling tools, and on the examination of data life cycle and technical, legal and economical frame. The aim of these two major parts is to determine the best framework and key success factors for tools conception, and to explore further interesting ideas to develop territorial data management and exploitation.

Introduction

Contexte de l'étude

La prise en compte des enjeux de développement durable s'est quasi-systématisée depuis une vingtaine d'années dans les collectivités, à travers les documents règlementaires ou volontaires tels que les Agenda 21, les PCET, les documents de planification urbaine. L'approche traditionnelle de ces problématiques, « en silo » (c'est-à-dire selon une segmentation sectorielle et thématique), est aujourd'hui dépassée par des démarches de planification globales. Les secteurs y sont traités de manière de plus en plus transversale et les thématiques se rejoignent.

Pour répondre aux enjeux de développement durable, la nécessité d'une meilleure compréhension des phénomènes en jeu se fait jour. Plus particulièrement, s'impose le besoin de comprendre et d'intégrer les interrelations et phénomènes de rétroaction entre les différents comportements, les impacts énergétiques et environnementaux, et leur environnement (mobilité, comportements de consommation énergétique, émissions de CO₂, évolution des formes urbaines, etc.), pour être en mesure de déployer des actions. Cela implique une meilleure maîtrise de l'ensemble des informations propres à renseigner et nourrir cette compréhension, notamment par la diversification des sources et la mobilisation des nouvelles possibilités offertes par le numérique.

Cette dynamique se reflète au niveau des outils de collecte, traitement et valorisation des données environnementales à l'échelle territoriale. Ces outils répondent à un besoin des acteurs locaux (services techniques des collectivités, opérateurs de services urbains...) dont les décisions, stratégiques et opérationnelles, doivent s'appuyer sur des données, observées ou simulées.

Cette recherche d'une réponse au besoin d'aide à la décision territoriale en matière d'environnement ou d'appui à la recherche a entraîné la création d'une multitude d'outils de modélisation territoriale, portés par une grande variété d'acteurs : bureaux d'études et organismes vendeurs de données ou d'outils de modélisation, infrastructures publiques centralisées, nationales ou régionales, fournisseurs de services, organismes de recherche.

Par ailleurs, ce développement de nombreux outils s'appuie sur des données territoriales de l'environnement toujours plus nombreuses et diversifiées en raison de la multiplication des sources. L'explosion des données produites présente un immense potentiel encore sous-exploité, et nécessite un renforcement des cadres technique, juridique et économique pour la mise en place de systèmes de gestion des données performants.

L'ADEME a développé une expertise en matière d'analyse environnementale et énergétique transversale d'une part et de planification territoriale et urbaine d'autre part, ce qui l'a amenée progressivement à s'intéresser à la problématique de la gestion des données dans une perspective d'aide à la décision des acteurs territoriaux.

Dans ce contexte d'évolution des modes et des moyens de collecter et de traiter les données territoriales, l'ADEME peut tenir plusieurs rôles :

- > Participation à la production d'outils
- > Connaissance, étude des outils existants et grille d'analyse pour la conception d'outils
- > Recommandations à l'attention des collectivités territoriales dans la mobilisation des outils
- > Participation aux réflexions sur la gestion des données territoriales

Objectifs de l'étude

Cette étude avait pour objectif d'aider l'ADEME à déterminer dans quelle mesure et de quelle manière elle peut s'impliquer dans la production d'outils d'aide à la décision en appui à la planification territoriale et urbaine, le soutien aux outils produits par des tiers, l'appui aux collectivités territoriales dans la mobilisation des outils et l'organisation de la gestion des données territoriales, qui permettent, via des indicateurs, d'orienter, de suivre et d'évaluer les politiques publiques.

L'étude s'est appuyée sur un premier travail exploratoire d'identification et d'examen des problématiques liées aux outils de modélisation existants, ce qui a permis d'aborder les besoins des collectivités en termes d'aide à la décision et à l'action. Les problématiques liées aux outils et au traitement des données que ces outils utilisent et produisent ont pu être identifiées et ont ainsi mené, d'une part à l'élaboration d'une grille d'analyse des outils d'aide à la décision à

vocation territoriale à destination des concepteurs, et d'autre part à la formulation de pistes d'approfondissement pour la gestion des données territoriales.

Ce rapport d'étude présente dans un premier temps les éléments de cadrage des outils existants (partie 1) et les résultats de l'étude de cas des 11 outils analysés (partie 2). Après la définition et la description des pratiques en matière de données territoriales (partie 3), sont abordés les enjeux rencontrés tout au long du cycle de vie de la donnée, c'est-à-dire aux étapes de la collecte, du traitement et de la valorisation (partie 4). Sont ensuite étudiées des structures de gestion des données territoriales dans leurs modes d'organisation et leurs modèles économiques (partie 5), avant d'ouvrir des pistes de réflexion plus générales pour la gestion des données (partie 6).

Compte tenu du large champ couvert et des fortes interactions entre les problématiques, la priorité a été donnée à un regard englobant. De ce fait, ce rapport d'étude ne prétend pas apporter des réponses définitives aux enjeux soulevés, mais souhaite au contraire alimenter de façon éclairante les discussions futures.

1. Panorama des outils de modélisation et d'observation territoriale

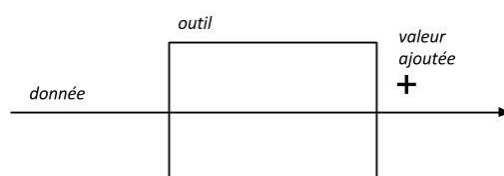
1.1. Un recensement large des outils existants

1.1.1. Qu'est-ce qu'un outil ?

Définition générale : élément d'une activité qui n'est qu'un moyen, un instrument (Larousse)

Dans le cadre de cette étude : les outils étudiés sont des processus calculatoires basés sur des logiciels (existants ou ad hoc) qui permettent d'appliquer un traitement à des données pour répondre à une problématique d'aide à la décision environnementale.

L'information contenue dans la donnée décrit une réalité territoriale, par exemple la population, le revenu, la consommation d'énergie d'un appareil, d'un logement ou d'un secteur tout entier. En passant au travers d'un outil, la donnée acquiert une valeur ajoutée, c'est à dire une utilité supplémentaire du point de vue du bénéficiaire de l'information produite.¹



L'outil comme générateur de valeur ajoutée pour la donnée

1.1.2. Le recensement : plus de 170 outils

Au cours de la phase 1, une recherche extensive d'outils de modélisation territoriale a été menée ; la démarche a croisé les apports de deux sources :

- > La sollicitation des chercheurs et experts mobilisés au sein du Comité Scientifique
- > L'analyse de publications se donnant pour objectif de dresser un état de l'art des outils ou modèles

Au travers de ces deux méthodes, 176 outils ont été identifiés.

1.1.3. Le constat d'une vaste diversité d'outils

Ce recensement exhaustif a mis en évidence qu'il existe une grande variété d'outils, qui peut être comprise comme le résultat :

- > De la variété des thématiques couvertes : énergie, climat, GES, mobilité, biodiversité, ressources, etc. ;
- > De la variété des approches : économiques, techniques ;
- > De la variété des producteurs, en concurrence pour produire des outils répondant aux besoins des collectivités ;
- > De la variété des problématiques : observation territoriale, stratégie de réduction des émissions GES, etc. ;
- > De la variété des besoins : outils de recherche, outils d'aide à la décision.

Cette diversité d'outils a engendré des difficultés connaître précisément le fonctionnement et les caractéristiques de chaque outil et pour catégoriser les outils afin d'en analyser dix représentatifs de cette diversité.

¹ La valeur ajoutée fait de cette donnée une information, c'est-à-dire une donnée liée à un système d'interprétation.

1.1.4. Premiers critères de sélection des outils

A partir de la base de données initiale, l'étude a procédé à une première sélection afin d'éliminer les outils dont la nature ne correspond pas aux besoins de la mission. Pour cela, trois « filtres » successifs ont été appliqués.

- > **Les thématiques retenues** : seuls les outils centrés une ou plusieurs des six thématiques suivantes ont été retenus :
 - Environnementales
 - Energie
 - Précarité énergétique
 - Emissions de GES
 - Pollution de l'air
 - Sectorielles
 - Bâtiments
 - Transports
- > **L'échelle territoriale** : ont été écartés les outils qui se limitent à la représentation d'un seul bâtiment, ainsi que les outils qui ne s'appliquent qu'aux niveaux national ou international.
- > **Les types d'outils retenus** : s'agissant d'un premier filtre, l'étude a écarté les méthodes, lignes directrices, boîtes à outil non formalisées en un « produit informatique ».

L'existence d'informations publiques a constitué un critère supplémentaire.

Après application de ces filtres, il restait 66 outils.

1.2. Organisation et typologie des outils

Deux catégories principales de catégorisation des outils sont distinguées :

- > **Les outils intégrés** : c'est à dire capables de déployer des raisonnements complexes, faisant communiquer des données issues de registres variés pour faire apparaître les influences mutuelles de plusieurs phénomènes ;
- > **Les outils multicritères** : ils comparent les impacts d'un changement de situation sur plusieurs thématiques environnementales de manière simultanée.

Ces deux axes sont au cœur de la mission et constituent par conséquent les repères d'analyse et de classification des outils développés ici. En complément de ces deux axes, d'autres perspectives d'organisation du paysage des outils sont proposées ensuite.

1.2.1. Le caractère intégré des outils

Dans cette étude, l'intégration désigne *l'association et le croisement de données issues de processus de raisonnement distincts*. Le caractère intégré d'un outil sera donc qualifié sur la base de sa *capacité à faire converger ces raisonnements* ; le niveau d'intégration constitue ainsi un indicateur du degré de gestion de la complexité de la réalité qu'un outil propose.

D'après cette définition, on distingue deux grandes familles d'outils :

- > Les outils de **capitalisation**
- > Les outils de **modélisation de processus**.

Les outils de capitalisation répondent principalement à un besoin d'observer, et pour cela centralisent et standardisent les données territoriales, mesurent l'accumulation des faits sur le temps long, et assemblent en général les données par unités spatiales le plus souvent basées sur les découpages administratifs.

Parmi ces outils, on distingue :

- > Les infrastructures de données géographiques qui sont des plateformes d'organisation des données à destination des utilisateurs de systèmes d'information géographiques (SIG) ;
- > Les outils d'observation territoriale développés notamment dans le cadre des observatoires régionaux de l'énergie et du climat, ou dans la veine des bilans carbone territoriaux et des Plans Climat.

Les outils de modélisation de processus répondent principalement à un besoin de comprendre, de décrypter, d'analyser des phénomènes. Ils visent à représenter un processus du monde réel dans son déroulement temporel. Ils s'appuient pour cela sur un modèle, représentation simplifiée (privilégiant certains aspects) d'un système afin de réaliser une série d'opérations de traitement. Les outils de modélisation de processus permettent généralement de représenter les évolutions de l'objet étudié à travers des simulations ou des scénarios.

Le choix a été fait de définir pour l'étude trois degrés pour le caractère intégré des outils étudiés :

- > Les outils de **capitalisation** définis plus haut ont le caractère intégré le plus faible,
- > Viennent ensuite les outils de **modélisation peu complexe**, pour lesquels les calculs se font de façon linéaire et sectorielle,
- > Puis les outils de **modélisation avancée** ont le degré d'intégration le plus élevé avec la prise en compte des rétroactions entre les différents composants.

1.2.2. Le caractère multicritères des outils

Les décisions publiques d'aménagement du territoire et de transition écologique sont susceptibles d'avoir des impacts sur plusieurs aspects environnementaux, mais aussi dans d'autres domaines. Par exemple, la plupart des actions environnementales représentent un coût d'investissement qui est à comparer aux bénéfices attendus de ces actions. De plus, une action positive d'un point de vue environnemental peut avoir des conséquences inattendues et négatives sur d'autres domaines.

Ces réalités poussent la décision publique à se doter d'un regard « à 360° ». Les outils qui répondent à ce besoin présentent un caractère « multicritère ».

Pour mesurer le caractère multicritère des outils, il ne suffit pas de compter le nombre de thématiques abordées dans les résultats produits par l'outil. L'analyse a ainsi porté sur la manière dont les outils examinés venaient en aide à la décision publique, par exemple en proposant :

- > Des « tableaux de bord » de suivi de la décision
- > Des calculs de coût efficacité ou de coût bénéfice par rapport aux euros investis dans la politique publique
- > Une présentation « pédagogique » des résultats dans des interfaces « non-expertes » (pour parler directement au décideur ou aux citoyens)
- > La possibilité d'optimiser la décision publique grâce à la simulation de plusieurs scénarios.

Trois degrés ont été définis dans le cadre de l'étude pour représenter le caractère multicritère des outils :

- > Les outils orientés sur une **unique thématique environnementale** en sortie ont le degré multicritère le plus faible
- > Les outils présentant des résultats relatifs à **plusieurs thématiques environnementales** sont considérés d'un degré multicritère supérieur
- > Enfin, les outils les plus multicritères proposent également une **aide à la décision**.

1.2.3. Perspectives complémentaires

Dans le cadre de la mission, l'analyse est organisée sur les axes-clés que sont le caractère intégré et le caractère multicritère des outils de modélisation territoriale. D'autres registres de classification sont apparus au cours de cette analyse, venant en complément des deux axes clés : ils permettent d'éclairer la très grande diversité des outils existants.

Une approche par **degré de territorialisation** de la donnée impose tout d'abord de distinguer spatialisation et territorialisation :

- > La spatialisation désigne le fait de produire une donnée relative à un espace donné.
- > La territorialisation implique de fournir au travers ou en complément de la donnée des éléments de contexte et de compréhension de la nature des facteurs qui interviennent dans la réalité observée du territoire.

Une approche par la **finalité des outils** repose sur le regroupement des outils en fonction de leur objectif opérationnel annoncé, et permet notamment d'analyser les finalités suivantes :

- > L'observation et les bilans territoriaux, ayant pour objectif de renseigner sur l'évolution des territoires ;
- > L'étude des secteurs d'un point de vue technique, dans le but de mesurer des consommations énergétiques, des niveaux de production et/ou des rejets de CO₂ ;
- > L'analyse des liens entre urbanisme et transport (LUTI – Land use & transport interaction) ;
- > Les analyses de cycle de vie, pour quantifier les impacts environnementaux potentiels d'un produit ou d'un service ;
- > La planification et l'aménagement, afin d'aider la décision publique et plus spécifiquement la maîtrise d'ouvrage et la maîtrise d'œuvre de projets d'aménagement ou d'infrastructures de transports.

Enfin, il est également possible d'associer aux outils observés des attributs en fonction des traits saillants de leur **fonctionnement**. Cette approche est intéressante car elle vise directement le cœur de traitement des outils (modèle), mais elle s'avère complexe à mettre en œuvre dans la mesure où elle requiert d'avoir accès à l'information sur le fonctionnement de l'outil, ce qui est rare lorsque l'on part de sources publiques. Ont pu être distingués les modes de fonctionnement suivants :

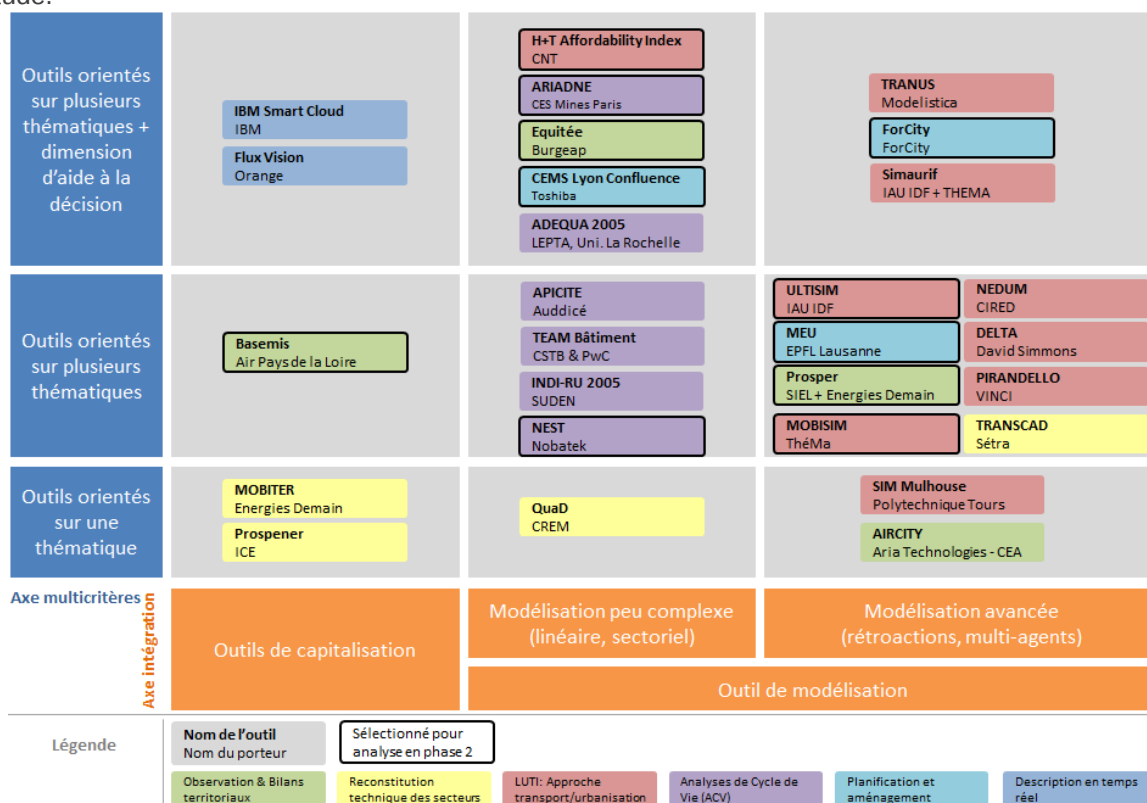
- > **Systèmes multi-agents** : outils basés sur une approche individu-centrée où le comportement global du système est le résultat d'interactions entre plusieurs comportements individuels ;
- > **Approche bottom-up ou top-down** : elle repose sur la manière d'ordonner les étapes d'un raisonnement d'analyse, des éléments individuels vers l'ensemble (approche ascendante ou bottom-up) ou de l'ensemble vers les éléments individuels (approche descendante ou top-down) ;

- > **Backcasting ou forecasting** : les outils peuvent fonctionner selon une approche backcasting, méthode d'exploration du futur dans laquelle les trajectoires des paramètres sont calculées de manière à atteindre un objectif fixé de long-terme. Le forecasting consiste à dériver le résultat final de la projection indépendante de chaque paramètre.
- > **Approche microscopique ou macroscopique** : Les outils de modélisation des transports peuvent être classés en deux types selon l'échelle utilisée dans la description des flux étudiés. L'approche macroscopique permet de caractériser le comportement global du trafic, à une échelle relativement importante, en ne représentant pas les véhicules individuellement mais comme un flux continu ; L'approche microscopique décrit plus finement l'écoulement des véhicules sur un trajet en s'intéressant aux comportements des usagers, aux phénomènes d'insertion et aux changements de voie.
- > **Approche cadastrale ou globale** : les outils peuvent appréhender les bilans et la prospective concernant les émissions de GES d'un territoire selon deux approches : l'approche cadastrale recense les émissions directes générées sur le périmètre administratif du territoire considéré ; l'approche globale prend en compte la totalité des émissions liées aux activités du territoire.
- > **Agrégation ou désagrégation** : L'agrégation désigne la combinaison des lignes d'une base de donnée pour retourner une valeur calculée sur toutes les lignes de la requête (nombre, moyenne, somme, maximum...). A l'inverse, les fonctions de désagrégation partent d'un résultat global (valeur calculée), et permettent de descendre vers des échelles plus fines (possiblement jusqu'aux données) en supposant la relation sous-jacente entre les éléments constitutifs du résultat global et ce résultat.

1.2.4. Vingt-huit outils présélectionnés

176 outils ont été recensés au cours de la première phase ; une présélection d'une soixantaine d'outils a ensuite été réalisée, sur la base de différents critères (accès possible à des informations suffisantes, respect de la thématique de l'étude, respect de l'échelle territoriale, outil en usage ou non...).

Parmi ces outils, un deuxième filtre a été appliqué pour conserver des outils permettant d'étudier le caractère intégré et multicritère. 28 outils sont apparus comme des candidats potentiels pour une analyse approfondie en seconde phase de cette étude.



Cartographie des 28 outils présélectionnés, selon leur caractère intégré et multicritère

2. Etudes de cas approfondies de 11 outils de modélisation et d'observation territoriale

2.1. Sélection d'un échantillon d'outils pour analyse approfondie

De cette liste consolidée d'une vingtaine d'outils, 11 ont été sélectionnés pour une analyse approfondie au cours de la phase 2 de l'étude.

La sélection de ces outils s'est appuyée sur le croisement des critères suivants :

- > Le caractère intégré et/ou multicritères des outils sélectionnés,
- > Une répartition équilibrée de l'échantillon en termes de finalité des outils (cf. 1.2.3), de secteurs observés ou de type de porteur (public, privé, etc.),
- > Le caractère opérationnel, robuste et reconnu des outils : ont été privilégiés les outils pour lesquels des applications concrètes ont eu lieu et où il est possible d'interroger les bénéficiaires et utilisateurs,
- > Modèle technique, juridique ou économique jugé innovant ou intéressant pour la suite de l'étude.

Sur la base de ces critères, les outils suivants ont été retenus :

Management Energétique Urbain (Ecole Polytechnique Fédérale de Lausanne EPFL)

1

L'outil MEU (Management Energétique Urbain) a pour objectif de fournir aux décideurs urbains et aux distributeurs d'énergie un outil d'aide à la décision permettant de planifier et de gérer des systèmes énergétiques en zones urbaines.

L'outil permet plus précisément d'établir un état des lieux énergétique des zones urbaines, de construire et d'évaluer des scénarios énergétiques sur la demande et la production, et enfin de construire une base de données sur les systèmes énergétiques urbains.

L'outil part du bâtiment comme entité de base, en utilisant des données réelles de consommation mesurées, ainsi que des données estimées de consommation.

MOBISIM (Laboratoire ThéMa)

2

MOBISIM est un outil de simulation individu-centré qui permet de simuler des scénarios prospectifs sur les mobilités quotidiennes et résidentielles (individus et ménages). Il s'agit d'un outil dit « LUTI » (Land-use and transport interaction) développé pour fournir une aide à la décision aux collectivités et aux aménageurs sur des questions d'urbanisation et de politiques de transports.

L'outil est conçu pour fonctionner sur un serveur de calcul (il est (actuellement sur le serveur du laboratoire ThéMa). Il est uniquement utilisé par des chercheurs du laboratoire, appliqué à un territoire principalement (Grand Besançon) et n'a pas fait l'objet d'une utilisation opérationnelle pour aider la décision publique.

Les outils des AASQA

3

Le réseau de mesures, dont les résultats sont à destination de l'Etat, des collectivités, industriels, association, de l'UE (rapport annuel), outils historiques des AASQA : le réseau comprend des stations de mesure, des camions-laboratoires, des campagnes de mesures, etc.

Les outils de modélisation, à destination des collectivités et de l'Etat (PPA) : modèles régionaux ou interrégionaux à une maille de 3-4-5 km; modèles urbains sur une agglomération ou une zone industrielle avec une maille de 1m à 10m environ; modèles à l'échelle du quartier ou du bâtiment.

Les outils d'inventaires, alimentant d'autres outils et pour les services de l'Etat et les collectivités (PPA, PCAET, SRCAE, PLUi, ...) : inventaires de quantification de toutes les consommations et de la production d'énergie liées aux activités humaines ainsi que des émissions de GES et de polluants sur les territoires

H+T (Centre for Neighborhood Technology)

L'outil H+T (Housing + Transport) propose une définition de l'accessibilité économique du logement en combinant une estimation du coût du logement et du coût des déplacements.

L'objectif est d'indiquer la meilleure localisation au regard du prix du logement et de la proximité des services, en y intégrant le coût des déplacements induits par cette localisation. Pour chaque zone d'une ville ou d'un territoire, les parts des coûts des déplacements d'une part, du logement d'autre part, et enfin des coûts cumulés logement + déplacements dans le budget moyen des ménages sont comparées à la part moyenne de ces coûts dans le budget des ménages du territoire considéré.

L'outil propose aussi des cartes d'émissions de gaz à effet de serre permettant de comparer les émissions par zone et par ménage.

CEMS Lyon Confluence (Toshiba)

CMS Lyon Confluence, pour Community Management System, est un outil de visualisation en temps quasi-réel de la consommation et des productions d'énergie à l'échelle du quartier. Il constitue un des 4 piliers du projet démonstrateur, qui comporte par ailleurs :

- > La réalisation d'un bâtiment-modèle en efficacité énergétique et intégration d'EnR au sein du quartier de la Confluence à Lyon,
- > Le déploiement d'une flotte de véhicules électriques en auto-partage,
- > L'installation de compteurs (« energy box ») pour assister les habitants dans la maîtrise de leur consommation énergétique.

CMS est un outil de gestion et de contrôle global des données liées à la consommation énergétique de ces 3 éléments listés ci-dessus. Les résultats sont calculés sur la base du signal transmis par les capteurs des bâtiments par pas de 10 min.

Equitée (Burgeap)

EQUITEE est un outil conçu pour accompagner les collectivités, agences d'urbanisme, aménageurs et services de l'Etat dans l'élaboration des stratégies de transition énergétique (PCEAT, planification des réseaux) et la territorialisation dans les documents d'urbanisme (SCOT, PLU, PLH, PDU). L'outil livre un diagnostic des enjeux énergétiques et climatiques de la région à l'échelle du quartier (IRIS), avec un focus sur les vulnérabilités, et des liens entre population, habitat, aménagement, transports. La fonctionnalité prospective de l'outil permet d'intégrer des données spécifiques du territoire et tester des scénarios d'aménagement ou des programmes d'actions sectoriels.

ForCity (ForCity)

ForCity est à la fois une société et le nom donné à la plateforme territoriale de simulation de l'évolution urbaine développée par cette start up. La plateforme entend répondre au besoin d'approche systémique du territoire, en intégrant les composantes de la ville, et surtout leurs interactions mutuelles à différentes échelles de temps et d'espace : la population, sa composition et sa répartition sur le territoire ; les transports, l'énergie, l'immobilier, etc.

L'outil ForCity propose une visualisation 3D des résultats et une utilisation collaborative entre utilisateurs. Il s'agit d'un simulateur spatial sur lequel on peut greffer différents processus linéaires (déplacements) ou réticulaires (réseaux). Les porteurs de l'outil insistent sur le caractère flexible et adaptable de la plateforme pour prendre en charge de nombreux raisonnements de modélisation et de simulation. Ils envisagent de continuer à y intégrer de nouveaux modules et fonctionnalités pour répondre aux besoins des acteurs publics et privés des territoires.

UltiSim (IAU IDF)

ULTISIM est un projet d'outil de simulation de l'interaction occupation du sol – transport (dit « LUTI » : Land Use and Transport Interaction) ambitieux et unique en Europe : complètement intégré qui ne résulte pas du couplage a posteriori d'un logiciel de transport et d'un logiciel d'urbanisation.

Le projet de l'outil a été initié dans le cadre d'une recherche financée par le MEDDTL dans le cadre du PREDIT 4. Il a été abandonné faute de financements pour la poursuite de son développement. Une première phase de conception avait été menée, qui a abouti à un rapport final décrivant l'architecture de l'outil et formulant des recommandations. L'outil ULTISIM aurait pu être utilisé dans le cadre de projets d'infrastructures ou de transport pour simuler sur le long terme les effets de ces projets.

NEST (Nobatek)

9

NEST (Neighborhood Evaluation for Sustainable Territories) est un outil d'évaluation environnementale quantitative (ACV) destiné à aider la prise de décision.

NEST est destiné aux phases de conception des Opérations d'Aménagement. Il n'a donc pas vocation à être utilisé à l'échelle du bâtiment par exemple, mais il permet à l'utilisateur de maîtriser l'impact environnemental du projet par la comparaison de scénarios.

NovaEQUER (CES Mines Paris Tech)

10

NovaEQUER est un outil d'aide à l'éco-conception des quartiers via l'évaluation des impacts environnementaux par l'application de l'analyse de cycle de vie aux quartiers.

L'outil résulte de l'évolution d'un outil d'ACV à l'échelle des bâtiments, EQUER. NovaEQUER fonctionne ainsi au sein d'une chaîne logicielle intégrant ce premier outil qu'il complète sur les aspects spécifiquement liés au quartier.

L'outil permet de comparer différentes variantes de projets de quartier selon différentes étapes du cycle de vie et selon une série de critères sous la forme d'éco-profils.

Il peut appuyer la décision publique à différents stades dans la conception d'un quartier et servir d'outil d'échange entre les parties prenantes du projet.

Prosper (SIEL + Energies Demain)

11

PROSPER est un outil de prospective énergétique territoriale permettant aux communes et intercommunalités de construire des scénarios énergétiques sur leur territoire aux horizons 2020, 2030 et 2050. Sur la base d'un état des lieux initial et multi-sectoriel en 2010 et d'un scénario d'évolution tendanciel de ce bilan tiré du SRCAE, l'utilisateur simule l'impact d'actions à la maille communale ou ECPI : maîtrise de l'énergie et substitution dans les bâtiments et l'appareil industriel, sensibilisation des usagers, mobilité douce, électrique ou commune, développement des EnR. L'impact des actions diffère en fonction des caractéristiques de chaque commune. Les résultats sont comparés aux objectifs fixés dans le SRCAE et évalués en termes économiques (coûts d'investissement, d'exploitation et impact sur la facture du territoire). Ces scénarios d'action sont ensuite consolidés à l'échelle départementale via l'intervention du SIEL, deuxième utilisateur clé de l'outil. Cela permet de visualiser l'impact global des politiques locales mises en œuvre sur le territoire sur les consommations énergétiques et leur impact GES, et de vérifier leur compatibilité avec les capacités du réseau électrique et gazier.

2.2. Méthodologie d'analyse

Les 11 outils sélectionnés ont été analysés de manière approfondie grâce à la réalisation d'entretiens auprès des porteurs et d'utilisateurs de l'outil. La trame suivie est présentée ci-après, elle permet de comprendre les besoins auxquels répond l'outil, ses spécificités techniques, les modalités de développement économique.

Trame d'entretien

Fiche technique de l'outil

- > Support informatique,
- > Périmètre d'application (échelle territoriale, secteurs modélisés, impacts environnementaux traités),
- > Participants à la chaîne de valeur de l'outil (porteurs, concepteurs, développeurs / utilisateurs, bénéficiaires)

Historique et finalité de l'outil

- > Contexte de développement de l'outil : besoin à l'origine, objectif d'appui à la décision publique, réalisation, sollicitation de potentiels utilisateurs, adaptation de l'outil à un territoire
- > Adaptation de l'outil au besoin utilisateur, avantages, limites
- > Exemples d'application
- > Evolutions passées et futures de l'outil

Connaissances et compétences nécessaires pour utiliser l'outil

Collecte des données de l'outil

- > Principales données entrantes et accessibilité
- > Qualité, représentativité et fiabilité des données collectées
- > Structuration du process de récolte des données
- > Apport de données complémentaires par l'utilisateur

Traitement (caractère intégré)

- > Positionnement de l'outil parmi les outils de capitalisation et de modélisation de processus
- > Simulations, comparaison de scénarios

Valorisation (caractère multicritère)

- > Indicateurs clés en sortie de l'outil
- > Validation des résultats en sortie
- > Mise à disposition des résultats pour l'utilisateur
- > Aide à la décision
- > Interopérabilité de l'outil

Modèle économique et pérennisation

- > Temps de travail, budget de développement
- > Pérennisation de l'outil
- > Modèle économique associé à la commercialisation

Enjeux juridiques

2.3. Le tableau synoptique des outils

Un tableau synoptique de synthèse de ces études de cas est présenté en Annexes (*Tableau synoptique des 11 outils*, p.68).

Y sont notamment indiqués les porteurs de l'outil, les utilisateurs cibles, les secteurs modélisés et les impacts estimés en sortie, ainsi que le caractère intégré et multicritère.

La réflexion menée a permis de mettre en évidence les facteurs de succès des outils de modélisation et d'observation territoriale, que ce soit pour le développement d'un outil répondant au mieux aux besoins des utilisateurs cibles pour les porteurs, ou pour le choix d'un outil pour les utilisateurs. Une grille d'analyse pour les outils de modélisation, destinée à accompagner la conception, est présentée en Annexes (*Grille d'analyse des outils de modélisation*, p.73).

Par ailleurs, avec cette analyse approfondie, les différents enjeux liés aux étapes de collecte, de traitement et de valorisation des données territoriales de l'environnement, intervenant au cours de la conception d'un outil, ont été soulignés et font l'objet des parties suivantes.

3. Les données territoriales de l'environnement : définitions, acteurs et législation

Les « données territoriales » ont acquis ces dernières années un statut stratégique dans les projets développés au sein des collectivités ; ce nouveau statut reflète trois dynamiques récentes.

D'une part, les exigences en matière de développement durable à l'échelle locale, notamment la production de documents de planification dédiés à - ou intégrant - l'environnement sont allées croissant depuis le Grenelle de l'environnement. Ces documents de planification se basent généralement sur un diagnostic détaillé des territoires sur lesquels ils portent, depuis les EPCI (Agenda 21, Plan Climat) jusqu'aux régions (SRCAE, SRE²). **Ce diagnostic est lui-même construit à partir de l'interprétation de données territoriales collectées, agrégées et modélisées** couvrant des secteurs (bâtiment, transports...) et thématiques (biodiversité, eau...) variés. Au niveau plus fin des grands ensembles urbains, agglomérations et métropoles, les données ont été mises au service de programmes d'optimisation des espaces et de réflexions prospectives.

D'autre part, un mouvement s'est engagé en Europe au cours des années 2000 (directives PSI et INSPIRE) pour **ouvrir largement au public les données collectées par les Etats et gouvernements locaux**. De ce point de vue, c'est une véritable révolution qui a eu lieu, dépassant largement le seul cadre des politiques de la ville ou de l'environnement ; les évolutions réglementaires actuelles en France confirment ce constat.

Enfin, ces évolutions dans la manière de diffuser et informer l'action publique ont été accompagnées par le **développement exponentiel des technologies numériques**, en entreprise et dans la vie quotidienne des individus. Aujourd'hui, les observateurs de ces évolutions s'accordent à dire que le potentiel généré par l'explosion des données territoriales reste encore largement sous-exploité. Des services nouveaux, basés sur la connectivité des objets et l'analyse en temps réel des grands ensembles de données, sont fortement attendus pour aider la décision publique de demain. Le développement de tels services nécessite des dispositifs performants de collecte et de gestion des données et partant, une évolution radicale des pratiques « historiques », un renouvellement des cadres techniques, juridiques et économiques traditionnels.

² Schéma régional éolien.

3.1. Que sont les données territoriales ?

Une donnée est le résultat brut de la mesure ou de l'observation du monde réel, effectué en référence à une échelle de perception des phénomènes, ou bien une information issue du traitement de données brutes. On appelle les données de la première catégorie "données primaires", issues d'un captage (ou collecte primaire), et celles de la seconde "données secondaires", issues d'une collecte secondaire.

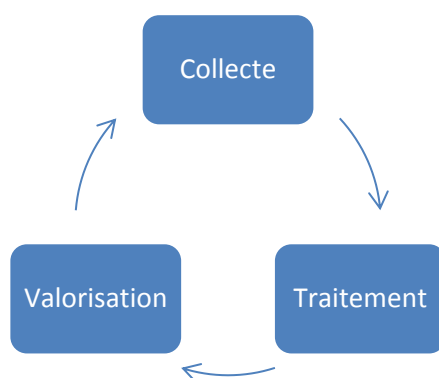
Les données sont le plus souvent de nature quantitative. Ce recours à la formalisation mathématique a l'avantage de clarifier la nature des hypothèses et de rendre possible une révision transparente par des observateurs extérieurs.

L'information contenue dans la donnée est une représentation d'un objet ou d'un phénomène réel ou imaginaire, passé, présent ou futur. Par exemple la population, le revenu, la consommation d'énergie d'un appareil, d'un logement ou d'un secteur tout entier, mais aussi justement une limite virtuelle telle un découpage administratif ou culturel. On parle d'information géographique lorsque l'objet ou le phénomène représenté est localisé dans l'espace à un moment donné et quelles qu'en soient la dimension et l'échelle de représentation. La donnée peut se rapporter à un flux, par exemple la fréquentation quotidienne d'un axe routier, ou un stock, par exemple le nombre de bâtiments chauffés au fioul dans une agglomération.

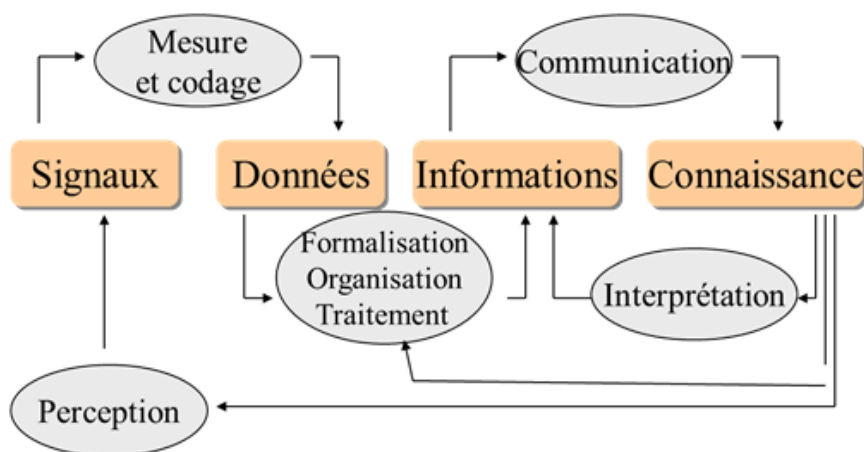
Les données sont traitées par des outils qui intègrent, transforment et valorisent ces données. Ce traitement confère aux données une valeur ajoutée par rapport à un problème donné. La valeur ajoutée à une donnée brute fait de cette donnée une information, c'est-à-dire une donnée liée à un système d'interprétation, qui peut à son tour faire l'objet de nouveaux traitements augmentant cette valeur ajoutée.

On considère ainsi les données territoriales à l'échelle de leur cycle de vie, c'est à dire de leur parcours initié par leur **collecte initiale** (captage), suivie par leur **traitement** (notamment via des outils de modélisation) qui comprend souvent plusieurs étapes et niveaux successifs, puis par leur **valorisation** (notamment sous forme de nouveaux services d'information et d'aide à la décision, proposés aux acteurs publics, privés et aux citoyens). La valorisation des données peut très bien se faire au travers d'un nouveau processus de collecte et de traitement, ce qui montre bien le caractère circulaire de l'usage des données territoriales.

Entre le captage et l'intégration par les outils interviennent des étapes intermédiaires de codage, de formalisation, d'organisation des données (modèles de données, ...). Elles mobilisent, implicitement ou explicitement, des systèmes de représentation propres aux acteurs territoriaux et sont très importantes car elles conditionnent les traitements ultérieurs.



Le schéma ci-dessous distingue la réalité observée, l'information qui en est extraite, la donnée elle-même et la compréhension du phénomène qu'elle favorise.

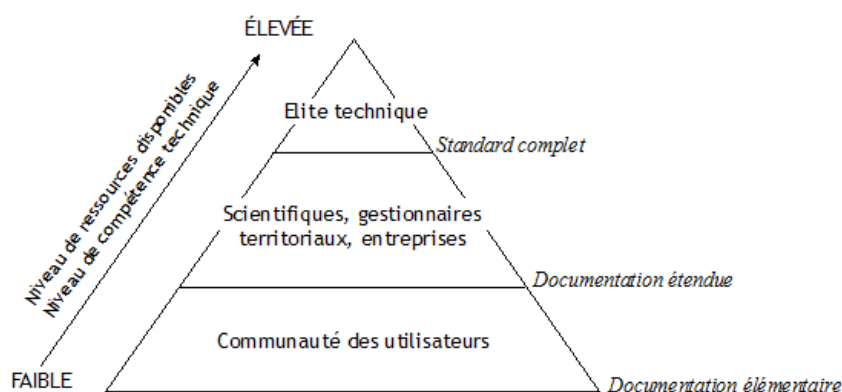


Processus de création de connaissance par une meilleure valorisation de la donnée³

Afin d'illustrer les concepts mobilisés sur ce schéma, on peut donner l'exemple d'un système de collecte de données de circulation:

- > le signal désignerait le passage de véhicules à un point d'une route,
- > la mesure et le captage de ce signal consisteraient en l'action d'un capteur qui relèverait et consignerait ce passage sous une forme codée, en faisant une donnée,
- > l'étape de formalisation-organisation-traitement serait la mise en forme dans une base de données du captage du signal sur une durée donnée, avec éventuellement des précisions attachées aux données renseignées, par exemple le moment (sous forme d'heure et de date) de leur captage, la vitesse des véhicules, etc.,
- > les informations produites pourraient alors être par exemple la fréquence de passage ou bien le flux de véhicules dans un intervalle de temps donné.

En complément des données elles-mêmes, existent également des "métadonnées", c'est-à-dire des données qui décrivent les données considérées. Il s'agit des éléments de description et de documentation des bases de données, qui déterminent en partie la valeur des données diffusées.



Documentation selon les bénéficiaires⁴

³ Source: de Sède-Marceau M-H. « Géographie, territoires et instrumentation : état des lieux, réflexions épistémologiques et perspectives de recherches ». Mémoire présenté en vue de l'obtention de l'Habilitation à Diriger des Recherches, Décembre 2002, laboratoire THEMA, UMR-6049 du CNRS, Université de Franche-Comté.

⁴ Sources: Foresman T.W., Wiggins H.V., Porter D.L. " Metadata Myth : Misunderstanding the Implications of Federal Metadata Standards. " First IEEE Metadata Conference, Silver Spring, Maryland, 1996.

Focus informatique #1 - les différents types et formats de données

On distingue trois niveaux de structuration de données.

Les données structurées

Ce sont des formats de données directement utilisables et intelligibles par des applications informatiques. L'exemple le plus simple est un *dataset* c'est-à-dire une collection de données et qui correspond typiquement à une série de mesures réalisées sur une population d'étude. Les *datasets* sont d'ordinaire exprimés dans des formats très basiques (formats textuels TXT, CSV).

D'autres exemples de données structurées usuelles sont :

- > Les tableaux de données à 2 dimensions (type Excel ou logiciels tableurs équivalents) ;
- > Les logs (flux automatiques de données, par exemple issus de capteurs, de robots ou d'équipements en fonctionnement), souvent générés à un format textuel simple ;
- > Les données cartographiques SIG (exemples de format : ESRI shapefile).

Les données semi-structurées

Ce sont des données avec des structures plus complexes et qui ne peuvent pas être représentées simplement dans un tableau à 2 dimensions.

Par exemple, on peut citer :

- > Des données dites « syntaxiques » qui permettent d'exprimer à la fois un contenu et sa mise en forme (exemple de format usuel : XML). Ce format est très utilisé sur le web et ainsi le format HTML des pages web en est dérivé.
- > Des données « orientées objet » qui permettent par exemple d'exprimer des données en structure arborescente (exemple de format usuel : JSON), très utilisée pour représenter des hiérarchies de données.

Les données non structurées

Ce sont les données les plus difficiles à exploiter informatiquement compte tenu des techniques de stockage et d'analyse actuelles. Leur enregistrement est possible sur un serveur de fichiers par exemple ou sous forme de BLOB (Binary Large Object) à l'intérieur d'une base de données, mais leur intelligibilité, et donc leur réutilisabilité de manière informatisée, est limitée.

Il s'agit typiquement :

- > De documents de travail divers (PDF, Word, Powerpoint, etc.), sur lesquels il est seulement techniquement possible de faire une analyse *full text* pour en extraire des mots-clés, mais pas encore d'obtenir la compréhension sémantique exacte.
- > De documents multimédias : images, vidéos, sons, sur lesquels des algorithmes d'analyse d'images peuvent identifier visages (appareils photos numériques) ou des objets communs (voitures, routes, etc. dans le cas de la Google Car) mais avec une phase d'apprentissage nécessaire (algorithmes de *machine learning*).

Du format des données va découler le mode de collecte et de stockage les plus appropriés et la facilité de leur réutilisation par des tiers.

De par leur représentation en 2 dimensions, les données structurées seront naturellement stockées dans des bases de données « classiques » (dites relationnelles, au standard SQL), alors que des données non structurées seront plus facilement prises en charge par technologies de bases de données plus récentes (dites big data, au standard NoSQL).

Du fait de la grande diversité des types informatiques de données, on perçoit aussi l'importance d'utiliser des données structurées ou semi-structurées à des formats standard (c'est-à-dire normalisés et documentés) et ouverts (au sens de non propriétaires et donc non dépendant d'un éditeur de logiciel), afin d'en faciliter à la fois la collecte et le stockage, mais aussi par la suite leur réutilisation par d'autres acteurs de la chaîne de valeur de la donnée territoriale (notion d'interopérabilité).

Conséquences de l'émergence du Big Data

Contrairement aux bases de données de type SQL qui définissent et sélectionnent la dimension des valeurs cherchées, le Big Data a pour objectif de garder le plus d'informations possibles. Si l'on considère l'exemple d'une image, celle-ci peut être définie par ses couleurs dans une base de type SQL, et ainsi divisés en pixels de certaines couleurs. Dans une base de données de type Big Data, l'image peut être définie par plusieurs variables comme les couleurs, les formes, les émotions ressenties, etc.

Le Big Data est donc constitué de données non structurées et significatives. De ce fait, il permet le traitement de flux différents. Le secteur de la sécurité tire par exemple un avantage certain au traitement de données météo, d'horaires d'ouverture des commerces, de l'usage des transports ou encore de l'historique des incidents, pour l'élaboration automatisée de stratégies de surveillance.

Le Big Data permet un gain d'opportunité en synchronisant mieux l'offre et la demande, par le traitement statistique et probabiliste de très gros sets de données collectées, permettant de mieux prédire les comportements des acteurs et d'optimiser l'offre.

L'émergence du Big Data pose néanmoins la question de la propriété des données produites. Les données appartiennent effectivement à ceux qui les ont créées. Un nouveau service américain, le « blue button » est parti de ce postulat et permet aux citoyens de récupérer, quand ils le souhaitent, l'ensemble des données les concernant. Se pose aussi la question de l'archivage ou au contraire du tri des données à conserver. La consommation d'énergie par les *data centers* (10% de la consommation nationale d'énergie en France) est aussi devenue une préoccupation. Des efforts considérables sont aujourd'hui réalisés à ce sujet, afin de développer des services qui soient le moins énergivore ou carbonés possibles.

3.2. Les données territoriales : une « matière première » mobilisée par de nombreux acteurs

Pour rendre compte du cycle de vie des données territoriales, on fait souvent l'analogie avec l'industrie des matières premières (bois, minerais, etc.). Les données doivent dans un premier temps être extraites de la réalité. Elles sont récoltées et agrégées, puis subissent une série de traitements au cours desquels elles sont raffinées, combinées avec d'autres données pour produire une information plus sophistiquée. Ces produits dérivés de la donnée sont alors valorisés et serviront à leur tour comme intrants dans de nouveaux processus.

3.2.1. Produire et collecter la donnée : une distinction de plus en plus ténue

1. Modes de production et de collecte des données

Il convient de distinguer la production de données, dont l'objectif est véritablement de produire des données non existantes, de la collecte qui a pour but de récolter des données déjà existantes (brutes ou déjà traitées) qui pourront être traitées afin d'être adaptées à la valorisation visée.

Les modes de production et de collecte des données sont variés et différent selon les acteurs ; certains modes se révèlent préférables à d'autres pour des raisons techniques, économiques ou d'accessibilité des données.

Les données peuvent provenir :

- > D'enquêtes statistiques diligentées pour documenter un objectif précis (exemple : recensement de la population)
- > De déclarations à l'administration dans le cadre d'actes administratifs ou juridiques (exemple : permis de construire)
- > De l'exploitation d'un patrimoine ou d'une infrastructure (exemple : réseau de distribution d'énergie, d'éclairage public)
- > De capteurs ponctuels ou diffus (exemple : compteur de consommation, capteur de fréquentation d'un axe routier)
- > Du suivi de comportements individuels (exemple : données de géolocalisation d'utilisateurs, données d'activité sur les réseaux sociaux).

2. Producteurs et collecteurs

Traditionnellement, on distingue les producteurs d'information et les collecteurs de données. Les producteurs sont à l'origine des actions ou faits mesurés par les données. Les collecteurs rassemblent et compilent les actions de plusieurs dizaines, centaines ou milliers de producteurs. Pour le producteur, l'enjeu est de savoir pourquoi et comment la donnée est collectée. Pour le collecteur, c'est l'agrégation et le volume des données qui sont fondamentaux.

Producteurs et collecteurs peuvent appartenir à tout organisme public ou privé :

- > Etat, agences, administrations et collectivités territoriales
- > les organismes institutionnels publics chargés l'observation et de la connaissance des territoires: INSEE, IGN
- > les délégataires de services publics (transports, énergie, eau, déchets) qui produisent des données à partir des infrastructures
- > les entreprises privées qui ont une activité sur le territoire (en particuliers fournisseurs d'accès à internet, opérateurs de téléphonie mobile, gestionnaires de flottes de véhicules partagés, etc.) produisent des données qui viennent directement de l'usage fait du service par les usagers ou des capteurs.

Les citoyens sont un producteur de données à part entière. Leur situation, leurs comportements sont en effet étudiés par tous les modes de productions cités auparavant (déplacements, consommation d'énergie, etc.) et sont décrits par les données produites. Ils peuvent également participer directement à la production de données grâce à leur rôle de "citoyen connecté" (ou « citoyen capteur »), production qui peut être faite de façon passive ou active. Les données produites par les citoyens sont collectées par un acteur différent (l'opérateur téléphonique dans le cas des données issues des smartphones par exemple).

3.2.2. La consommation des données : qui, quelles données, pour quoi faire ?

Au cours de la table ronde “La ville intelligente est-elle durable?” organisée par Le Club Ville de l’IDDRI en avril 2015, les enjeux des données territoriales ont pu être décryptés.

L’amélioration des espaces urbains, dans une perspective de développement durable, constitue une finalité de l’utilisation des données territoriales, au travers des moyens suivants :

- > optimisation des fonctions urbaines,
- > pilotage et gestion de la ville en temps réel,
- > diagnostic et planification de la ville à moyen ou long terme,
- > émergence ou renforcement des processus participatifs.

Dans ce cadre, les consommateurs des données territoriales sont les collectivités gestionnaires en charge de la politique de la ville, les services et agences chargées d’en optimiser le fonctionnement, ou les créateurs de nouveaux services urbains.

3.2.3. Qui sont les transformateurs de données ?

Les transformateurs de données peuvent être les producteurs de données eux-mêmes, qui vont transformer les données collectées avant même de les publier. Il peut également s’agir d’entreprises externes spécialisées dans le traitement des données, de bureaux d’études qui récupèrent des données brutes et les transforment pour leur propre usage, ou encore de citoyens qui produisent et transforment des données (VGI)⁵.

Le caractère malléable des données territoriales, leur capacité à être combinées les unes avec les autres pour en dériver de nouveaux résultats, par capitalisation ou modélisation, signifie que la transformation des données intervient à tous les niveaux d’utilisation.

Les modes de production de données qui visent une finalité précise (enquêtes, campagnes de mesures à l’aide de capteurs) sont généralement accompagnées d’un traitement et d’une transformation de ces données pour les rendre compréhensibles par d’autres acteurs.

Dans le cas des délégataires de service public, les données récoltées par l’usage du service permettent de tirer des tendances et des conclusions quant à cet usage: les producteurs de données sont en mesure de transformer les données brutes obtenues en résultats mis en forme. Cependant, le rôle attendu de cette catégorie d’acteurs n’est pas aujourd’hui très clair quant aux données qu’il convient de publier ou au degré de transformation attendu avant leur publication. L’exemple des nouveaux compteurs intelligents Gazpar mis en place par GrDF illustre bien cette problématique. Dès lors que cette nouvelle technologie sera mise en place sur un nombre significatif de territoires, le groupe deviendra collecteur de données relatives à la consommation de gaz et sera en mesure de les publier de façon brute, ou de les traiter pour y apporter une valeur ajoutée (cartes, graphiques, agrégation, etc.) avant publication.

3.2.4. Quels sont les services liés à la donnée territoriale ?

1. Un nouveau contexte pour l’émergence de nouveaux services

De nouveaux services territoriaux peuvent émerger et les existants se perfectionner avec le traitement des grandes quantités de données disponibles et à l’accès de plus en plus libre et ouvert. La mise à disposition de données travaillées et qualifiées, collectées et traitées en temps réel et parfois avec une dimension prédictive, apporte une valeur ajoutée.

⁵ L’information géographique volontaire (VGI) caractérise les contenus géolocalisés produits par les usagers (Goodchild, 2007). Ce concept a été introduit par le passage de la consultation des contenus de l’information géographique à l’interaction avec ces contenus par les citoyens qui « écrivent la carte ». (Mericskay, 2011).

Des démarches nationales visent à stimuler l'ouverture des données, en particulier publiques (portail data.gouv.fr), et la création de nouveaux services à partir de ces données (création d'une fonction d'administrateur général des données, mise en place des pôles de compétitivité, etc.), dans une stratégie de croissance, d'inclusion et de confiance dans le numérique.

Ainsi, la démarche EcoCité, programme de financement du Ministère du logement, de l'égalité des territoires et de la ruralité, vise à soutenir la croissance et l'attractivité des villes, en les rendant plus respectueuses de l'environnement, moins consommatrices d'énergie ou d'espace périurbain, et en répondant aux attentes de leurs habitants actuels et futurs. Ce programme a pour objectif de stimuler des innovations urbaines cohérentes et au service d'un projet de territoire, dans une approche systémique. Il a 5 dimensions, dont une consiste à favoriser l'intelligence urbaine (par le recours aux outils numériques, la création de nouveaux services innovants, l'expérimentation par les usages). Il est à l'origine de plusieurs expérimentations de nouveaux services urbains et territoriaux basés sur l'utilisation des données qui préfigurent les services du futur de la donnée territoriale qui pourront se développer de manière plus ou moins généralisée sur les territoires sur les thématiques de la mobilité, de la logistique, des énergies, de la communication, des activités, etc.

Dans ce nouveau contexte, l'élément déterminant n'est pas l'innovation technologique mais bien le rôle de la collectivité déjà évoqué et son positionnement en tant qu'orchestrateur plutôt que gouvernail. L'implication des acteurs des territoires est essentielle, en plus de la démarche politique pour construire un projet de territoire. La coopération entre acteurs divers ainsi que les pratiques de mutualisation seront déterminantes dans le développement de nouveaux services, utiles au territoire, à ses habitants et à son développement économique et durable.

2. Différentes catégories de services

Le Club Ville de l'IDDRI a classé les services liés à la donnée en quatre catégories : services d'optimisation des fonctions urbaines traditionnelles, services de pilotage et de gestion de la ville en temps réel, services pour le diagnostic et la planification de la ville, et enfin services visant à faire émerger ou à renforcer les processus participatifs.

La Fabrique de la Cité classe les services d'une manière différente, tout aussi intéressante. Les services de "comptabilité" ont pour objectif de compter ou de recenser une variable urbaine, comme l'illustre un service mis en place à Los Angeles qui comptabilise et renseigne la part d'énergie d'origine renouvelable dans le mix électrique. Les services "smart data" sont des services utilisant les données à des fins d'utilité publique, comme le font des capteurs mesurant la qualité de l'air mis en place sur des lampadaires de Chicago et publiant en temps réel ces informations via une application dédiée au grand public. Les services d'intelligence collective se fondent sur des données produites par les citoyens, comme les exemples d'applications "Citizen connect" ou "Fix my street" qui permettent aux citoyens de signaler des points endommagés dans la ville.

En plus de ces trois types de services, la Fabrique de la Cité a noté des initiatives de partenariats développées autour des données, comme un partenariat entre la Ville de Boston, Uber et Waze qui a permis de mettre en place un système de priorisation des feux tricolores.

En conclusion, les services développés sur un territoire et liés aux données territoriales peuvent être classés en deux grandes catégories :

- > les services d'observation territoriale utilisés pour effectuer un diagnostic du territoire ou en faire la planification à moyen ou long terme,
- > les services de gestion en temps réel de la ville pour son utilisation ou son optimisation.

Des exemples de services produits grâce aux données sont proposés ci-après.

3. Exemples de services produits grâce aux données

Des services recueillant des données

Le pass pour accéder aux services urbains : un recueil de données

Le pass urbain rassemble les données d'usage des services de la ville et de mobilité des citoyens au niveau individuel, ce qui donne accès à divers services existants, notamment publics.

Le pass urbain, mis en place à Lyon Confluence, permet de bénéficier d'une carte unique pour utiliser et pour payer à la fois les transports collectifs lyonnais (tous les TC et les vélos en libre-service), mais également les parkings et les péages autour du Grand Lyon. A terme, ce pass pourra intégrer d'autres services tels que ceux proposés par les bibliothèques municipales, l'accès aux piscines municipales, des services d'auto-partage ou d'autres services à définir.

Gestion énergétique en fonction des besoins : adaptation en fonction des données recueillies⁶

Un système intégré de gestion de l'alimentation énergétique en fonction des besoins et de la présence humaine dans un logement : l'équipement serait intégré au logement et consiste en un système électronique intégré reliant toutes les installations électriques et objets connectés du logement. Ce système est couplé à des détecteurs de présence et à des prises électriques « intelligentes » qui détectent la pertinence ou non d'un branchement électrique.

Une carte recueillant des données personnelles de consommations alimentaires⁷

Un service de carte pour s'alimenter de façon saine et responsable à un prix contrôlé et avec une connaissance des conséquences des choix alimentaires : il s'agirait de créer un réseau de centres de restauration collective qui proposeraient un abonnement mensuel permettant de consulter son impact alimentaire (à partir du traitement des données relatives à ses consommations et aux produits proposés) en lien avec des objectifs personnels renseignés et en termes d'impact environnemental et social des plats proposés et choisis.

Des services issus de l'utilisation des données

Services d'information liés à la mobilité basés sur le traitement de données sur les pratiques

Le secteur de la mobilité et des transports est particulièrement favorable au développement et à l'expérimentation de *services innovants basés sur le traitement des données*, parce qu'il allie enjeux technologiques, économiques, et expérience utilisateurs. Les données en grand nombre et en temps réel issues du big data, si elles peuvent être collectées, permettent en effet de développer une compétence situationnelle, c'est-à-dire de développer des outils qui vont aider la gestion d'une situation de pic ou de crise, voire la gérer automatiquement⁸.

L'exemple de l'application Waze montre le grand potentiel des big data dans le secteur des transports : grâce à la contribution passive (traces des itinéraires empruntés) et active (signalement d'un accident par exemple) des automobilistes, la communauté de conducteurs peut bénéficier d'informations trafic en temps réel et adapter son trajet à ces éléments.

La mise en œuvre d'un parking mutualisé après analyse des données pour évaluer les besoins

Pour assurer le fonctionnement d'un parking mutualisé, le *traitement des données* liées à l'occupation du parking est central. Il permet de limiter l'espace alloué à la voiture en optimisant l'offre de place par rapport au besoin. Cela constitue une condition de réussite du projet.

Un tel parking a pu être mis en place à Lyon Confluence avec une obligation imposée aux promoteurs (un ratio de 0,7 places de parking par logement). De plus, ce ne sont pas les places de parking qui sont vendues, mais le droit d'utiliser un parking via une concession, par exemple sur 20 ans (substitution d'un droit d'usage à une propriété, selon les principes de l'économie de la fonctionnalité). Le parking mutualisé a été rendu acceptable auprès des habitants notamment par l'ajout de services à la mobilité proposés dans le cadre de l'accès au parking. Il s'agit par exemple de voitures et vélos électriques en libre-service, de tarifs préférentiels pour l'accès aux transports collectifs, etc.

La gestion d'un chantier furtif urbain⁹ : l'analyse de données pour optimiser le fonctionnement

A Lyon, dans le quartier de la Confluence, un projet de gestion des chantiers de manière mutualisée unique en Europe s'est mis en place. L'initiative consiste d'abord à *analyser l'ensemble des flux entrants et sortants* du chantier (camions, eau, déchets, etc.) avant de déterminer dans quelle mesure les flux entrants peuvent être mutualisés. La mutualisation est ensuite mise en place via un cadencement et une obligation aux parties prenantes du chantier de collaborer pour

⁶ Ce service n'est pas un exemple réel, il a été imaginé dans le cadre de l'étude prospective sur l'économie de la fonctionnalité conduite en 2008 par des étudiants d'HEC pour la fondation Nicolas Hulot, qui présente des exemples de nouveaux services innovants qui consistent en des fonctions permettant de répondre aux différents besoins humains (habiter, se déplacer, communiquer, se nourrir, etc.). <http://www.inspire-institut.org/etude-prospective-sur-leconomie-de-fonctionnalite-en-france.html>

⁷ Idem.

⁸ Source : Conférence « La ville intelligente est-elle durable ? Décryptage à partir des enjeux de données » du 17 avril par le Club Ville IDDRI

⁹ Selon l'ANR, la « furtivité de chantier », qui s'adresse aux maîtres d'ouvrage et maître d'œuvre, vise à outiller leur capacité à optimiser l'organisation et la planification d'une opération de chantier dans un lieu et un temps donné, en l'intégrant le mieux possible à son environnement urbain.

mutualiser ces flux. La collectivité a joué un rôle de tiers de confiance dans son effort pour mettre en relation des entreprises privées et garantir que les données collectées auprès de ces acteurs (bilan exhaustif de tous les flux entrants) seront partagées avec les autres entreprises uniquement dans l'intérêt du chantier qui n'y auront pas accès pour d'autres usages.

La gestion d'un parc de résidences secondaires mutualisées¹⁰

Un service de gestion des résidences secondaires de manière mutualisée les rendant énergétiquement performantes : il s'agirait pour un particulier de confier sa résidence secondaire à une société de gestion locative en échange de l'accès à un parc de nombreuses autres résidences de même standing et de bénéficier de services supplémentaires. Les gains environnementaux consistent à limiter les consommations d'énergie inutiles générées par les résidences secondaires inoccupées et à l'amélioration des performances énergétiques des résidences sous la gestion de la société tiers.

Vers de nouveaux services basés sur les usages avec l'essor de l'économie de la fonctionnalité

Dans le domaine de la ville comme dans beaucoup d'autres, on est entré dans un fonctionnement de plus en plus basé sur une économie de la fonctionnalité où l'usage des biens prend le dessus sur leur propriété. Le développement de services basés sur les données territoriales de l'environnement gagne à s'inscrire dans cette évolution pour deux raisons :

D'une part, l'économie de la fonctionnalité constitue un levier fort pour le développement durable, dans la mesure où l'intensité de l'usage d'un bien est à la fois un facteur de profit et un facteur de réduction des consommations et des ressources en matières premières et énergie ainsi que des impacts environnementaux associés.

D'autre part, l'utilisation de grandes quantités de données spécifiques à un territoire constitue un fort potentiel pour le développement de services basés sur l'usage de biens grâce à certaines spécificités des traitements permis par ces types de sets de données : localisation des individus et des services en temps réel (pour des services d'auto-partage par exemple), développement du *peer-to-peer*, prise en compte du renseignement de données personnelles de tous types et de manière consentie, etc.

4. Les nouveaux objets connectés : nouveaux instruments de production de données

La révolution numérique s'accompagne de l'émergence des objets connectés, terme qui désigne des objets munis de capteurs variés et reliés à internet, ce qui permet un partage en temps réel des données générées.

L'usage de ces objets ou capteurs connectés est aujourd'hui très vaste :

- > Des capteurs reliés aux compteurs électriques ou de gaz permettent également de mesurer la consommation des foyers.
- > Dans le secteur des transports, des projets visant une meilleure gestion du stationnement voient le jour avec l'installation de capteurs détectant le caractère libre ou non d'une place de parking. Des services utilisent ces données en temps réel pour orienter le conducteur vers la place libre la plus proche.
- > Le secteur agricole, par exemple au Qatar qui vise l'indépendance alimentaire d'ici 2025. Des capteurs infrarouges sont largement utilisés pour mesurer le taux d'humidité des plantes et détecter plusieurs heures avant que l'œil humain ne puisse le faire le moment de les arroser.

Le principal frein à l'utilisation de ces objets connectés réside aujourd'hui dans le manque de confiance qu'y accordent les individus, qui y voient un risque de publication ou d'utilisation de leurs données personnelles sans leur avis. En plus des protections juridiques à mettre en place pour les données, la communication sera nécessaire à l'établissement d'un nouveau climat de confiance.

¹⁰ Service imaginé dans le cadre de l'étude prospective sur l'économie de la fonctionnalité conduite en 2008 par des étudiants d'HEC pour la fondation Nicolas Hulot.

3.3. Le cadre juridique de la collecte et de l'utilisation de la donnée territoriale

La collecte, le traitement et la valorisation des données territoriales posent des problèmes juridiques et éthiques. A ce titre, les données territoriales liées aux enjeux environnementaux ne font pas exception. Les problèmes posés appartiennent principalement à deux catégories: les données qui touchent aux individus et celles qui révèlent des informations stratégiques sur les entreprises.

3.3.1. Un premier cadre : la loi sur le secret statistique

Historiquement, ce sont les enjeux liés aux informations stratégiques sur les entreprises qui ont été réglementés en premier. La loi sur le secret statistique de 1951 définit des limites non pas dans la collecte mais dans la diffusion des données. Ainsi, les agrégats collectés sur un territoire ne peuvent pas être divulgués s'ils concernent moins de 3 entreprises sur le territoire. L'exemple le plus courant concerne la publication des données de consommation d'énergie ou d'émissions de polluants dans l'industrie, où la concentration des activités au sein de quelques entreprises est très marquée.

Le secret statistique peut, en théorie, être contourné si l'utilisateur des données respecte lui-même la règle de non-divulgaration. La loi prévoit alors que les données doivent être transmises dans des conditions très sécurisées. Par exemple, les supports contenant les données ne peuvent pas être connectés à un réseau, et doivent faire l'objet d'une protection physique particulière (coffre-fort).

Aujourd'hui, **le secret statistique reste un élément essentiel de la protection des données stratégiques** sur l'activité des entreprises, en particulier dans le domaine industriel.

Le plus souvent, les données territoriales sur l'environnement ne sont pas couvertes par le secret statistique, ni directement relatives à des individus. Il arrive cependant que les porteurs d'outils et les producteurs de données doivent composer avec le droit en place. Par exemple, le besoin de données à échelle très fine pour les outils d'inventaire des consommations ou des émissions se heurte au secret statistique: les données doivent être estimées (avec une marge d'erreur inévitable) plutôt que mesurées ou enquêtées. Dans le domaine de l'instrumentation des consommations d'énergie des bâtiments, ou de reconstitution des déplacements individuels des ménages, le droit en matière de données personnelles peut intervenir.

Au-delà des conditions générales déterminées par la réglementation, l'accès, l'usage et la circulation des données sont conditionnés par des contrats bilatéraux de droit privé. Les contrats d'études ou les partenariats de soutien aux projets de collecte de données précisent généralement le statut des résultats, leur propriété, les conditions dans lesquelles les acteurs peuvent les utiliser ou les faire circuler à des tiers. Dans le domaine de l'énergie, les données de consommation énergétique des clients des réseaux de distribution sont souvent sollicitées par les collectivités qui souhaitent engager une démarche de planification territoriale. Des conventions spécifiques régissent la mise à disposition et l'utilisation de ces données, généralement transmises à des niveaux d'agrégation qui échappent au périmètre des données personnelles.

3.3.2. Les perspectives d'évolution de la juridiction française

Dans les années 1970, la généralisation des fichiers statistiques, des enquêtes d'opinion, des listings en entreprise et les débuts de l'informatique ont encouragé la législation sur les données relatives aux individus, appelées données personnelles. La loi "informatique et libertés" de 1974 s'inscrit avant tout dans la perspective de prévenir les dérives du fichage des citoyens par les pouvoirs publics ou les intérêts privés. La loi définit les données personnelles et isole une catégorie particulière de données sensibles, telles que celles se rapportant à la religion, les préférences politiques ou l'orientation sexuelle des individus. Sous le régime de la loi, la collecte des données sensibles est interdite (sauf autorisation particulière à des fins très encadrées, de recherche principalement). Contrairement à la loi de 1951, c'est bien la collecte et non plus seulement la divulgation des données qui est prohibée. Pour les données personnelles non-sensibles, le fichier doit être déclaré à la CNIL de manière à garder un suivi de l'existence des fichiers.

Le contexte français récent est marqué par la volonté de continuer à se doter d'outils juridiques pour la gestion des données. La première moitié de l'année 2015 a été particulièrement riche sur le sujet. En parallèle de la remise des rapports Jutand et Lepage (voir encadré), la loi n° 2015-990 pour la croissance et l'activité (dite Loi Macron) a intégré plusieurs mesures portant sur l'Open Data. Adoptées sous la forme d'amendements, et en dépit de l'avis défavorable du gouvernement qui souhaitait les réserver au projet de loi sur le numérique, ces mesures portent sur l'ouverture des données de transport et du registre du commerce.

- > La première impose la mise en ligne des « principales données des services réguliers de transport public de personnes », et ce « sous un format ouvert et librement réutilisable ». Cette mesure oblige les entreprises concernées à ouvrir leurs données relatives à leurs horaires théoriques, à leurs arrêts ou bien encore aux conditions d'accessibilité pour les personnes handicapées, et vise notamment à faire émerger des plateformes de comparateur de coûts des trajets en fonctions des différents modes de transport.
- > La seconde mesure prévoit de rendre gratuites et réutilisables par tous les données publiques contenues dans le registre national du commerce et des sociétés. Ces données seront mises en ligne par l'Institut national de la propriété intellectuelle (INPI).

Le rapport Jutand

Remis en mars 2015 au Secrétaire d'Etat aux transports, ce rapport préconise l'ouverture des données de transport, en insistant sur les retombées bénéfiques à attendre pour l'information aux usagers, la promotion des modes de transport alternatifs, ainsi que la transparence de l'action publique. Il préconise que cette ouverture se fasse de manière souple, en classifiant les données. Le rapport retient ainsi trois modes d'ouverture des données : *données ouvertes libres de réutilisation* déterminées par décret ; *données ouvertes avec condition de réutilisation* incluses dans une liste gérée par l'État (la restriction étant par exemple liée à des cas de gestion de crise) ; *données ouvertes à des fins d'études et de recherche* qui relèvent de la décision de l'autorité organisatrice (la restriction s'explique par le risque d'atteinte à la vie privée, au secret commercial...).

Afin d'élargir le périmètre des informations réutilisables et pour combler certains manques de la loi CADA¹¹, le rapport se prononce en la faveur de la création d'un statut « de « données d'intérêt général » pour faciliter leur mise à disposition.

Le rapport Lepage

Ce rapport sur « L'Economie du Nouveau Monde » remis en juin 2015 à la ministre de l'Environnement, et dirigé par Corinne Lepage, s'inspire des principes de la 3^{ème} révolution industrielle et propose plusieurs mesures structurantes afin de favoriser l'initiative économique. L'ouverture des données et la valorisation du Big Data font notamment partie des solutions proposées pour soutenir les PME innovantes, en particulier dans les domaines de la mobilité durable, de la rénovation énergétique ou encore de l'agriculture.

Une loi sur le numérique est attendue depuis plusieurs années pour venir renouveler les principes énoncés par la jurisprudence communautaire et par la loi de confiance dans l'économie numérique (LCEN) de 2004. Le projet de « loi pour une République numérique » en cours d'examen par l'Assemblée Nationale en janvier 2016¹² est axé autour de trois volets centraux, que sont la circulation des données et du savoir, la protection des données personnelles et l'accès de tous au numérique.

Concernant la circulation des données, le projet de loi prévoit d'ouvrir gratuitement l'accès aux données publiques communicables au sens de la loi CADA, et de mettre en place un service public de la donnée afin de faciliter leur exploitation. Il prévoit également la mise en place de la notion de « données d'intérêt général », couvrant les données produites par les délégataires de service public ou les organisations recevant plus d'un million d'euros de subvention

¹¹ Loi n° 78-753 du 17 juillet 1978, posant le principe de la liberté d'accès aux documents administratifs

¹² Une première version du projet de « Loi pour une République numérique » a été présentée au mois de septembre 2015 par la secrétaire d'Etat au numérique pour l'ouvrir à une consultation publique, puis remaniée. Le projet de loi (version 2) a été adopté par le conseil des Ministres le 9 décembre, à l'issue de son examen par le Conseil d'Etat. Il a été transmis à l'Assemblée nationale, qui commencera son examen en commission en décembre puis en séance publique en janvier 2016.

ainsi que certaines données privées produites dans les secteurs de l'énergie ou de l'environnement, dans le but de permettre leur réutilisation par tous.

Le projet de loi affirme le principe de portabilité des données personnelles, qui donne le droit aux internautes de demander une copie des données collectées à leur sujet lors de leur inscription sur des sites ou bien lors de leurs transactions en ligne. Les notions de droit à la mort numérique et de droit à l'oubli pour les mineurs font également partie du projet de loi. La mise en place de mesures en faveur de la protection des données personnelles a créé un climat de confiance de la part des internautes dans leur navigation, ce qui favorise par effet de ricochet la création de données.

Concernant l'accès de tous au numérique, le texte prévoit le maintien de la connexion internet, indispensable pour communiquer, travailler, s'informer, chercher un emploi ou établir des actes administratifs, en cas de défaut de paiement. Cela induit donc la question du droit à une connexion minimale, comme c'est le cas pour l'électricité par exemple.

3.3.3. La législation européenne

La directive européenne 95/46/CE inscrit en droit européen les dispositions concernant la collecte et le traitement des données personnelles. La directive prolonge la loi de 1974 en énonçant notamment les principes de consentement et le droit d'accès (modification, rectification ou suppression) des individus au sujet desquels sont collectées les données. De même, la divulgation des données à des tiers est strictement encadrée et soumise à un consentement spécifique.

La directive PSI 2003/98/CE fixe un ensemble minimal de règles et de moyens concernant la réutilisation des informations du secteur public. Elle vise à mettre en place un des premiers fondements de la réutilisation par les opérateurs économiques des informations publiques, qualifiées de « ressources essentielles ». Les dispositions de cette directive ont été transposées en droit français à travers une ordonnance et un décret venant modifier la loi CADA du 17 juillet 1978. Cette directive est cependant peu normative du point de vue des Etats, auxquels elle n'impose pas d'autoriser la réutilisation de leurs informations publiques. Devant le constat qu'elle n'a pas permis la création d'un véritable droit de la réutilisation opposable aux Etats, elle a récemment été révisée par la directive 2013/37/CE, qui impose une obligation générale pour les Etats membres d'autoriser la réutilisation, à des fins commerciales ou non, de leurs informations publiques librement accessibles, à moins qu'elles ne soient protégées par un droit de propriété intellectuelle au bénéfice d'un tiers. Cette directive a été transposée en droit français par l'ordonnance n°2148 du 16 juillet 2014.

La directive INSPIRE 2007/2/CE du 14 mars 2007 établit l'infrastructure d'information géographique dans l'UE en vue de la protection de l'environnement. Elle s'adresse aux autorités publiques diverses « fournissant des services publics en rapport avec l'environnement » et vise à améliorer l'information et les services aux usagers, ainsi qu'à favoriser la croissance et l'innovation. La directive impose que ces données soient partagées et accessibles sur Internet ; et que les autorités publiques les partagent entre elles. Parmi ces données géographiques, la directive établit une distinction entre les référentiels géographiques, les objets géographiques et les données. Pour faciliter le respect de ces obligations par les autorités publiques concernées, la directive leur impose de recenser les données concernées par le texte, d'établir des métadonnées exploitables à partir de cette collecte, de mettre ces métadonnées sous forme utilisable et de les publier sur Internet.

Big data : les enjeux sociétaux et économiques du changement d'échelle dans le volume des données rendent nécessaire le cadre juridique

Le phénomène du Big Data (traduit parfois en français par « mégadonnées ») désigne la croissance exponentielle du volume des données disponibles sous forme numérique aussi bien dans les entreprises que sur l'internet¹³. Cette masse de donnée est gérée par des technologies et algorithmes de tri en temps réel et des outils d'analyse permettant notamment de faire émerger des comportements, des profils de consommateurs, etc. Cette évolution technologique accompagne et alimente une profonde mutation culturelle, centrée sur le partage d'informations et l'explosion de nouveaux usages de celle-ci. De par les informations qu'il recèle et révèle, le Big Data possède un potentiel économique et sociétal jamais égalé.

Aujourd'hui la maîtrise des masses de données produites est détenue par un nombre restreint d'acteurs essentiellement privés (Google, Amazon). Quelques acteurs publics et quelques universités possédant les puissances de calculateurs requises intègrent ce cercle très fermé (le CEA par exemple). Ce paysage des acteurs du big data pose la question de l'accès à l'information et de la concentration des pouvoirs que cette configuration implique. Les liens entre intérêts publics et intérêts privés de l'exploitation du Big Data, notamment dans les domaines de la santé et de l'éducation, la gestion démocratique de ces masses de données, sont autant de questions posées aujourd'hui aux décideurs comme aux citoyens.

D'un point de vue économique, le Big Data a initié une « démarche visant à faire des données un mode de décision, un actif stratégique et une façon de créer de la valeur C'est un véritable renversement de paradigme d'organisation, [...] dans lequel l'entreprise est guidée par les données¹⁴ ». Au-delà des informations commerciales classiques permettant l'optimisation des activités des entreprises, la multiplication des objets connectés ouvre la voie à des applications et des usages innovants dans le domaine des services et du géomarketing. Dans ce contexte, la maîtrise de l'information devient un élément clé de la compétitivité des entreprises (et pourquoi pas des Etats).

Les intérêts qu'il suscite imposent donc la mise en œuvre d'un cadre juridique qui doit porter à la fois sur la propriété des données et sur les garanties autour des usages qui en sont faits. Il convient alors d'identifier ce qui relève de la propriété intellectuelle et de la protection de la vie privée pour les données privées. La question se pose également pour les données publiques qui relèvent de l'open data et dont l'usage est régi par la licence ouverte¹⁵.

Enfin il convient de protéger le patrimoine informationnel et les savoirs faire qui s'y rattachent. Dans ce cas, et selon les experts, aucun dispositif de protection n'existant à ce jour, cette protection repose sur les règles de protection des logiciels, des bases de données, etc.

¹³ 12èmes journées de la propriété intellectuelle, nov. 2014

¹⁴ Source : <http://www.data-business.fr/>

¹⁵ Voir http://wiki.data.gouv.fr/wiki/Licence_Ouverte/_/Open_Licence

4. Le cycle de vie des données territoriales

Cette partie présente les données territoriales par une analyse en cycle de vie : collecte, traitement, valorisation. Elle présente notamment, pour chaque étape du cycle de vie, les enjeux techniques, économiques et juridiques qui y sont spécifiques.

4.1. La collecte primaire ou captage, obtention initiale des données

4.1.1. En quoi ça consiste ?

La collecte primaire (captage) concerne l'obtention initiale de la donnée. La donnée peut être obtenue en captant directement certains aspects de la réalité que l'on veut décrire (par exemple, une campagne de mesure à l'aide de wattmètres pour mesurer la consommation d'énergie d'habitants d'un échantillon de logements) ou bien en interrogeant en premier les objets et phénomènes de la réalité (par exemple, une enquête téléphonique sur les pratiques de déplacement urbains).

4.1.2. Quelles sont les méthodes de captage ?

Les enquêtes (enquêtes ménages déplacements, enquête nationale transports et déplacements, enquête INSEE) peuvent être réalisées de façon régulière ou ponctuelle. Généralement, cela prend la forme d'une liste de questions auxquelles répond un échantillon de la population étudiée. Les enjeux sont donc d'avoir une bonne représentativité de l'échantillon considéré par rapport à la population étudiée, ainsi que d'éviter et de savoir reconnaître les biais de réponses: en répondant à une enquête sur la santé ou l'environnement, les personnes vont gonfler leurs parts de modes actifs.

Les instruments de mesure sont très variés. Il peut s'agir de caméras, compteurs (électricité, eau, gaz), de capteurs de présence, de capteurs magnétiques de comptage autoroutier, des péages, etc. Ils sont destinés à mesurer la réalité et l'enjeu principal est donc d'avoir des capteurs performants techniquement, ce qui se teste ex ante. De la même manière que pour les enquêtes, les instruments de mesure représentent un échantillon qui sera ensuite traité pour obtenir une vision globale. L'exemple de la caractérisation de la qualité de l'air illustre bien ce point : la mesure des concentrations de polluants en un nombre fini de points par des capteurs permet une estimation de la concentration en tous points de la zone considérée.

Stratégies et intentions des acteurs territoriaux dans la collecte de données

Il est important de faire la nuance entre **une donnée d'origine privée** - qu'il s'agisse d'une donnée produite par une entreprise, une association, un individu - et une **donnée qualifiée de publique**. Une donnée publique est une donnée collectée ou produite, dans le cadre de sa mission, par un service public, sur fonds publics¹⁶). Dans les faits, les données publiques sont généralement le fruit de l'action d'un certain nombre d'institutions reconnues, le terme d'institution désignant un organisme visant à maintenir des normes ou pratiques sociales ayant valeur officielle. Cette première distinction est importante puisque, selon les législations, elle va influencer la notion de propriété intellectuelle et les cadres de diffusion.

Cependant au-delà de ces implications juridiques, il convient d'introduire la notion d'**intentionnalité de la collecte**. Dans le cas des données publiques, la collecte est essentiellement conditionnée par les orientations et politiques publiques mises en œuvre. Les référentiels (type données IGN, INSEE,...) sont par exemple issus de véritables choix d'observation et véhiculent des représentations de composantes territoriales vues par les institutions pour répondre à leurs besoins. Pour ce qui concerne les données privées acquises par des entreprises, des associations voire des individus, cette intentionnalité sera liée aux objectifs et aux stratégies propres à chacun de ces producteurs de

¹⁶ Circulaire du 14 février 1994 relative à la diffusion des données publiques, J.O., 19 février 1994.

données.

Il en va tout autrement des **données issues d'approches collaboratives** dont le caractère intentionnel, même s'il est présent, ne peut être rapporté aux objectifs et stratégies d'un seul acteur. Le projet OpenStreetMap, dont l'objectif est de produire une carte collaborative du monde, s'inscrit dans ce cadre et met à disposition de tous des données brutes sous licence libre.

Le développement et la montée en puissance de la téléphonie mobile conduit aujourd'hui à l'explosion de données que l'on pourrait qualifier de non intentionnelles, générées par le captage des signaux émis. Dans ce type de configuration, le producteur de la donnée (l'utilisateur du téléphone portable) est à dissocier du collecteur (l'opérateur de téléphonie mobile), ce dernier pouvant en revanche, en fonction de ses objectifs et stratégies, réintroduire un caractère intentionnel dans les modes de structurations et de diffusion mis en œuvre.

La collecte des données est en lien avec la production de données (types de données nécessaires, modes de collecte) **et avec leur utilisation** (construction d'un indicateur composite avec des données particulières). Disposer des données est nécessaire pour produire des informations qui permettent le pilotage des systèmes.

► Pour aller plus loin : <http://openstreetmap.fr/> et <http://lesrencontres.decryptageo.fr/grand-temoin-valerie-peugeot/> (conférence de Valérie Peugeot aux journées Décryptageo 2015 : réflexions autour de la notion de « communs » appliquée à la connaissance).

Focus informatique #2 - les solutions techniques pour collecter et stocker des données

Préambule

Le stockage en propre des données, territoriales ou autres, n'est pas obligatoire. Dans un contexte où il existe déjà de nombreux catalogues de données publiques dites plateformes *open data*, il n'est pas nécessaire pour un territoire de dupliquer une donnée existante et librement accessible, mais simplement de proposer les moyens d'y accéder au moyen d'un moteur de recherche et d'un outil de requête adapté. Un territoire doit donc concevoir son catalogue de données territoriales en tenant compte d'un écosystème existant, de nombreuses sources de données pertinentes, afin de compléter l'existant.

► Pour aller plus loin : <https://www.data.gouv.fr>, <https://data.sncf.com>

D'un point de vue technique, il existe 3 grandes familles de systèmes de bases de données, et chaque technologie sera plus ou moins adaptée en fonction du type de donnée à gérer.

Les bases de données relationnelles

C'est le système de bases de données le plus répandu et le plus utilisé. Il est particulièrement adapté pour stocker des données structurées dans des tableaux à 2 dimensions, avec en colonnes les attributs caractérisant la donnée et en lignes les enregistrements de cette donnée. Pour interagir avec ce type de bases de données on utilise le langage standard SQL (Structured Query Language).

Des solutions de bases de données relationnelles sont proposées aussi bien par des éditeurs de logiciel établis et bien connus (Microsoft SQL Server, Oracle Database) que par la communauté open source.

Les bases de données multidimensionnelles

Ce système de bases de données multidimensionnelles (dites OLAP) est adapté pour stocker des données caractérisées par plusieurs axes ou dimensions. Prenons un exemple concret : sur un territoire donné, la consommation d'énergie du secteur résidentiel peut ainsi être décomposée selon les dimensions suivantes : une composante géographique (code postal, département, région), le type de logement (maison, appartement), le type d'usage

(chauffage, ECS, cuisson, climatisation, etc.), le type d'énergie consommée (électricité, gaz, bois, etc.)
Plusieurs éditeurs sont spécialisés sur ce type de bases de données : IBM Cognos, SAP NetWeaver, Microsoft Analysis Service, Jedox OLAP, Pentaho Mondrian.

Les bases de données orientées big data

Ces bases de données sont optimisées pour stocker et gérer une quantité de données beaucoup plus grande (plusieurs millions d'enregistrements). Il s'agit d'une technologie récente et pour laquelle aucun standard unique ne s'est imposé. Il existe donc plusieurs sous-familles de bases big data, chacune plus ou moins adaptée selon les besoins en termes de traitement de données et de performance : il y aura des bases orientées « documents » particulièrement adaptées pour un moteur de recherche qui cherche à indexer le contenu de millions de pages web, ou encore des bases orientées « graphes » pour permettre l'analyse de données issues des réseaux sociaux par exemple. Actuellement ces différents types de bases de données big data sont désignés sous le terme générique de « NoSQL » (Not Only SQL) pour indiquer leur complémentarité avec les bases de données SQL traditionnelles. Citons plusieurs solutions performantes qui émergent : MongoDB, Cassandra, CouchDB.

» Pour aller plus loin : solutions open source MySQL (<https://www.mysql.fr>) et PostgreSQL (<http://www.postgresql.org>) qui sont les plus utilisés sur le web actuellement

Enfin, pour aller plus loin que la base de données, il est nécessaire d'installer un programme qui va permettre d'informatiser la gestion des données, que ce soit la collecte, la publication en ligne ou la gestion du *workflow* de mise à jour de ces données avec des systèmes d'horodatage, d'archivage et d'envoi d'alerte automatique aux utilisateurs concernés. On peut désigner ce type de logiciel par le terme de plateforme web de partage des données (ou portail web, catalogue web, etc.).

Il existe actuellement 2 approches possibles pour mettre en œuvre une telle infrastructure, selon que le territoire souhaite avoir la complète maîtrise de sa solution technique ou bien au contraire la « sous-traiter » auprès d'un prestataire technique spécialisé :

- > L'utilisation d'un logiciel open source de type *data portal platform* et la gestion de sa propre plateforme. Le logiciel open source le plus complet et abouti actuellement est CKAN très utilisé par des sites open data à l'échelle nationale (France, Etats-Unis, Grande-Bretagne) ou à une échelle plus locale (Berlin, Amsterdam)
- > Le recours à une plateforme d'hébergement propriétaire et gérée par une entreprise privée. En France plusieurs éditeurs existent (OpenDataSoft, Data Publica). De manière générale, la rémunération de ces éditeurs se fait en fonction de la quantité de données hébergées ou au nombre de requêtes faites à la base de données. Par exemple OpenDataSoft indique un tarif de 700€/mois pour 10 millions d'enregistrements gérés dans leur base de données ou 100.000 requêtes faites sur la base par jour (et 200€/mois pour 100.000 enregistrements ou 20.000 requêtes par jour).

Dans une perspective d'interconnexion avec des plateformes open data existantes et par souci d'évolutivité, il sera plus intéressant de privilégier une solution basée sur un *data portal platform* en open source comme CKAN.

» Pour aller plus loin : <http://ckan.org>, <https://www.opendatasoft.com>, <http://www.data-publica.com>, <http://www.data.gov>, <http://data.gov.uk>, <http://daten.berlin.de>, <http://www.amsterdamopendata.nl>

4.1.3. Enjeux techniques, économiques et juridiques associés à la collecte primaire des données

Les enjeux présentés ci-dessous, rencontrés au moment de la collecte primaire des données, peuvent être techniques, économiques et/ou juridiques.

Des pratiques de collecte de plus en plus individualisées

TECH

ECON

JURD

Cet enjeu est très récent et est particulièrement lié à la montée en puissance des objets connectés.

Le smartphone en fait partie et est aujourd'hui utilisé par une part de la population en constante augmentation (27 millions de personnes équipées en 2014 selon la MMA¹⁷). C'est un véritable outil connecté qui fonctionne à l'aide de nombreux capteurs (GPS, accéléromètre, caméra, etc.) et est régulièrement connecté à internet ce qui permet à l'opérateur de collecter les données issues des mesures effectuées par les capteurs.

L'usage d'internet sur un ordinateur permet également la collecte passive de données qui peuvent être à usage commercial (centres d'intérêt, nombre de clics sur une page web).

Plus récemment, les montres connectées (et plus largement les objets connectés du domaine de la santé), permettent également de récolter des données relative à la santé ou à l'activité physique des individus.

Ces données peuvent être récoltées de manière plus ou moins passive (géolocalisation) et sont alors la propriété de l'opérateur de téléphonie mobile, ou active (services en *peer-to-peer* comme Waze ou Fix my street).

Harmoniser, normaliser et standardiser la collecte

TECH

ECON

JURD

Selon les capteurs et la méthode de transmission (carte à puces, internet) des données utilisées, différents formats de données existent. L'enjeu est de favoriser une harmonisation des formats de données pour permettre aux utilisateurs de ces données d'en croiser de sources différentes et de ne pas rester captifs d'une solution logicielle particulière, avec des coûts minimum et une efficacité maximale.

Le secteur des transports illustre parfaitement cette variété de capteurs, avec l'utilisation de capteurs intrusifs, définis comme nécessitant un aménagement spécifique de la chaussée, ou non intrusifs (comme les capteurs embarqués par exemple sur des smartphones par exemple)¹⁸.

Cet enjeu d'harmonisation est aujourd'hui bien reconnu et des initiatives ont lieu afin de normaliser les méthodes de collecte et par conséquent les formats des données obtenues. C'est le cas des acteurs comme le CITEPA ou des laboratoires associés au sein des AASQA (Associations Agréées de Surveillance de la Qualité de l'Air) qui souhaitent aujourd'hui capitaliser sur les différentes méthodes de collecte et modèles développés afin d'aboutir à des résultats compatibles d'une AASQA à l'autre.

Un autre exemple est mis en avant par le Centre Scientifique et Technique du Bâtiment, la définition de standards pour représenter tous les types de bâtiments.

Par ailleurs, il ressort que les données non structurées ne sont pas assez exploitées. Ces données non structurées peuvent être définies comme les données liées aux différents services des collectivités qui existent à l'heure actuelle comme données relevées, objectives, mais encore non mises à disposition sous une forme signifiante. Ces données peuvent concerner le contenu informationnel de la ville, des services et des intervenants et avoir trait aux thématiques de l'eau, l'air, la santé publique, les services sociaux, etc. Si elles étaient structurées, et donc normalisées, ces données pourraient encourager le développement de nouveaux services et nourrir des outils de modélisation.

¹⁷ La Mobile Marketing Association (MMA) France publie le Baromètre trimestriel du Marketing Mobile.

¹⁸ NB : ce vocabulaire propre au domaine de l'équipement pourrait prêter à confusion, la collecte d'informations via les appareils embarqués étant considérés comme intrusifs par certains acteurs (ONG...).

Structurer les données en catégories significantes : l'enjeu de la construction de nomenclatures adaptées

TECH

ECON

JURD

En matière de qualité des données territoriales, un enjeu particulièrement important est celui de la structuration et de l'agrégation des données en catégories significantes et utiles. Cela facilite l'accès à la donnée « source » et générique pouvant être retravaillée. Le plus souvent, cette structuration renvoie au jeu de catégories employées pour classer les données au moment de la collecte.

Ces catégories peuvent être plus ou moins adaptées aux enjeux actuels des données. Par exemple, beaucoup d'enquêtes sur le chauffage des logements (EPTB, Sitadel, fichier logement) considèrent une seule catégorie pour les énergies de chauffage renouvelables. Or, avec le développement de ces modes de chauffage, il devient pertinent de distinguer les différents sous-ensembles pour mieux les analyser. Cet exemple souligne l'importance du choix et de la construction des nomenclatures, certains ensembles étant exclusifs ou constitutifs d'autres ensembles. On parle alors de métamodèle, c'est à dire d'un niveau de raisonnement préalable à la collecte de données qui conditionne la catégorisation et la labellisation des données.

Dans certains secteurs, plusieurs nomenclatures existent selon la finalité, l'approche retenue, la pratique courante ou les acquis historiques. Très souvent, des catégories concurrentes se recoupent sans se confondre. Par exemple, dans le domaine des transports, on distingue transports collectifs, transports en commun et transports urbains, chaque ensemble renvoyant à sa propre définition.

Pour les porteurs d'outils ou producteurs de services, la qualité de la donnée et des nomenclatures est souvent un paramètre exogène: il n'est pas possible d'éclater ou d'interpréter le contenu des catégories élaborées en amont. Cette situation pousse les outils en aval à adapter leur raisonnement aux catégories existantes, plutôt qu'à faire évoluer les catégories en accord avec le besoin¹⁹. Lorsque les nomenclatures évoluent, elles créent des ruptures de série dans la manière d'interpréter des données, sauf à procéder à de complexes corrections rétrospective. Cette difficulté accroît l'enjeu de travailler en amont des enquêtes sur des nomenclatures claires et précises applicables aux sujets étudiés.

La granulométrie spatiale et temporelle de la collecte de données : un besoin de régularité et de stabilité

TECH

ECON

JURD

Certaines enquêtes ou mesures sont effectuées de façon régulière, et traitées régulièrement également. C'est le cas par exemple des mesures effectuées par les stations de mesures des AASQA qui captent les concentrations des polluants réglementés chaque jour, soit en fond urbain soit près de sources précises (axe routier, industrie).

De la même manière, des enquêtes sont réalisées de façon régulière, généralement sur une base annuelle. On peut citer par exemple les enquêtes de l'INSEE comme l'ESANE (statistiques annuelles d'entreprise), EACEI (consommation d'énergie dans l'industrie), le recensement de la population, le fichier SITADEL (fichier national où sont enregistrés tous les permis de construire instruits pour le suivi des constructions neuves).

Cette première stratégie, couplée à une analyse régulière des résultats bruts, permet rapidement de dégager des tendances et d'observer des ruptures. De façon journalière ou hebdomadaire pour les études ou mesures les plus fréquentes, ou de façon annuelle. Elle aide également à la conception de modèles de prévision des tendances futures, tendances qui peuvent être confrontées aux mesures effectivement obtenue en temps réel afin d'ajuster et d'affiner le modèle en question.

Une deuxième stratégie est de réaliser des enquêtes de façon ponctuelle, comme c'est le cas pour les EMD (Enquête Ménage Déplacement) ou ENTND (Enquête Nationale Transport Déplacement). Ces enquêtes sont de grande ampleur, et leurs coûts importants ne permettent pas d'en réaliser plus d'une tous les 10 ans. Le manque de moyens conduit

¹⁹ La modélisation doit néanmoins se baser sur l'application que l'on veut mettre en œuvre plutôt que s'appuyer sur les données existantes pour orienter les services à proposer.

généralement à une analyse peu approfondie des résultats.

Cette deuxième stratégie n'est pas idéale, et de nouveaux services émergent pour pallier le manque de régularité, la granularité temporelle trop importante et le manque de moyens d'analyse de ce genre d'enquêtes.

Cette question de la granularité s'étend aussi au domaine spatial. Une enquête nationale ne pourra pas rendre compte avec précision et robustesse des variations locales d'un phénomène. Un exemple typique de cette situation est l'enquête nationale transports et déplacements, dont le fichier source ne peut pas être exploité à une échelle plus grande que celle de la région, car le nombre de réponses à l'enquête devient trop faible pour en tirer des conclusions significatives. Ce problème de granularité spatiale inadaptée est également important dans le domaine énergétique.

La granularité de la collecte de données, à la fois dans l'espace et dans le temps, doit être considérée au regard du besoin que l'on cherche à couvrir.

La couverture des nouvelles problématiques environnementales : orienter la collecte de données vers les secteurs peu documentés

TECH

ECON

JURD

La connaissance statistique et la production de données organisées est très inégale selon les secteurs. Traditionnellement, la production de données est la plus forte dans les secteurs fortement régulés comme l'industrie, le résidentiel ou l'agriculture. Dans les secteurs diffus et hétérogènes, comme les transports ou le tertiaire, les données sont généralement plus parcellaires. Ce constat peut s'étendre aux méthodes innovantes de traitement des sujets environnementaux, très transversaux et qui nécessitent des approches systémiques et des données complexes, à l'interface entre les domaines classiquement renseignés : bases de données ACV, données sur l'énergie grise des infrastructures et des équipements, pollutions transfrontalières et émissions du "scope 3".

Pour réduire ces écarts et améliorer la couverture des zones d'ombre, il est important de définir les priorités en matière de collecte, de mobiliser les acteurs institutionnels et privés sur les pratiques de partage et de remontée d'information, et de définir des méthodes pour capter les informations difficiles à obtenir, notamment les informations comportementales. La réalisation d'enquêtes ponctuelles est un premier pas pour comprendre où se situent les enjeux du secteur, mais elles présentent souvent le défaut de ne pas être reconduites dans le temps. D'autres méthodes complémentaires de suivi sont alors nécessaires.

4.2. La collecte secondaire : réutilisation de données existantes

4.2.1. En quoi ça consiste ?

La collecte secondaire consiste en l'obtention de données auprès de producteurs qui les ont déjà collectées, traitées et valorisées. Ces sources peuvent être d'accès libre ou d'accès restreint (club, achat, droit de propriété), et ce indépendamment du statut juridique du fournisseur de cette donnée : les structures publiques peuvent restreindre l'accès à certaines données et inversement, des structures privées peuvent choisir de diffuser librement les leurs.

4.2.2. Enjeux techniques, économiques et juridiques associés à la collecte secondaire des données

Les enjeux présentés ci-dessous, rencontrés au moment de la collecte secondaire des données, peuvent être techniques, économiques et/ou juridiques.

La connaissance et la diffusion des sets de données existants

TECH

ECON

JURD

L'accès aux données territoriales de l'environnement est conditionné par la connaissance des données existantes et de leur localisation. Pour répondre à ce besoin, existent en France :

- > Des catalogues de données, permettant de renseigner sur l'existence et la localisation de données. Ces catalogues ou plateformes ne mettent pas ces données en accès mais les documentent et peuvent en expliquer les conditions et lieux d'obtention. L'utilisateur doit ensuite se diriger vers la localisation désignée par le catalogue pour récupérer les données en question.
- > Des bases de données et outils de capitalisation dédiés, dont le rôle est d'organiser les données collectées de manière plus ou moins agrégée. Dans ce cas, la recherche de données et leur récupération se fait sur le même support. Ces bases de données ou plateformes peuvent rassembler des données plus ou moins hétérogènes et en ouvrir de manière plus ou moins large l'accès.

On répertorie aujourd'hui différentes structures ou outils de ce type :

- > Les infrastructures de données géographiques (IDG) : plateformes d'organisation des données à destination des utilisateurs de systèmes d'information géographiques (SIG) et plus généralement dispositif de partage des données territoriales.
Ces plateformes existent au niveau national et sont déclinées à tous les échelons territoriaux. Elles facilitent l'accès sur un même territoire à des données et des services géographiques issus de fournisseurs différents. Elles sont au cœur de l'écosystème des données en libre accès dans la mesure où elles structurent depuis plusieurs années les données géographiques institutionnelles (une part essentielle des données publiques libérées).
- > On parle de l'Infrastructure Globale Nationale française pour désigner la somme de plateformes nationales (de ministères comme Géo-IDE, de plateformes thématiques, etc.) et de plateformes territoriales (régionales, départementales, ou simplement locales, etc.). Ces plateformes regroupent un catalogue de métadonnées avec service de recherche, une plateforme cartographique (avec services de visualisation) et un hébergement de données à télécharger.
- > Au niveau national en France, le Géoportail, créé en 2006 sous la maîtrise d'ouvrage du MEDDE et mis en œuvre par l'IGN (Institut national de l'information géographique et forestière) et le BRGM (Bureau de recherches géologiques et minières) donne accès à des référentiels pour l'information géographique via le rassemblement de nombreuses données publiques géographiques. Il est construit dans une logique d'ouverture et d'interopérabilité des données. Il comprend un Géocatalogue, qui inventorie les différentes sources d'informations géographiques publiques et offre également des services aux particuliers, organismes publics et professionnels.

Il y a néanmoins un enjeu important lié à la prise de connaissance des données existantes rendue parfois difficile dans la mesure où les données environnementales sont dispersées dans des structures de mise à disposition nombreuses. Il

peut ainsi être difficile de savoir où chercher et où trouver une donnée particulière. En effet, en dehors des catalogues de l'infrastructure globale nationale, d'autres données environnementales, non uniquement géographiques, sont produites par des organismes variés sans être toujours recensées dans des catalogues permettant de les identifier et localiser rapidement.

Dans le cadre de la collecte de données hétérogènes pour alimenter un outil de modélisation multithématiques par exemple, le collecteur de données pourra être amené à devoir récupérer des données issues de nombreuses bases de données différentes, ou à devoir choisir où récupérer une donnée particulière produite par différentes sources. Par exemple, les rôles respectifs des AASQA (associations agréées de surveillance de la qualité de l'air) et des OREGES (Observatoire de l'énergie et des gaz à effet de serre) peuvent parfois se chevaucher et apporter de la confusion pour les acteurs extérieurs à ces structures.

Par conséquent, il existe un besoin de cadre juridique adapté pour répondre à cet enjeu de l'éparpillement des données, afin de faciliter la prise de connaissance des données existantes et de leur localisation.

La notion de qualité des données géographiques²⁰

Le développement considérable de l'information géographique, que ce soit dans un contexte professionnel ou pour une utilisation « grand public », nous oblige à reconsidérer de plus en plus sérieusement le problème de la qualité des données, qui a une influence directe sur la fiabilité des analyses spatiales produites et des décisions qui en découlent. Un cadre législatif se met progressivement en place pour veiller à la bonne documentation des données et en particulier à la description de leur qualité (géométrique et sémantique)²¹.

Les outils actuels qui permettent de mesurer, qualifier et communiquer la qualité des données semblent inadaptés aux besoins, aux enjeux et à la diversité des acteurs. Les régies ou infrastructures de données jouent ainsi souvent un rôle clé dans la validation et la qualification de la qualité des données produites et diffusées. Le monde de la recherche (Devilleers et Jeansoulin, 2005) propose deux définitions pour la qualité des données géographiques qui permettent d'envisager des protocoles de contrôles et de recette différents :

- > La qualité interne se mesure au travers de critères (précision géométrique, exhaustivité, précision sémantique, cohérence logique, actualité, etc.) définis dans la norme ISO 19113 à l'aide de méthodes explicitées dans la norme ISO 19114 et de mesures spécifiques dans la norme 19138. Pour communiquer cette évaluation de qualité on utilise les métadonnées de la norme ISO 19115. Ainsi la qualité interne vise à répondre à la question : « comment puis-je mesurer la qualité de mes données et comment le faire savoir ? ».
- > La qualité externe se définit plutôt comme une adéquation aux besoins que l'on peut résumer ainsi : « quels sont les besoins des utilisateurs en termes de données et d'informations de qualité et comment puis-je les leur offrir pour qu'ils évitent les utilisations abusives de ces données ? ». Les anglophones parlent de *fitness for use*.

Enjeux juridiques et pratiques de l'accès aux données territoriales

TECH

ECON

JURD

L'obtention des données territoriales peut se faire selon différentes modalités, certaines moyennant rémunération, d'autres gratuites ; via un contrat contraignant ou de manière libre et ouverte.

Ainsi, certains utilisateurs choisissent de contourner des sets de données payants, au risque que les données qu'ils récupèrent soient de qualité moindre par rapport à l'offre disponible. D'autres mettent en place des conventionnements ou contractualisent afin de négocier un accès aux données qui leur soit favorable. De manière générale, la question de la rémunération de la donnée peut être traitée de différentes manières et être gérée par le producteur des données lui-même (vente de sets de données, contractualisation, ...) ou bien par des structures tierces en charge de rassembler

²⁰ Référence bibliographique : Rodolphe Devillers, Robert Jeansoulin, *Qualité de l'information géographique*, Paris, Hermès, 2005, 350 pages.

²¹ En particulier, la directive européenne 2007/2/CE du 14 mars 2007, dite directive INSPIRE : <http://www.developpement-durable.gouv.fr/IMG/pdf/Inspire.pdf>.

(via l'achat, l'échange contre service, le conventionnement pour l'obtention initiale des données, etc.) et de mettre à disposition les données (selon un système de forfait, d'achat de sets, etc.) selon des modalités qui peuvent viser à normaliser plus ou moins le coût des différents sets de données.

Un mouvement en faveur de la mise à disposition libre et gratuite des données des collectivités se développe actuellement (en France notamment). Ce phénomène désigné sous le nom d'open data a commencé en 2010 à Rennes Métropole et s'est diffusé sur le territoire. Aujourd'hui, de nombreuses villes ont mis à disposition leurs données publiques sur des plateformes ouvertes et plus encore se sont engagées dans des démarches visant l'amélioration de la transparence de l'action publique et la création de services et de nouveaux modèles économiques à partir de données libérées. Au niveau national, on peut citer l'exemple de www.data.gouv.fr.

Sur la mise à disposition de données, il est aussi intéressant de citer les bases de données des observatoires régionaux de l'énergie et du climat (OREGES) et des Associations agréées de surveillance de la qualité de l'air (AASQA) qui sont également destinées à s'ouvrir. Par exemple, Basemis est une base de données territorialisée de l'énergie et des GES en Pays de la Loire, portée par l'AASQA Air Pays de la Loire.

Cependant, l'ouverture des données territoriales reste un enjeu en raison des réserves d'acteurs de la sphère publique pour partager les données :

- > Des acteurs historiques ont investi dans la constitution de bases de données, et sont devenus les maillons essentiels de la chaîne pour leur diffusion. Des opérateurs comme l'IGN ont des activités rémunératrices tout en étant incités à diffuser leurs données gratuitement (l'IGN a participé à l'élaboration de la *Charte Ville numérique* avec le CSTB, qui prévoit une mise à disposition de données et collaboration avec d'autres organismes). Ces acteurs sont positionnés entre une activité de service public et une activité commerciale, ce qui nécessite reconfiguration.
- > Mettre en accès libre les données permet de faire du contrôle a posteriori, ce qui démultiplie les acteurs capables de réagir et de critiquer les données et constitue un facteur limitant la propension de certains à diffuser leurs données.

Par ailleurs, l'accès à des données produites par des acteurs privés et venant majoritairement des usagers des services qu'ils proposent ou de capteurs utilisés dans le cadre de leur activité industrielle ou commerciale, est également restreint. En effet, ces acteurs veulent garder le bénéfice des données qu'ils produisent. Entrent par exemple dans cette catégorie les gestionnaires de plateformes *peer-to-peer*, de flottes de véhicules partagés, les opérateurs de réseaux télécom et mobiles, les fournisseurs d'énergie, les gestionnaires de centres commerciaux ou encore des acteurs comme Google. Souvent, ils ont conscience de la valeur des données qu'ils traitent et parfois développent une offre de service spécifiquement basée sur le traitement des données produites par leurs activités principales.

Par exemple, Orange Business Services a lancé le service FluxVision, offre consistant à récupérer d'énormes quantités de données issues de l'usage des téléphones mobiles pour fournir des indicateurs de fréquentation et de flux de déplacements sur le territoire.

On assiste cependant de plus en plus à des initiatives de conventionnement pour du partage de données entre organismes publics. Ce genre de partenariat se développe également progressivement entre des organismes publics et privés afin de mutualiser la production et le partage de certaines données.

Par exemple, le projet BANO de Base d'Adresses Nationale Ouverte initié par OpenStreetMap France, ayant pour objet la constitution d'une base la plus complète possible de points d'adresse à l'échelle de la France, est le fruit d'un partenariat entre L'IGN, OpenStreetMap et La Poste. Cette base est disponible selon un double modèle de licence : gratuite en repartage et tarifiée sans repartage.

La sécurisation des données : un enjeu qui dépasse le cadre des données territoriales de l'environnement

TECH

ECON

JURD

La sécurisation des données renvoie au besoin de limiter la diffusion des données. Par exemple, lors de transactions bancaires, ou de partages d'éléments relevant de la vie privée, il est important de disposer d'un cadre technique et juridique, empêchant et interdisant que les données transmises soient diffusées.

Cette sécurisation passe le plus souvent par des processus techniques tels que le chiffrement des échanges : les informations sensibles sont transmises de manière à n'être intelligibles que par le récipiendaire intentionnel de l'échange. Du point de vue du stockage, des mesures de sécurité de nature informatique sont indispensables. Toutefois, les réseaux ne sont pas infaillibles et sont exposés à des attaques de nature malveillante.

Dans le domaine des données territoriales de l'environnement, les données les plus sensibles sont les données de production industrielle et celles qui se rapportent aux comportements individuels. A ce titre, les opérateurs de réseaux de distribution, ou les gestionnaires de bases de données relatives aux déplacements des individus sont concernés.

La documentation : une pratique indispensable pour assurer la transparence des sets de données et de leurs traitements

TECH

ECON

JURD

Le constat a été fait que la documentation sur les outils existants est très faible, et est généralement une étape négligée au cours de l'élaboration de ces derniers. De la même manière que pour le développement des outils, les étapes de collecte, de traitement et de valorisation des jeux de données destinés à être publiés sont aujourd'hui rarement documentées. Il y a un enjeu de coût de cette documentation (moyens humains et financiers nécessaires) et de savoir-faire.

Pour aller plus loin, la documentation de la source des données peut être considérée comme l'enjeu le plus important. En effet, les données peuvent être incomplètes, de format non standardisé, pas assez ou trop agrégées : l'utilisateur, s'il est sûr qu'elles sont de source fiable, parviendra en général, par traitement, à les transformer pour pouvoir en faire un usage convenable dans le cadre du service développé. A l'inverse, des données a priori parfaitement adaptées à un usage, si elles ne sont pas documentées, ne peuvent garantir les résultats qui en sont obtenus.

Les données documentées peuvent être partagées et réutilisées.

Par ailleurs, plus les données sont élaborées, avancées, plus la documentation est importante, pour compenser le risque d'effet « boîte noire ». A partir du moment où tout est documenté, les données brutes ou retraitées sont tout aussi intéressantes et signifiantes, et le choix entre les deux dépend de l'usage qui en est fait ensuite. Le travail mené actuellement par le CERDD Nord-Pas-de-Calais (porteur de l'observatoire régional de l'énergie et des GES) en matière de cadrage des indicateurs, par le biais de fiches métadonnées, apparaît à ce titre particulièrement intéressant.

Désagrégation et niveau de détail des données : un besoin récurrent de la part des porteurs d'outils et de services

TECH

ECON

JURD

Le créateur de service peut avoir des difficultés pour atteindre le niveau de désagrégation dont il a besoin pour les données nécessaires. Les données peuvent exister sans être publiées, ou publiées uniquement auprès d'un public restreint. Elles peuvent également exister sous forme brute sans être exploitables en l'état. Si les données ont déjà été traitées, il est également probable que le traitement apporté par un acteur différent de l'utilisateur confère aux données brutes un niveau d'agrégation par défaut qui ne correspond pas aux besoins de ce dernier, ainsi qu'aux utilisateurs potentiellement intéressés.

Pour reprendre l'enjeu précédent (documentation), si les données sont documentées, et donc qualifiées, alors l'atteinte du niveau de désagrégation nécessaire et suffisant par l'utilisateur lui-même sera plus facile.

Cet enjeu est en lien direct avec la phase de traitement des données produites ou collectées, explicitée en particulier dans la suite du document et notamment sur le sujet de « l'accès aux données à un niveau de détail et de désagrégation adaptés aux usages ».

4.3. Le traitement : enrichissement et intégration des données

4.3.1. En quoi ça consiste ?

Le traitement des données est un ensemble d'étapes par lesquelles la donnée change de forme, de nature ou de signification pour son utilisateur. On peut distinguer deux types de traitement des données :

- > La capitalisation des données vise à centraliser, organiser et standardiser les données territoriales dans le but d'être utilisées directement par d'autres utilisateurs (diagnostic, observation) ou de passer par des étapes de traitement supplémentaires (par exemple de modélisation) ;
- > La modélisation consiste à utiliser dans des calculs (principalement effectués via des processus informatiques) les données dans le cadre d'une représentation simplifiée d'un système (ou modèle). L'objectif du traitement par la modélisation est de comprendre, décrypter et analyser des phénomènes à partir des données qui leurs sont liées.

Focus informatique #3 - les traitements appliqués aux données

L'objectif des traitements informatiques réalisés sur les données reçues est de rendre ces données intelligibles et pertinentes pour les potentiels réutilisateurs. En effet, une donnée « brute » est souvent difficile à utiliser dans son format initial (informations manquantes, libellés à harmoniser, etc.) ou incomplète (par exemple on dispose d'une code postal mais en fait on a besoin des coordonnées GPS) et donc difficilement exploitable en l'état, soit par un utilisateur individuel soit par un programme informatique.

L'amélioration de la qualité des données repose sur plusieurs types d'actions :

- > Ajout de métadonnées pour donner plus d'informations contextuelles sur la donnée gérée : désignation précise, identifiant unique dans la base de données, catégorisation, descriptif textuel, historique des versions, type de licence, tags (mots-clés), etc.
- > Normalisation, c'est-à-dire la vérification de l'intégrité et le « nettoyage » des données. Par exemple : traitement automatique des caractères accentués, technique de filtres pour associer automatiquement des termes orthographiés différemment, etc. Des outils *open source* gratuits existent pour faciliter ce type de manipulation sur les données (par exemple OpenRefine).
- > ETL (*Extract, Transform & Load*) : terme informatique qui désigne l'import de données provenant de sources hétérogènes et réalisé sous forme d'un *workflow* automatisé. Cela permettrait par exemple d'importer et agréger à la volée des données de consommations kWh provenant de plusieurs fournisseurs d'électricité ou de plusieurs départements.
- > *Data mapping* : mise en correspondance de champs (colonnes) de données identiques, présentes dans plusieurs fichiers de données mais éventuellement nommés différemment.
- > *Geocoding* ; ajout automatique des coordonnées latitude et longitude, sur des adresses ou des noms de communes.
- > Traduction automatique des données afin de proposer des *datasets* multilingues.

► Pour aller plus loin :

OpenRefine : <http://openrefine.org>

Screenshot issu du logiciel ETL de référence Talend : <http://www.talend.com>

4.3.2. Enjeux techniques, économiques et juridiques associés au traitement des données

Les enjeux présentés ci-dessous au sujet du traitement des données, peuvent être techniques, économiques et/ou juridiques.

L'accès aux données à un niveau de détail et de désagrégation adaptés aux usages

TECH

ECON

JURD

Il est rare que les données territoriales soient utilisées directement telles qu'elles sont collectées ou mesurées. De la même manière qu'en industrie où les matières premières extraites de l'environnement sont progressivement purifiées et améliorées en vue des traitements intermédiaires, les données territoriales sont enrichies par différents procédés tels que :

- > La construction d'indicateurs statistiques permettant de caractériser le jeu de données : calcul de la moyenne, médiane, écart-type, variance, d'une série de données ;
- > La correction des valeurs incohérentes, des valeurs non pertinentes, des extrêmes, des valeurs limites ;
- > La correction géographique, climatique ;
- > La pondération d'un échantillon en vue d'une extrapolation à l'échelle d'une région ou d'un territoire ;
- > L'agrégation (totaux) des cas de l'échantillon selon des catégories prédéfinies ;
- > La modélisation de données manquantes pour compléter l'échantillon de données mesurées ou enquêtées.

On notera qu'à chaque étape, des hypothèses explicites ou implicites sont posées quant à la signification des données traitées. Plus la donnée est enrichie depuis sa source initiale, plus les hypothèses prises en chemin pèsent dans le résultat final. Cette tendance à l'enrichissement des données justifie le besoin de documentation sur cet enrichissement et plus généralement sur les opérations effectuées sur les données.

Outre les traitements techniques et statistiques, les données peuvent être croisées entre elles pour produire des indicateurs composites ou pour aboutir à d'autres sets de données. Par exemple, un set de données décrivant les consommations énergétiques des ménages d'un territoire liées aux transports peut être le fruit d'un croisement entre une base de données des déplacements du territoire, recueillis sur un échantillon par une enquête et une modélisation du parc de véhicules des ménages sur la base d'un registre des ventes annuelles. Dans cet exemple, on comprend que le set final sur la consommation d'énergie accumule les hypothèses posées dans l'élaboration des deux sets initiaux. Le caractère plus ou moins élaboré des procédures de croisement de données implique, dans le cadre d'outils élaborés de type observatoire, de développer des structures d'accueil de données adaptées permettant le traitement de données hétérogènes.

La qualification des données

TECH

ECON

JURD

En Europe, la Directive INSPIRE, s'appliquant aux informations liées à un cadre géographique détenues sous format électronique par des autorités publiques ou en leur nom (à la fois données de références ou données plus spécifiques à la thématique environnementale), indique que ces données doivent être assorties de métadonnées complètes, notamment relatives aux conditions applicables à l'accès et à l'utilisation des informations géographiques, à la qualité et à la validité de ces informations, aux conditions d'accès ainsi qu'aux autorités publiques qui en sont en charge. Ainsi, alors que les textes antérieurs formulaient des obligations et des modalités administratives d'accessibilité des données publiques, INSPIRE oblige à la création de dispositifs à dimension technique importante et nécessitant notamment la définition de nouvelles normes concernant la qualification des données.

La qualification des données et la constitution de métadonnées sont essentielles pour être en mesure d'avoir une traçabilité des données (savoir d'où provient un résultat, revenir à la source de l'information) et constituent des gages de qualité des données produites et diffusées. La qualification des données, qu'elles soient brutes ou enrichies, constitue une plus-value importante pour les données. En raison du volume gigantesque de données produites et mises à disposition, cette qualification permet aussi de gagner du temps dans la recherche de données.

Les métadonnées, éléments de description, concernent en grande partie la qualification et la documentation des données. Par exemple, il peut s'agir, pour la photographie aérienne d'une route, d'informations nécessaires à la compréhension de la photographie (moment de la prise de vue, localisation géographique, auteur, etc.).

Plus les données considérées sont croisées et/ou enrichies, plus leur qualification est importante pour éviter un effet "boîte noire" de ce type de données. De même, lorsque ces données sont intégrées aux modèles de traitement d'outils de modélisation, il est important de pouvoir en connaître la source et les déterminants pour éviter cette fois un effet boîte noire de l'outil et pour faciliter la vérifiabilité de ses résultats.

Exemple : Concernant les données géographiques en France, les catalogues de données de l'infrastructure globale nationale sont interopérables et permettent, par un système de moissonnage (i.e. d'interconnexion des catalogues entre eux) de faire remonter des fiches de métadonnées dans le Géocatalogue. Les plateformes régionales ont ainsi un rôle de fédérateur intermédiaire des plateformes sous-jacentes.

Par ailleurs, les pratiques en matière de qualification des données ont évolué parallèlement aux pratiques de mise à disposition des données dans le cadre du développement de l'open data. Ainsi, les premières plateformes open data se voulaient neutres face aux données qui devaient être publiées rapidement et sans traitement et n'étaient pas forcément assorties de métadonnées. Ensuite, les collectivités ont commencé à porter un effort important pour la qualité des données et des métadonnées qu'elles publiaient.

Le modèle de l'initiative www.data.gouv.fr comporte des caractéristiques de ces deux approches : les données sont publiées sans traitement et sont doublées d'un système de notation sur leur qualité et les manières de participer à leur amélioration.

Le contrôle et la vérification des données

TECH

ECON

JURD

La vérification des données territoriales est essentielle : ces données peuvent être utilisées pour le diagnostic et la décision en matière de planification urbaine, en temps réel ou à moyen ou long terme. Les créateurs de services basés sur les données territoriales ne peuvent se permettre de travailler avec des données non fiables ou dont la fiabilité est méconnue.

Il y a deux façons de contrôler et vérifier des données. Pour les données issues de mesures, ce sont les performances techniques des capteurs qui doivent être contrôlées de façon scientifique. Quant aux données estimées ou modélisées, elles peuvent être comparées à la réalité (elle-même "mesurée par des capteurs") en temps voulu et donc plusieurs années après l'estimation. Il est donc important de vérifier la robustesse des modèles produisant des données, ce qui peut être réalisé en comparant les résultats du modèle testé aux résultats obtenus par un autre outil ou une autre méthode, déjà validés, avec les mêmes données d'entrée et hypothèses.

4.4. La valorisation : compréhension, diffusion et modèles économiques

4.4.1. En quoi ça consiste ?

Certains acteurs s'intéressant au potentiel rendu par le développement de l'ouverture des données font le constat d'un paradoxe : bien que de nouveaux usages des données apparaissent (notamment chez les acteurs privés), l'accessibilité des données ces dernières années n'a pas toujours conduit au développement de nombreux nouveaux usages et services à partir de ces données et en lien avec le développement territorial tel qu'anticipé par ces acteurs. De plus, les services proposés ne sont pas tous innovants et adaptés aux besoins.

L'encadrement et l'animation concernant l'ouverture, le partage et la mise à disposition des données sont donc essentiels pour permettre l'émergence de nouveaux services participant au développement des territoires. Cela repose notamment sur le développement de l'acceptabilité du partage des données pour les entreprises et les individus. Cependant, l'ouverture des données n'est pas l'unique levier pour développer l'usage des données et n'est pas toujours possible. D'autres systèmes pourraient être envisagés, par exemple de mise à disposition restreinte des données ou encore de mise en réseau d'acteurs.

D'autre part, l'identification des usages et des usagers est un facteur essentiel pour le développement de services pertinents. Il est aujourd'hui difficile, mais crucial, d'identifier les usages, à cause de la diversification et la multiplication des usages possibles et des usagers.

Focus informatique #4 - techniques informatiques de partage ou de mise à disposition des données

Dans la perspective de faciliter au maximum la réutilisation des données par des tiers, plusieurs solutions informatiques peuvent être mises en œuvre pour optimiser le partage des données gérées :

- > Disposer d'un **moteur de recherche** multicritères et *full text* performant sur la plateforme de partage des données et indexant l'intégralité des bases utilisées.
- > Proposer des **API** (*Application Programming Interface*) stables et documentés : ce sont des bouts de code informatique réutilisables librement par des tiers au sein des applications qu'ils développent et facilitant/automatisant les requêtes aux données de la plateforme.
- > Bénéficier de la **connectivité** et de l'**interopérabilité** avec d'autres plateformes « open data » existantes en s'appuyant par exemple sur un standard open source tel que CKAN.
- > Faire du **SEO** (*Search Engine Optimization*) : important pour permettre de trouver facilement les données au moyen de moteurs de recherche standard (Google, Yahoo, Bing).
- > Proposer des **téléchargements de fichiers** sous une palette de formats standards la plus large possible (CSV, XML, SHP, etc.) à partir des enregistrements originaux ou recomposés. La plateforme doit aussi proposer des outils de conversion automatique entre différents formats de données standard pour laisser plus de libertés à l'utilisateur de choisir le format qui lui convient le mieux.
- > Proposer des possibilités variées de **datavizualisation** (sous forme de tableau de bord, de cartographie, etc.) pour faciliter la représentation graphique des données de manière agréable pour l'utilisateur (graphique téléchargeable ou à afficher sur un site tiers).
- > Mettre à disposition des **outils de data mining** (analyse des données) et de traitement big data classiques : ACP (Analyse en Composantes Principales), *K-means*, *clustering*, etc. : de cette manière la puissance de calcul de la plateforme est partagée et mise à disposition des utilisateurs pour les attirer et les inciter à partager leurs données directement sur la plateforme.
- > Respecter la **norme Linked Data** (ou « Web des données » en Français) qui est le standard du web sémantique pour structurer et publier des jeux de données en ligne en les reliant les uns aux autres et ainsi constituer un réseau global d'informations. Cette norme s'appuie sur le format RDF (*Resource Description Framework*).
- > Laisser la possibilité de **faire du mashup** avec les données provenant de la plateforme, c'est-à-dire pouvoir afficher sur des sites tiers des données ou des visualisations provenant de cette plateforme, à la manière d'une vidéo Youtube par exemple.

► Pour aller plus loin : CKAN : <http://ckan.org>

4.4.2. Enjeux techniques, économiques et juridiques associés à la valorisation des données

Les enjeux liés plus particulièrement à la valorisation des données, peuvent être techniques, économiques ou juridiques, et sont donc classés en ce sens.

La responsabilité du publicateur de données : une exigence qui fait parfois obstacle à la diffusion des résultats

TECH ECON JURD

Pour un organisme détenteur de données collectées, les publier, les valoriser ou les mettre à disposition implique d'assumer publiquement leur qualité et leur justesse. Avant de procéder à cette publication, il est souhaitable de prendre le temps de s'assurer que les données publiées sont de bonne qualité et ne seront pas remises en question par les éventuels utilisateurs.

Cette étape demande des moyens techniques et financiers que les institutions qui détiennent les données ne peuvent ou souhaitent pas toujours consacrer à cette tâche. Il en résulte deux attitudes problématiques: la non-publication de sets de données valables ou la publication de sets de données de mauvaise qualité. Dans les deux cas, cela entraîne des difficultés pour les utilisateurs en aval.

En plus de cet enjeu technico-économique, on peut relever la réticence d'organismes à publier des données qui pourraient être utilisées par des observateurs extérieurs afin de critiquer la gestion d'une problématique ou d'un service.

Faire connaître et comprendre : appréhension de certains aspects de la réalité par la valorisation et la diffusion des données territoriales

TECH ECON JURD

La Directive INSPIRE formule des obligations relatives à la valorisation des données publiques territoriales. Notamment, elle énonce l'obligation pour les collectivités de mettre à disposition des services normalisés de visualisation, de traitement et de téléchargement des données. Par ailleurs, les collectivités ont intérêt à mettre en œuvre des stratégies pour favoriser la valorisation et la diffusion des données de leur territoire, afin de bénéficier de ces données utiles pour de l'aide à la décision.

La valorisation des données est encouragée par les capacités d'agrégation et de traitement de ces données. Elle est opérée à travers **la création d'outils d'aide à la décision pour les décideurs territoriaux et économiques et le développement de nouveaux services sur les territoires.**

Elle peut prendre différentes formes comme :

- > Un traitement appliqué à des données brutes avant leur diffusion (agrégation, amélioration de la qualité, etc.)
- > La production d'un ou plusieurs rendus visuels capables de mettre en évidence la compréhension de certains aspects d'une situation réelle par des données (graphiques, cartes figées ou interactives, etc.);
- > La mise à disposition en ligne sur des supports interactifs (par exemple : outils d'observation et/ou analyse du territoire avec interface graphique interactive);
- > La publication au sein d'études, de rapports ou de plaquettes de présentation d'information produites par les acteurs des collectivités;
- > La mise à disposition des données de sortie d'un outil à destination d'utilisateurs futurs, que ce soit au travers d'un accès ouvert ou d'un accès restreint;
- > Le soutien à la prise de décision ou la régulation, y compris aux travers de projections, estimations, scénarios, simulations, hypothèses, etc. Cela peut concerner aussi bien une décision humaine qu'un acte de régulation par une machine, comme pour les outils de régulation du trafic routier;
- > L'élaboration d'un indicateur composite (exemple : IDH) ;
- > L'établissement d'un bilan, de tableaux de bord, suivant des règles de comptabilité précises, en permettant de mettre à jour une série (par exemple: série annuelle ou mensuelle des installations d'énergie renouvelables).

Les formes et les stratégies de valorisation des données varient en outre selon les objectifs et les publics visés par la

mise à disposition des données.

On peut en particulier distinguer trois objectifs pour cette mise à disposition.

Usage des données par des organismes publics et entreprises privées du territoire :

Dans le cadre du développement de l'open data, on remarque une évolution de la stratégie de valorisation des données, au fil du temps de plus en plus tournée vers l'incitation à la réutilisation des données produites :

- > Au début, l'objectif se limitait à l'appropriation de la donnée brute par la visualisation, qui impliquait l'utilisation de données très propres pour avoir des visualisations pertinentes²².
- > Ensuite, l'accent a été mis sur les récepteurs et les réutilisateurs des données. Il s'agit s'appuyer sur des programmes d'information et de médiation autour des données pour stimuler leur réutilisation et valorisation (via des outils, applications, services, etc.). Par exemple, à travers le programme Infolab, la Fing explore les médiations possibles pour accroître la capacité des acteurs d'exploiter les données dans leurs activités.

Participation citoyenne :

La valorisation via des outils interactifs à destination des citoyens peuvent permettre de favoriser la participation du citoyen dans la gestion et le développement de sa ville. Le processus participatif se base alors sur des diagnostics et des mesures plus complets et l'intelligence collective pour la coproduction de projets est facilitée par des dispositifs numériques. Ainsi, le citoyen participe à la valorisation des données par deux biais :

- > Il révèle l'état de la ville (capteurs environnementaux citoyens, application de signalisation d'incident, etc.) ;
- > Il participe à la co-construction de l'action publique à partir des informations issues des données auxquelles il a accès plus facilement.

Aide à la décision publique :

La valorisation des données territoriales, en particulier via des outils d'aide à la décision, peut permettre de relayer et véhiculer des messages directement auprès de l'utilisateur (services techniques de la collectivité) et à destination du décideur (Directeurs généraux, élus...). Les formes ludiques que peuvent prendre ces rendus (exploration 3D, localisation des "points chauds" du territoire, animations, données en "temps réel", etc.) influencent les décisions et plus généralement la manière dont un utilisateur ou un bénéficiaire envisage la nature d'un problème environnemental. Par exemple, des figurés réticulaires (faisant apparaître des réseaux) pour représenter la mobilité sont plus illustratifs de la nature systémique des transports qu'une série d'histogrammes. Les sorties 3D donnent l'illusion de voir en direct les problématiques du territoire.

L'interopérabilité des services et des outils élaborés à partir des données

TECH

ECON

JURD

Le sujet traité ici est bien celui de l'interopérabilité des services créés à partir des données territoriales, et non de l'harmonisation des données servant à la création de ces services par la mise en place de standards (cf. *Harmoniser, normaliser et standardiser la collecte*, en 4.1.3).

Il s'agit pour le concepteur de faire en sorte que le service proposé réponde à un besoin et soit interopérable avec d'autres services répondant à d'autres besoins. Par ailleurs, l'interopérabilité des services permet également que les résultats d'un service ou d'un outil soient les données entrantes d'un autre service.

Le caractère générique d'un modèle dépend du niveau de définition primaire de ses composants, c'est à dire à du niveau de généralité auquel le composant du modèle est initialement conçu. Par exemple, pour représenter une route et une rivière, le modèle moins générique utilisera deux entités, "route" et "rivière", tandis que le modèle plus générique utilisera l'entité "segment", dont les attributs particuliers en feront une route ou une rivière. Généralement, plus un modèle est générique, plus il est adaptable, ouvert et réceptif à l'introduction de nouveaux modules ou composants. Mais les modèles génériques demandent plus de temps et d'investissement de départ, puisqu'on passe beaucoup de temps à décrire des éléments génériques dont seulement quelques attributs serviront dans les premières mises en œuvre du modèle.

²² Ce type de valorisation est lié au mode de mise à disposition des données qui favorise la publication de données de qualité associées à des métadonnées.

5. Pour une utilisation renforcée des données : une gouvernance et une organisation à mettre en œuvre

5.1. Principes généraux des structures de gestion des données existantes

Le projet de mettre en place un modèle de gouvernance des données répond à un besoin identifié d'organiser un écosystème autour des données territoriales qui favorise la mise en réseau des acteurs, le partage des données et l'amélioration de leur qualité. Par ailleurs, l'organisation d'une gouvernance des données doit être un moteur pour l'émergence de nouveaux services territoriaux à partir de ces données ou l'amélioration de services existants.

Parmi les exemples de modes de gestion des données identifiés dans le cadre de cette étude, certaines adoptent l'appellation de *régie* de données (ou de gestion des données), qui nécessite quelques éclaircissements. En effet, *régie* désigne en droit public français un établissement public chargé de la gestion d'un service public ainsi que le mode de gestion de ce service. Une régie est donc un mode d'organisation permettant aux collectivités de prendre en charge une activité dans le cadre de leurs services²³.

Dans le cadre des structures de gestion des données existantes, le concept de *régie* selon le droit public français a ainsi été utilisé pour la Régie de Gestion des Données des Pays de Savoie, tournée vers les acteurs publics du territoire, où il désignait le statut de régie à caractère industriel et commercial de la structure. Un concept de « régie de données » ne correspondant pas à la définition ci-dessus a récemment été formalisé avec les travaux de la FING²⁴ et du programme Datact (mené par le groupe Chronos et le Hub agence), à partir du développement de portails open data et du constat que la libéralisation des données ne conduisait pas automatiquement au développement de nouveaux usages et services autour des données territoriales ou bien que les services proposés n'étaient pas très innovants et manquaient d'ambition. Dans le projet du programme Datact, le modèle de gouvernance appelé *régie* de données est lié à l'open innovation (innovation ouverte) et vise notamment à faire fonctionner les départements de recherche des entreprises en réseau. Aujourd'hui, la *régie* de données telle que décrite dans le programme Datact n'a pas encore rencontré de mise en œuvre concrète, bien qu'elle soit actuellement dans une phase d'expérimentation sur le territoire de Plaine Commune (projet Données communes). En effet, on n'observe pas d'exemple concret de régie de données qui, en s'appuyant sur de l'innovation ouverte, propose un dispositif totalement défini.

Par ailleurs, d'autres structures de gestion des données existent et permettent d'avoir un aperçu de la diversité des visions et des fonctionnements de ces structures. Les paragraphes qui suivent s'attachent à décrire ces différentes initiatives, afin de présenter le panorama de l'existant en matière de structures de gestion des données territoriales.

5.1.1. La Régie de Gestion des Données des Pays de Savoie

La Régie de Gestion des Données des Pays de Savoie (RGD 73-74) est un exemple de mise en place d'une structure de gestion des données territoriales conçue comme un service public bi-départemental et mise en place pour les collectivités locales et les organismes publics, ayant le statut de régie à caractère industriel et commercial. Elle n'a ainsi pas pour vocation première, à la différence du concept de régie de données élaboré par Datact, de stimuler la création de nouveaux services par des acteurs privés. Elle est uniquement tournée vers les acteurs publics et les abonnés à la régie comprennent ainsi des services de l'Etat (DDAF, DDE, gendarmerie, RTM, etc.), des services départementaux (voirie, transport, aménagement du territoire, etc.), des communes, des EPCI et des structures à l'échelle départementale (Chambre d'Agriculture, SDIS, SAFER, SEDHS, SEA, etc.) avec en tout plus de 650 organismes. Cette régie existe depuis environ 20 ans. Sa mise en place a été initiée par le Conseil Général de Haute Savoie et motivée par la prise de conscience de la nécessité de disposer d'un outil de connaissance du territoire à disposition des collectivités pour les aider à répondre aux enjeux territoriaux en se basant sur des éléments objectifs. Les données diffusées comprennent des données de base telles que cadastres, photographies aériennes, POS-PLU, cartes

²³ Il existe trois catégories de régies : la régie directe (intégrée aux services), autonome (dispose d'organes de gestion et d'un budget propres) et personnalisée (placée sous le contrôle de la collectivité mais sous régime comptable privé).

²⁴ Ex. : Programme Infolab, qui vise à mettre les acteurs en capacité d'exploiter les données dans leurs activités et projets par des médiations et en particulier un dispositif collaboratif pour aider à la compréhension, la manipulation et l'exploration des données numériques.

topographiques et géologiques, occupation réelle des sols, référentiel d'adressage, fond de plan voirie, plans de ville, altimétrie, etc. D'autres données sont en outre apportées et actualisées par chaque partenaire public (données d'urbanisme, réseaux de distribution ou de communication, périmètres réglementaires, etc.).

La mise en place de la régie a permis plusieurs réalisations. Par exemple, une base de données agricole a été constituée à moindre coût et un Observatoire des Territoires de la Savoie a vu le jour²⁵. La régie de données a favorisé le rapprochement d'acteurs qui n'avaient pas pour habitude de travailler ensemble sur des projets communs de territoire basés sur l'utilisation des données. Par exemple, la Chambre de l'agriculture et la direction départementale d'agriculture ont travaillé en collaboration sur la base de données agricole.

Organisation et gouvernance

La régie de gestion des données des pays de Savoie connaît une structure stable et une gouvernance bien établie propres à nourrir les réflexions quant aux structurations à définir pour des initiatives de régie de données à venir. Elle est supervisée par un conseil d'administration et par son statut de régie à caractère industriel et commercial qui respecte donc le cadre réglementaire pour ce type de structure.

La RDG 73-74 est avant tout guidée par un principe de mutualisation, via le regroupement des compétences et des conseils, permettant des économies d'échelle et une forme de péréquation entre petites et grandes structures.

La régie fonctionne ainsi selon un système d'abonnements et de droits d'entrée qui permet de couvrir les frais de fonctionnement et permet depuis quelques années de faire des investissements (intégration de toutes les données, consolidation des bases de données, évolution des outils). Ce sont les coûts d'abonnement qui font l'objet d'un système de péréquation entre les grands et petits organismes. Ainsi, une petite commune peut payer jusqu'à cinq fois moins cher qu'une grande.

Deux ensembles de données sont traités par la régie, et nécessitent des types de contractualisation différents :

- > les données de base : toutes les données faisant l'objet d'une délégation à la régie ou dont la constitution ou l'actualisation sont l'initiative de la régie (cadastre, données d'affectation des sols, etc.);
- > les données partenariales : les données gérées par les partenaires et transmises à la régie dans le cadre d'une convention de partenariat établie avec les abonnés ainsi que quelques données apportées dans le cadre de partenariats avec des structures non abonnées. Dans le cas des structures non abonnées, l'accès aux données se fait via une mise à disposition moyennant un coût fixé par la régie.

Le modèle de gouvernance et la structuration de la RGD 73-74 sont intéressants à étudier dans la mesure où la RGD 73-74 est stable, indépendante financièrement et réalise environ 300 000 euros d'investissements annuels qui lui permettent d'actualiser son réservoir de données et d'améliorer ses services sans faire appel à des subventions publiques. De plus, on peut la qualifier de gouvernance vertueuse dans la mesure où des services se sont développés grâce à elle (base de données agricole, observatoire économique, etc.).

Cependant, cette régie est considérée par certains acteurs comme autoritaire et nécessite la signature de nombreuses conventions pour en faire partie. L'ouverture des données est ainsi conditionnée à ce cadre strict et n'est en outre pas destinée aux acteurs privés territoriaux. Enfin, elle n'a pas vocation à stimuler le développement de nouveaux services liés aux données issues du Big Data et/ou produites par des services proposés par des acteurs privés de la ville. Au contraire, elle génère elle-même des services et son modèle de diffusion n'a pas permis à d'autres acteurs - issus des sphères privée et associative - de développer des services ou usages innovants à partir de ces données. Ce type de gouvernance pourrait intégrer plus de flexibilité et s'ouvrir aux acteurs privés pour susciter l'innovation.

²⁵ L'Observatoire porte sur les thématiques de l'agriculture, de la nature et biodiversité, de la forêt, de l'eau, de l'aménagement, de l'urbanisme, des transports, des risques et du climat. Il comprend plusieurs entrées selon le niveau de collectivité concerné et propose des fonctionnalités de cartographie interactive. Lien vers l'Observatoire : <http://www.observatoire.savoie.equipement-agriculture.gouv.fr/>. Sa création a été initiée par la DDT de la Savoie et a associé les autres services de l'Etat (DDE, préfecture, DIREN principalement), des établissements publics tels que le parc national de la Vanoise et l'ONF, ainsi que la Chambre d'Agriculture et le Conseil Général.

5.1.2. Les IDG : structures de gestion des données de l'information géographique

Les infrastructures de données géographiques (IDG) décrites plus en amont de ce rapport d'étude investissent un périmètre spécifique au sein de l'ensemble des données territoriales : celui des données géographiques publiques. Elles couvrent des thèmes variés ayant trait à la gestion de l'espace naturel et à celle de l'espace urbanisé, comme la mer et le littoral, les zones humides, la gestion des risques, l'urbanisme, l'aménagement numérique, l'adresse, etc.

Ces structures présentent des similitudes avec le concept de régie de données développé par Datact (Cf. 5.1.4.) et la RGD 73-74 (cf. 5.1.1.). Ces infrastructures rassemblent en effet les données, les réseaux informatiques, les normes et standards, les accords organisationnels et les ressources humaines nécessaires pour faciliter et coordonner le partage, l'accès et la gestion des données géographiques. Elles participent en outre à l'animation de communautés de pratiques autour de l'information géographique et animent des groupes de travail visant notamment à co-produire ou harmoniser des données territoriales thématiques (urbanisme, environnement, aménagement numérique, etc.). Elles sont donc à la fois des outils de capitalisation des ressources numériques et des lieux de rencontre des producteurs institutionnels de données.

Les IDG nationales et régionales s'inscrivent dans l'effort (guidé par la directive européenne INSPIRE) de standardisation des données et des modèles par de nouvelles règles de mise en œuvre. En particulier, ces règlements incluent des modalités de géoréférencement, la définition et classification des objets géographiques et des spécifications sur la structuration des futures bases de données. Le but de cette standardisation est de faciliter l'interopérabilité des données et des systèmes pour.

Organisation et gouvernance

Depuis les années 1990 se développent des Infrastructures de Données Géographiques nationales, déclinées ensuite à tous les échelons territoriaux.

- > **Les IDG nationales** permettent aux services de l'Etat principalement d'organiser les connaissances sur une thématique donnée et d'améliorer l'accès aux données. Elles répondent à des attentes politiques formulées sous forme de décrets, directives, livres blancs, etc. liées à l'environnement.
- > **Les IDG régionales** existent sous deux formes. Elles peuvent être portées par l'Etat et servir de guichet régional entre les différents services déconcentrés ou bien co-pilotées par l'Etat et la Région. Ensuite, il existe quelques IDG portées uniquement par des structures régionales. Le mode de gouvernance de ces IDG régionales est en outre souvent partagé entre les services déconcentrés de l'Etat (SGAR ou DREAL) et les collectivités territoriales (CR ou CG).
- > **Les IDG départementales** existent et certaines fusionnent avec des IDG interdépartementales, dans la logique de la réorganisation territoriale. D'autres IDG couvrent plusieurs territoires : deux départements, un bassin versant, une chaîne de montages ou des pays.

L'articulation entre les différentes plates-formes se fait de manière horizontale (plates-formes de même échelon territorial) et verticale (plates-formes d'une même thématique). Cette articulation consiste en des échanges techniques (métadonnées, données, services Web,...) et par des échanges (rencontres thématiques, réunions de travail, partenariats, etc.).

Les IDG nationales sont ainsi construites par une grande variété d'acteurs, ce qui les fait apparaître comme des innovations organisationnelles. Elles font évoluer les rôles des administrations centrales qui passent d'organisations accumulant les données publiques à organisations au cœur des réseaux de bases de données²⁶. Les données sont ainsi gérées à l'échelle locale à différentes étapes de leur cycle de vie (collecte, stockage, traitement). Les administrations centrales ont pour mission d'agrégier ces données au niveau national ou international, qui restent localisées au niveau local. L'articulation entre les niveaux national et régional est encouragée par des réseaux comme celui des CRIGES²⁷ qui permet aux IDG régionales de dialoguer et de s'exprimer collectivement auprès de structures nationales.

²⁶ Matthieu Noucher, « Infrastructures de données géographiques et flux d'information environnementale », *Netcom* [En ligne], 27-1/2 | 2013, mis en ligne le 11 juin 2014

²⁷ Réseau des centres régionaux d'information géographique animé par l'AFIGEO.

Les IDG de niveau régional sont particulièrement nombreuses et jouent un rôle important d'intermédiaire entre le niveau national et le niveau local dans l'application de la directive INSPIRE. Les IDG régionales sont en effet au centre de la production et diffusion des informations géographiques, et animent des réseaux thématiques locaux sous forme de groupes de travail appelés « pôles métiers » ce qui encourage l'adaptation de l'usage des données au territoire via la réutilisation ou la coproduction des données territoriales. Une partie de ces groupes traitent des questions d'organisation (gouvernance, plateforme,...) et l'autre de problématiques spécifiques (urbanisme, aménagement, agriculture,...).

Les IDG départementales sont au même niveau que les IDG régionales et conservent leur légitimité grâce aux relations privilégiées qu'elles entretiennent avec des acteurs locaux.

5.1.3. ForCity : une version privée de structure de gestion des données

Les porteurs de l'outil de modélisation ForCity le positionnent comme une plateforme territoriale de simulation de l'évolution du territoire permettant à différents acteurs de l'écosystème urbain d'y représenter leur métier et de faire des simulations de scénarios communs dans une démarche systémique et multi-thématique (mobilité, énergie, utilisation des sols, eau, déchets, bâtiments, etc.). ForCity s'appuie sur un modèle de données en open source, favorisant la substituabilité de l'outil pour ses utilisateurs et la reproductibilité des applications d'un territoire à l'autre. Le modèle économique de la société repose sur l'abonnement à une plateforme de simulation fournissant une aide à la décision en matière d'aménagement (impacts des choix à moyen-long terme).

La mise en place de plateformes d'intermédiation a été envisagée pour encourager le partage des données sur un territoire. Cette vision d'une structure « privée » de gestion de données se matérialiserait par des interfaces permettant de partager des données et de développer des axes pour le traitement de ces données partagées. François Grosse, président de la société ForCity, rapproche cela de l'idée d'un *marketplace* de la donnée territoriale qui hébergerait des données de toutes sortes sans dimension de simulation mais propres à générer des initiatives d'activités de B to B ou B to C à partir de ces données. Néanmoins, cette piste n'est plus envisagée dans les évolutions ultérieures de l'outil.

Organisation et gouvernance

Les modalités précises de la gouvernance qui sera mise en place autour de cette structure ne sont pour l'heure pas définies. La structure, matérialisée par l'outil ForCity, présentera les caractéristiques suivantes :

- > L'outil est basé sur une utilisation collaborative entre utilisateurs, notamment d'échelles territoriales différentes. Il s'intéresse en priorité à l'échelle de l'agglomération mais pourra être déployé à l'échelle régionale (voire nationale) dans le futur.
- > Pour sa construction, l'outil est développé par l'entreprise ForCity et une structure annexe de l'entreprise basée à Rennes pour la partie logicielle. ForCity a noué des partenariats académiques, notamment avec le Laboratoire Aménagement Economie Transports (LAET) à Lyon avec lequel ForCity co-développe l'outil de simulation urbanisme/transport de l'agglomération de Lyon pour l'intégrer dans sa plate-forme de simulation. Un autre partenariat est mis en place avec le laboratoire THÉMA pour l'intégration à la plateforme de l'outil de modélisation LUTI ("Land-use and transport interaction") MobiSim.
- > Les publics cibles de la plate-forme sont à la fois des villes ou des autorités publiques territoriales (syndicat intercommunal, département, région) et des entreprises (agences d'urbanisme, entreprises de l'énergie, des réseaux, de l'immobilier, de la construction, etc.), qui utiliseront l'outil et l'alimenteront également en données.
- > La plate-forme devrait permettre de développer de nouvelles formes de collaboration dans les territoires, entre les différentes échelles et types d'acteurs (entreprises, administrations publiques et autres acteurs publics). La plate-forme est en effet collaborative et permet de faire des scénarios sur une ville ou un territoire virtuel, adaptables en fonction du besoin et des thématiques intéressant chaque utilisateur et spécifiques aux territoires. Elle doit être accessible simultanément par ces différents utilisateurs et permet à chacun de partager une partie de leurs informations et résultats.

5.1.4. La régie de donnée selon le programme expérimental Datact

Selon l'analyse des porteurs du programme Datact, une volonté s'est développée depuis 2010 de dépasser une vision "historique" de l'open data pour penser le développement de services territoriaux par la coopération et le partage des données entre acteurs publics, privés et individus. L'idée est également d'amener les acteurs des collectivités à imaginer de nouvelles formes d'action publique et de mise à disposition d'outils basés sur le partage des données dans des nouveaux modèles de coopération. Le programme Datact vise à modéliser une Régie de données afin de faire qu'une économie de la donnée soit viable pour tous les acteurs parties prenantes de la production de services territoriaux. L'objectif affiché est de dépasser l'open data pour étudier les conditions d'acceptabilité du partage de la donnée pour les grands groupes privés, les PME/TPE et les individus, pour qu'émergent d'autres modèles économiques et de nouveaux moyens de développement territorial.

Entre le printemps 2013 et l'automne 2014, le programme a visé à élaborer une vision partagée et prospective des services basés sur le partage de données entre les parties prenantes et à établir un cahier des charges d'une régie de données multi-sources et multi-sectorielle. Ce dernier a nécessité de réfléchir au rôle à jouer pour les différents acteurs, aux modalités partenariales acceptables (droits et obligations) et aux modèles économique et de gouvernance de la régie.

A la suite de cette phase, le projet d'expérimentation de la régie de données sur le territoire de Plaine Commune a débuté en 2014 avec le soutien de l'Ile-de-France. Les services développés à partir de ces données viseront à accompagner l'implantation de grandes entreprises sur le territoire, avec le besoin de nouveaux services que cela implique. Aujourd'hui dans une phase de mise en place d'un démonstrateur pour tester une première version de la régie, le projet d'expérimentation se focalise en premier lieu sur la thématique de la mobilité. Celle-ci est centrale dans la mesure où l'offre de transports collectifs actuelle, dans l'attente de l'arrivée de la ligne 14 du métro, n'est pas en mesure de répondre aux besoins en termes de mobilité et d'accessibilité créés par l'implantation de nombreuses nouvelles entreprises.

Organisation et gouvernance

La structuration et la gouvernance de la régie de données développée par le programme Datact ne sont pas à l'heure actuelle entièrement définies. Le projet d'expérimentation de la régie sur le territoire de Plaine Commune vise notamment à déterminer les modes de gouvernance et de structuration optimaux pour un tel projet. Ainsi, la Région Ile-de-France intervient dans ce dossier en étroite collaboration avec Plaine Commune, le groupe Chronos et le Hub agence pour que ce laboratoire d'expérimentation devienne un modèle répliquable dans plusieurs territoires. Le choix d'une démarche partant de la ville vers la Région, en passant par la communauté d'agglomération et le département, vise également à déterminer comment développer ce type de modèle d'innovation par le partage des données. Il ne s'agit pas de déterminer l'échelle idéale mais plutôt les modalités de connexion et de coopération entre les différentes échelles.

Aujourd'hui, les acteurs parties prenantes du projet sont en recherche de la structure précise à donner à la régie, et le démonstrateur en cours de formalisation vise à servir de test à un élargissement à d'autres thématiques que la mobilité. Ainsi, des réflexions sont en cours sur :

- > le statut de la régie qui pourra être une régie opérée par Plaine Commune, une délégation de service public ou encore un partenariat public-privé. Le statut devra notamment prendre en compte d'autres enjeux liés aux acteurs privés et aux individus (secret statistique, loi de protection des données individuelles, etc.);
- > le mode d'accès aux données qui pourra être un système de licences en fonction de la taille et de la nature de la société en demande de données et en fonction du service public imaginé. L'accès aux données pourra également être gratuit dans le cas d'un service au bénéfice entièrement redistribué aux citoyens;
- > les niveaux de traitement des données diffusées, qui pourront être multiples (données brutes, données traitées, croisées, etc.);
- > la définition précise de l'acteur « tiers de confiance » de la régie, qui pourra être l'équivalent d'une chambre de compensation dans le système bancaire, un organe de type comité de pilotage public-privé, une délégation de ce rôle sous le contrôle d'un comité de pilotage, etc.

- > Les ressources financières, humaines et techniques nécessaires pour accompagner la régie, mettre en place des protocoles et s'assurer que les données soient fiables, documentées, remises dans des formes réutilisables, compatibles et anonymisées, via notamment la recherche de partenariats (SFR et Orange implantés sur le territoire, etc.).

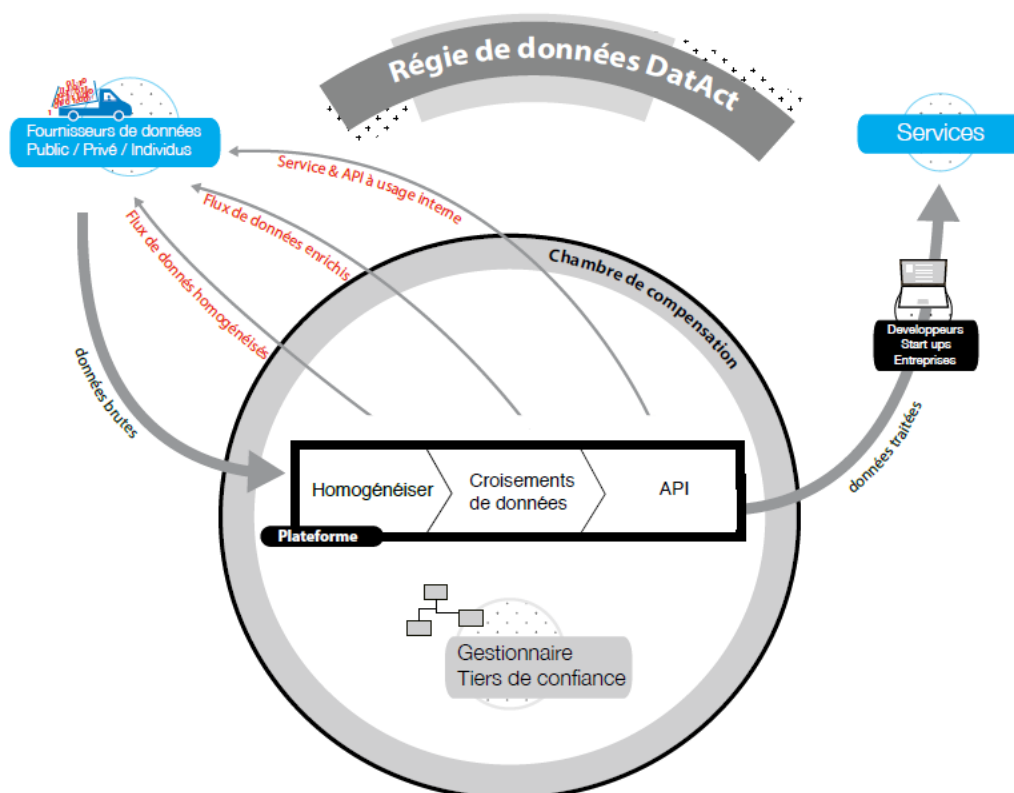


Schéma de la régie de données telle que conçue par Chronos et le Hub
(Source : La Gazette des communes, Cahier détaché n°2, 2014)

5.2. Missions et points de vigilance formulés à partir de l'étude de l'existant

Les missions des structures existantes de gestion de données varient en fonction des initiatives. Les missions listées ici ont été identifiées en particulier dans les initiatives de la RGD 73-74 et le projet de régie de données du programme DatAct, mais aussi pour certaines dans les initiatives plus spécifiques des IDG (centrées sur les données géographiques) et de la plateforme ForCity (structure privée).

Ces missions constituent une base pour améliorer la qualité, la documentation, le partage des données territoriales, pour faciliter leur traitement ultérieur, pour permettre leur utilisation pour différents usages voire leur visualisation directe. **Des points de vigilance** relatifs aux modes de gouvernance de telles structures ont d'ores et déjà été identifiés ; ils seront à compléter au fil des expérimentations faites en la matière, notamment suite à celle du projet de régie de données du programme DatAct sur le territoire de Plaine Commune.

5.2.1. Missions associées aux structures de gestion des données

1. Constituer un service territorial de base de données

Il s'agit de gérer un système informatisé de banques de données en vue de la diffusion de ces données à différentes catégories d'acteurs.

Ces acteurs sont définis pour chaque structure de gestion de données avec des variations dans le niveau d'ouverture aux acteurs privés : la RGD 73-74 est destinée aux structures publiques alors que le programme DatAct envisage la régie de données comme avant tout destinée aux entreprises.

La RGD 73-74 se donne ainsi pour mission de mutualiser toutes les acquisitions de données au bénéfice des acteurs publics, d'organiser et de structurer le développement de l'information en fournissant un cadre harmonisé sur le territoire des Pays de Savoie et d'accompagner les utilisateurs d'information géographique du territoire dans leur démarche de respect de la directive européenne INSPIRE.

Dans le cadre du projet DatAct, la régie a vocation à constituer un réseau d'acteurs divers autour des données du territoire, constitué de chercheurs, d'usagers, d'entreprises, etc. qui se connaissent, s'impliquent et se lient afin de participer au développement d'une culture de l'innovation nourrie par les données mises à disposition.

2. Etre une plateforme technique de partage, de traitement et de croisement des données

Le partage n'est pas un réel moteur pour la plupart des acteurs concernés. Cependant la fonction première d'un tel système peut être pour chaque acteur de normaliser ses données et de les configurer pour une possible interopérabilité avec celles d'autres acteurs qu'ils pourraient utiliser. En effet, les données proposées dans le cadre d'une structure de gestion des données doivent être obligatoirement documentées et qualifiées. De plus, ces données peuvent être brutes ou traitées, agrégées, hétérogènes, l'idéal étant de proposer différents niveaux de traitement des données afin de répondre aux besoins de tous types d'acteurs.

Dans le cadre de la Régie des Pays de Savoie, sont par exemple effectués des contrôles qualité et un suivi des mises à jour sur les données réglementaires comme le cadastre et les Plans Locaux d'Urbanisme.

De plus, ce type de structure de gestion de données a pour rôle de se porter garant des échanges et des transferts de données pour garantir à chacun des acteurs l'équilibre général des transactions.

Par ailleurs, la nature des données partagées varie en fonction des buts recherchés. La régie de données du programme DatAct, dans le cadre de son projet démonstrateur à Plaine Commune, n'a pas vocation à acheter des données au-delà des partenariats qu'elle cherche à mettre en place avec les acteurs du territoire. Au contraire, la régie de données des Pays de Savoie a mis en place un système d'achat et de péréquation dans le repartage des données via des abonnements. Elle cherche à élargir le champ des données mises à disposition et à continuellement enrichir le socle de ces données en s'adaptant aux demandes et contraintes de ses utilisateurs.

3. Faire office de tiers de confiance pour favoriser les relations et coopérations entre acteurs

En général, ce tiers de confiance a le rôle de garantir l'intégrité des apports en données, de leurs traitements et de leurs croisements. Il a également un rôle lié à la sécurisation des données (vérification, création de bases réutilisables, anonymisation...) et à la rémunération éventuelle de leurs fournisseurs.

Son statut juridique et ses missions précises restent néanmoins à définir et varient en fonction des initiatives.

Dans le cadre du projet démonstrateur de Plaine Commune (DatAct), le tiers de confiance est considéré comme un acteur partie prenante de la régie, agissant sur le traitement des données des parties prenantes, pour en améliorer la qualité et favoriser leur interopérabilité et assurant le rôle de sécurisation, notamment lié à l'anonymisation, des données sensibles. Dans la mesure où les données impliquées sont pour beaucoup des données individuelles voire sensibles -notamment de par la nature des services envisagés-, la régie doit constituer une garantie pour les usagers d'un usage éthique et sécurisé de leurs données personnelles et d'un accès à des services personnalisés.

Pour les porteurs de la régie de données des pays de Savoie, le statut de tiers de confiance de la régie désigne avant tout la reconnaissance de sa légitimité par les acteurs locaux.

4. Proposer des services autour des données

Ces services dépendent des utilisateurs bénéficiaires de la structure de gestion de données et de leurs besoins. Dans le cas de la RGD 73-74, ils sont destinés aux acteurs publics uniquement et permettent le traitement des données proposées. Parmi eux, un géoservice permet de visualiser les données territoriales, des services pour gérer les différents métiers de la commune (droits des sols, voirie, eau et assainissement, etc.), un service de représentation du terrain naturel en 3D, un service pour réaliser des cartes thématiques, etc.

5. Etre un générateur de nouveaux services développés par des acteurs privés

Une structure de gestion des données peut être un outil pour accompagner l'innovation en matière de nouveaux services en grâce à une structuration des données ouvertes qui constituent la matière première pour la création de ces services. Par exemple, la régie de données du programme DatAct est une plateforme générant des flux de données et destinée avant tout aux entreprises et entrepreneurs pouvant utiliser les données disponibles et périodiquement mises à jour pour améliorer un service existant ou en créer de nouveaux.

6. Enclencher une dynamique d'attractivité territoriale

Certaines structures peuvent également avoir pour ambition de connecter les acteurs territoriaux par la donnée, perçue comme un levier à la mise en relation d'acteurs pour le développement de services et, par-là, du territoire. Les nouveaux services visés peuvent par exemple concerner l'offre culturelle du territoire ou encore le développement d'une offre de mobilité alternative à celle des réseaux de transports collectifs (auto-partage en temps réel, etc.) et de services participant au lissage des arrivées et départs au travail. C'est notamment le cas de la régie selon le programme DatAct.

5.2.2. Points de vigilance associés aux structures de gestion de données

1. Des risques liés à une structuration de régie trop centralisée

Des modes de structuration trop fermés peuvent aller à l'encontre de l'objectif de mise à disposition des données, être contre-productifs ou avoir des impacts négatifs.

Par exemple, un modèle de régie obligeant ses adhérents (communes ou autres organismes publics en particulier) à lui fournir ses données sans les laisser libres de les diffuser par ailleurs et fournissant des services développés à partir de ces données (applications cartographiques métiers) :

- > peut conduire à fermer tout développement économique de la géomatique sur un territoire puisque les éditeurs n'y vendraient plus de logiciels (les collectivités passeraient exclusivement par les services de la régie);

- > limite de manière importante le partage des données des acteurs du territoire;
- > repose sur une logique de rendre des services payants allant à l'encontre de l'idée qu'on puisse diffuser des données pour que d'autres puissent développer des services librement.

2. Proposer une structure de gestion des données peut nécessiter des moyens techniques et financiers importants

Les différentes fonctions assurées par une « structure de gestion des données » décrites plus haut exigent, pour être menées à bien, des ressources financières et humaines significatives pour un territoire.

Les acteurs qui mettent en œuvre une structure de gestion des données doivent ainsi anticiper les partenariats à développer et les expertises à mobiliser. En particulier, le rôle et le statut du tiers de confiance est à définir de manière rigoureuse. Ce dernier doit présenter une expertise technique poussée dans la mesure où il sera en charge d'effectuer l'anonymisation des données en accord avec la loi et d'incarner une garantie de sécurité pour les fournisseurs de données, en particulier les individus et entreprises qui acceptent de diffuser leurs données au sein d'un périmètre précis.

5.3. Quels modèles économiques pour les structures de gestion des données?

La mise en place d'une structure de gestion de données territoriales constitue un projet souvent inédit pour les acteurs locaux (producteurs de données, collectivités...). Ce projet vise à répondre à des besoins nouveaux en matière d'aide à la décision dans le domaine de l'aménagement et la gestion durable du territoire, dans un contexte de croissance très forte de données mobilisables. Si l'utilité des structures de gestion de données est reconnue, la difficulté de trouver un modèle d'affaires pertinent (en particulier dans le domaine de l'environnement) freine leur développement. Hors financement public, ces systèmes ne rencontrent pas aujourd'hui une demande leur permettant de financer leur fonctionnement. La présente partie s'attache à dessiner les contours d'un possible modèle économique pour la gestion de données, selon la valeur représentée par la donnée et à partir d'analogies à des modèles existants.

5.3.1. La valeur attribuée à la donnée : un fondement du modèle économique

Il apparaît possible d'attribuer une valeur à la donnée, en fonction de la manière dont la donnée sera considérée, en lien avec l'approche en cycle de vie. Si le modèle économique ne se résumera pas à la façon d'attribuer de la valeur à la donnée, il sera en revanche fortement influencé par celle-ci. On se rapproche ici des grandes étapes du cycle de vie de la donnée présentées dans la partie précédente, où la donnée va être collectée (la valeur sera associée au stock de données), va circuler (valeur associée au flux) et enfin être traitée (valeur associée à l'usage ou au service associé à la donnée elle-même).

1. La valeur de la donnée comprise comme un stock

Une structure de gestion des données peut fonctionner comme une organisation centralisatrice et redistributrice des données comprises comme un stock et dont la valeur est lissée par le rôle centralisateur de la structure. Elle se positionne alors comme un acteur central intermédiaire qui achète des données de manière centralisée, parfois via des achats groupés, auprès de sources diverses, et les rend ensuite accessibles à ses abonnés de manière plus ou moins ouverte selon l'abonnement contracté ou selon le type d'acteur intéressé. On se situe généralement ici sur un maillon relativement « amont » de la chaîne de valeur de production de la donnée : il s'agit généralement de marchés en « BtoB » où les clients ou cibles sont eux-mêmes de producteurs de services autour de la donnée et doivent acheter cette dernière, comme un produit brut.

Les prix des données peuvent ainsi être les mêmes quels que soient les sets de données téléchargés par les abonnés et quel que soit le prix initial payé par la structure de gestion de données pour l'obtention de ces données. Le prix de l'abonnement est lui souvent corrélé à la taille de la collectivité abonnée, ou de l'entreprise intéressée par un set de données : de grosses entreprises à moyens importants peuvent avoir un coût d'accès aux données plus important qu'une startup dont la régie peut viser à favoriser le développement. Dans ce type de modèle économique, la plus-value apportée par la régie réside dans le contrôle de la donnée diffusée et éventuellement dans son traitement (qualification, métadonnées, etc.).

Ce type de modèle peut se décliner ainsi selon différents types d'accès à des données centralisées par une structure de gestion de données :

- > Une mise à disposition gratuite des données, de manière différenciée ou non selon les acteurs ;
- > Un abonnement mensuel ou annuel offrant un accès illimité à tout ou partie des données mises à disposition ;
- > L'achat d'un set de données à l'unité avec un coût variable selon l'acteur considéré.

La régie de données des Pays de Savoie fonctionne selon un modèle économique de ce type : elle réalise des achats de données en consolidant progressivement la base de données disponibles et propose un système d'abonnement permettant de lisser les coûts de récupération des données.

Le contrôle de qualité des données est en outre important dans les IDG qui achètent des données à l'IGN et doivent s'assurer des caractéristiques de ces dernières (en particulier leur robustesse, origine...). Cela demande des compétences techniques pour contrôler la donnée et faire remonter les problèmes au niveau de l'IGN, compétent en la matière. L'encadré page suivante fournit un exemple d'un modèle économique d'IDG départementale.

2. La valeur de la donnée comprise comme un flux

Un modèle économique considérant la donnée comme un flux consiste par exemple à proposer des plateformes qui connectent entre eux des systèmes existants, à rendre interopérables les données et les outils techniques pour les traiter. Il s'agit souvent des plateformes financées dans le cadre de projets état-région.

Ainsi, le programme Dataact conçoit la régie de données comme une « chambre de compensation » de la donnée qui se porte garante des échanges et transferts de données qui se font selon des flux continus avec un mouvement vers un équilibre des transactions. L'interopérabilité est également au centre de cette vision afin de favoriser et fluidifier les échanges.

De la même manière, le concept de *market place* de la donnée territoriale consiste à proposer un espace hébergeant des données de toutes sortes, alimenté par divers acteurs et accessible de manière très ouverte. Ce système est également générateur d'un flux de données continu en vue de stimuler des initiatives d'activité ou de création de services.

3. La valeur de la donnée comprise comme un usage

Le modèle économique d'une structure de gestion des données peut également considérer la valeur des données à partir de l'usage qui en est fait ou du service créé à partir de ces dernières. La cible est alors un utilisateur final du service, qui peut être un particulier (ex. : application Waze, informations produites dans le cadre de la mise en place des compteurs communicants Linky ou Gazpar...) ou une organisation. Ce type de service est basé sur la fourniture d'une information de nature à produire un changement de comportement : modification de parcours routier, écogestes en matière de consommation d'énergie...

Ce type de modèle apporte une plus-value aux données comprises comme un stock ou un flux, et propose en plus des services « à la carte » qui constituent l'essentiel de ses revenus. Ces services s'appuient généralement sur des modèles économiques spécifiques, qui peuvent eux-mêmes s'adapter aux différents utilisateurs cibles en proposant une personnalisation du service.

5.3.2. Analogie avec des modèles économiques existants

La gestion des données territoriales n'est pas assez mature pour avoir trouvé son (ses) mode(s) de développement économique. Cependant, le modèle économique étant une condition de réussite et de pérennité de cette activité, il est intéressant d'explorer des modèles d'activités connexes à la gestion des données ou présentant des caractéristiques communes, qui peuvent être dans chacun des cas basés sur la donnée comme un stock, un flux ou un usage.

1. Modèles économiques de logiciels

Le domaine du logiciel offre une voie pour approcher les différents modèles d'affaires possibles pour la gestion des données, cette dernière présentant des enjeux liés à ceux des outils informatiques. Le logiciel voit coexister aujourd'hui deux modèles « historiques » :

- > Le premier correspond à une souscription ponctuelle (« one-time-fee ») associée à la vente d'un produit (logiciel) : il y a une concession de nature perpétuelle d'une licence de logiciel entre le titulaire du droit d'auteur et l'utilisateur du logiciel. L'exemple classique est celui de la suite Office de Windows.
- > Le deuxième répond à des outils nécessitant la mise à jour régulière de données : le produit est alors vendu en récence, commercialisée sous la forme d'une licence mensuelle ou trimestrielle, avec mises à jour régulières (ex. : logiciels d'analyse de cycle de vie).

Dans les deux cas, pour une structure de gestion de données, cela correspond à l'obtention par le client (via une cotisation, un achat ponctuel) d'un set de données. L'encadré ci-dessous illustre ce type de fonctionnement.

Exemple de modèle économique d'IDG départementale²⁸

Certains dispositifs sont financés à 100% par des subventions (collectivités, Etat) et proposent des services gratuits à leur membre (accès sécurisé) ou à tous (accès ouvert).

D'autres plateformes développent des prestations sur des projets particuliers (par exemple la mise en place de systèmes d'information auprès de différents partenaires). L'infrastructure mutualisée de données est alors financée par ces partenaires (qui l'alimentent et s'y alimentent en données).

Enfin, d'autres systèmes proposent des formules payantes. Deux modèles types de tarification sont mis en place dans les départements qui ont mis en œuvre une infrastructure de données géographiques : une tarification « à la parcelle » répondant bien souvent à un objectif de numérisation du plan cadastral (cas des Landes) ou une tarification « au nombre d'habitants » (cas du Jura, de la Savoie...). Ces modes de tarification sont souvent complétés par la mise en place d'abonnement « à la carte » à des GéoServices métier et pondérés pour le groupement de type EPCI, départements, SDIS, SDE, etc.

Le tableau ci-dessous présente quelques exemples de tarifs pour les Départements de la Nièvre et du Jura. Le calcul a été effectué sur les GéoServices de base de chacune de ces plates-formes.

	ADACL 40 Landes	SIEEN (Nièvre)	GéoJura Jura
Commune de 150 habitants	Dépend du nb de parcelles.	DE : 170 € AA : 60 €	DE : 324 € AA : 91 €
Commune de 550 habitants	Dépend du nb de parcelles.	DE : 270 € AA : 150 €	DE : 857 € AA : 223 €
Commune de 3500 habitants	Dépend du nb de parcelles.	DE : 600 € AA : 400 €	DE : 4780 € AA : 1190 €
Com. Com. de 10.000 habitants	Dépend du nb de parcelles.	DE : 600 € AA : 400 €	DE : 5575 € AA : 1692 €

DE : Droit d'Entrée – AA : Abonnement Annuel

En complément de ce tableau qui fournit des indications sur des GéoServices de base, certaines plates-formes proposent des GéoServices métier dédiés à des utilisations spécifiques. C'est le cas de Géo Vendée qui propose un outil pour la gestion technique et financière des phases de travaux de voirie. Il permet la saisie et le suivi des travaux de voirie (travaux neufs et d'entretien), année par année, et les restitutions statistiques et cartographiques avec une gestion intégrée de l'historique et des photographies. L'utilisation de ce service est proposée sous forme d'un abonnement annuel de 150 € TTC par commune. Cet abonnement est parfois pris en charge par les intercommunalités pour l'ensemble des communes.

On distingue souvent :

- > Un droit d'entrée (payé une seule fois) et qui inclus l'initialisation et le paramétrage des comptes utilisateurs, les formations et l'assistance à la prise en main des outils.
- > Un abonnement annuel qui dépend du nombre et du type de GéoServices utilisés.

La création d'une formule « acteur-relais » est parfois envisagée sur le modèle de ce qui se pratique en Savoie : un EPCI qui s'abonne pour ses services et pour l'ensemble de ses communes se voit attribuer une réduction de 20% du montant global d'abonnement. Le risque est de faire payer plusieurs fois un même territoire à travers la commune, l'intercommunalité et le syndicat mixte ; certains territoires de Savoie payent ainsi les GéoServices de la RGD plus cher que s'ils s'équipaient directement.

²⁸ Source : Matthieu Noucher

De nouveaux modèles sont apparus :

- > « Cloud computing » : ce modèle (qui n'est pas exclusif des modèles précédents) repose sur le paiement d'un abonnement uniquement à l'utilisation, sans que l'outil ne soit lui-même installé sur le site de l'utilisateur (ex. : Dropbox).
Dans ce cas, pour une structure de gestion de données, à l'obtention par le client (par une cotisation, par exemple) d'un droit d'accès aux informations produites par les données ou à des outils basés sur les données (de type logiciel SAAS).
- > « Open source » : ces logiciels (ex. : Linux), comme leur nom l'indique présentent un code libre de diffusion ou dont la source est disponible gratuitement. La création de valeur économique pour l'éditeur du logiciel se fait donc ailleurs, à travers des services complémentaires (informations telles que la documentation, support d'utilisation...) contractualisés.
Dans ce cas, pour une structure de gestion de données, cela correspond pour le client/utilisateur à un accès gratuit aux données, et à une utilisation payante de services basés sur ces données (indicateurs, outils). Les utilisateurs peuvent contribuer à l'amélioration et la précision des données.

Le domaine du logiciel ne suffit cependant pas à illustrer les possibilités en termes de modèles économiques qui s'offrent à l'activité de gestion de données. Le rôle d'une structure de gestion de données est en effet multiformes (cf. partie 5.2.1.) et dépasse le cadre d'une offre de produit/service. Il faut en particulier envisager les possibilités de relations autres que client-fournisseur pour intégrer la multiplicité des acteurs et la diversité des rôles que chacun peut jouer.

2. Modèles économiques émergents en lien avec l'utilisation des TIC

Le numérique fait émerger de nouveaux modèles économiques, en offrant un potentiel de création de valeur et une diversité de modalités de diffusion d'un service²⁹.

Certains modèles reposent sur une « déconstruction » de la relation classique « fournisseur-client » en plaçant l'entreprise dans un schéma d'interactions avec différents types de clients (modèles bifaces). C'est le cas des services de type « plateforme » pouvant à la fois toucher des bénéficiaires finaux avec un service gratuit (ex. : information commerciale sur un trajet donné) et des entreprises intéressées pour diffuser leur information commerciale à une population donnée. La valeur de ces services dépend bien entendu du nombre d'utilisateurs et partant, de la capacité de la plateforme à les toucher.

D'autres modèles se fondent sur une différenciation d'accès (à une base de données) ou de qualité (de service) pour le consommateur final, dont l'accès de « premier niveau » est gratuit et devient payant à mesure que la donnée ou le service fournis intègre de la valeur ajoutée. Ce type de modèle (« Freemium ») permet ainsi de faire payer la gestion d'un service dans son intégralité par ses utilisateurs les plus assidus / exigeants.

Appliqués à l'activité de gestion des données territoriales, ces différents modèles offrent des pistes de réflexion intéressantes : peut-on distinguer des niveaux d'accès différents selon les acteurs ou clients finaux ? L'accès aux données relatives à des profils de consommateurs de la donnée représente-t-il une valeur en soi ?

3. Modèles économiques intégrant une dimension collaborative

Les modèles économiques associés à l'économie collaborative offrent un autre regard sur la gestion des données, par certaines caractéristiques partagées, par exemple : existence d'une multiplicité de contributeurs, accord pour partager certaines informations ou services, relations d'échanges / de coopération nécessitant une confiance vis-à-vis du porteur du service bien entendu, mais également entre utilisateurs de celui-ci³⁰.

L'économie du partage est en effet basée sur le partage d'actifs sous-utilisés comme des espaces ou des compétences ; elle peut générer des bénéfices monétaires ou « non captés » (bénéfices sociaux, gain de temps...).

²⁹ Paragraphe « Modèles économiques émergents en lien avec l'utilisation des TIC » issu d'une étude de l'ADEME finalisée en 2015. Rapport d'étude : L'approche économique des projets d'aménagement : enjeux, mécanismes et acteurs

³⁰ On retrouve ici l'idée du tiers de confiance exposée dans le cadre de l'activité de gestion des données.

Elle s'est développée en particulier grâce aux outils numériques et un fonctionnement collaboratif basé sur les contributions de chacun pour améliorer une connaissance, un service, un produit (c'est le cas de certains logiciels). Les structures de gestion de données pourraient présenter un fonctionnement similaire au sens où les contributeurs sont multiples et où le succès de la structure repose sur le nombre et la qualité des informations collectées.

L'économie du partage implique généralement une organisation moins pyramidale et plus participative : un parallèle peut être établi avec la gestion des données dont le fonctionnement implique une relation entre la structure de gestion et les producteurs et consommateurs de données très participative.

Sur le plan économique, on constate que certaines applications au départ gratuites sont devenues payantes (et ont repris ainsi un fonctionnement plus classique). L'accès aux données pourrait l'être jusqu'à ce que l'activité des utilisateurs soit stabilisée.

Un potentiel économique à valoriser ?

L'activité de gestion de données territoriales de l'environnement présente un potentiel de développement (économique) encore très faiblement exploité.

Plusieurs raisons peuvent expliquer ce constat :

- > Faiblesse de l'expression du besoin :
 - Du côté des collectivités : la prise en compte de l'environnement par les collectivités dans les politiques et projets est bien souvent insuffisante au regard des enjeux (voire parfois des obligations réglementaires) ; une prise en compte rigoureuse exigerait des données plus nombreuses et de meilleure qualité. Dans ce sens, les collectivités seraient amenées à payer un service de gestion des données territoriales à sa juste valeur. Le cas de ForCity, dont le modèle repose sur une capacité à payer des collectivités pour un service de « prédiction » du fonctionnement de la ville autour de services nouveaux, constitue aujourd'hui une exception. De plus, le recours à la modélisation, qui nécessite de nombreuses données et confirme l'importance de leur gestion, n'est pas systématique pour orienter les décisions.
 - Du côté des individus : l'intérêt des données environnementales pour le grand public reste encore très faible malgré la variété des informations portées (ex. : qualité de l'air, zones d'intérêt écologique, niveau de consommation d'un logement...). Il est probable que le « consentement » à payer pour une information environnementale soit associé à la qualité du service associé à la donnée elle-même.
- > Faiblesse des services aujourd'hui associés à ces données : si la gestion des données ne constitue pas une activité facilement « monétarisable » aujourd'hui, il est probable que son modèle économique décolle via les outils et services associés aux données elles-mêmes³¹. La partie suivante permet à ce titre d'explorer quelques pistes.

Ainsi, au-delà du caractère innovant des modes de collecte et de traitement de la donnée, ou encore des outils et services associés à celle-ci, la création de modèles économiques constitue en soi un champ d'investigation en matière d'innovation. Sans exploration spécifique de ce champ (par les promoteurs d'outils, leurs bénéficiaires...), il apparaît peu probable de voir la dynamique relative à la gestion des données environnementales pour l'aide à la décision locale se renforcer à l'avenir.

³¹ Le caractère innovant de ces outils ou services sera à cet égard déterminant, associé au caractère innovant des modèles économiques eux-mêmes (nouvelle création de valeur, nouvelle relation producteur du service / bénéficiaire du service).

5.4. Organisation des collectivités et des acteurs territoriaux pour la gestion des données

5.4.1. Evolution de la place du numérique et de la gestion des données dans l'organisation des collectivités

1. Organisation interne des collectivités

Pour les collectivités, la prise en compte des enjeux du numérique et le développement de projets y ayant recours nécessite une organisation en interne spécifique et adaptée aux enjeux de la gestion des données territoriales et de la création de nouveaux services. Encore récemment, la configuration dominante au sein des collectivités pour traiter ces sujets était un service dédié à la gestion des données et au numérique. Ce rôle était souvent incarné par la direction des systèmes d'information (ou des services informatiques) qui coordonnait l'utilisation des outils numériques dans des projets spécifiquement du domaine du numérique. L'approche qui se développe progressivement est celle d'une approche intégrée et transversale, dans laquelle les différents directeurs de services d'une collectivité sont sensibilisés et formés à l'utilisation du numérique dans leur domaine. Cette organisation vise une prise en compte systématique du numérique dès la phase de conception des projets sectoriels.

Les nouveaux outils numériques et les données massives désormais accessibles permettent de mieux définir une politique publique et d'en améliorer le pilotage et l'efficacité. De plus, le numérique fait évoluer l'organisation des systèmes sociaux. Les services de la ville œuvrent avec des citoyens qui passent de consommateurs à acteurs des services urbains.

Concernant plus particulièrement le traitement de l'open data dans les collectivités, il est nécessaire que les différents services d'une collectivité s'approprient cette thématique afin qu'elle soit traitée de manière transversale dans les projets. La création de services via des projets initiés par la collectivité elle-même ou proposé par des acteurs privés nécessite souvent le croisement de données issues de nombreuses sources. Cela implique que les acteurs publics, privés et les citoyens soient mobilisés sur le partage de ces données et engagés dans une coopération pour développer de nouveaux services propres à participer à construire la ville.

A Plaine Commune, un centre de prospective territoriale dédié à ces questions a été créé. Il s'agit d'un organisme externe dont la vocation est d'aider les élus à se positionner et à décliner leur prospective en matière de politique numérique territoriale. La diffusion dans les différents services de la politique numérique de la collectivité se fait donc en aval de la définition d'un véritable projet de territoire en la matière. Le centre de prospective mène également une réflexion avec le philosophe Bernard Stiegler sur le rôle du numérique dans la transformation des métiers et de l'action publique.

2. Moyens mis à disposition par les collectivités pour la révolution numérique³²

La ville connectée repose sur les apports d'informations, de données ou de connaissances, provenant des responsables politiques comme des citoyens. Les collectivités territoriales sont sollicitées aujourd'hui par de nombreux acteurs qui souhaitent mettre en place des nouveaux services sur le territoire : entreprises du numérique, acteurs de l'énergie, entrepreneurs, acteurs du domaine de la santé, du domaine des transports, ou citoyens eux-mêmes.

La révolution numérique en cours nécessite une évolution de l'image du numérique par les parties prenantes. Cette évolution a été lente à se mettre en place en France, comme le montre l'utilisation encore très faible des nouveaux services numériques dans le domaine de la santé, comme le DMP (Dossier Médical Personnalisé) par exemple. Néanmoins, un éveil des régulateurs est en cours, ainsi qu'un retournement notable en France de l'image du numérique par les différentes parties prenantes. Cela se manifeste par une évolution des compétences chez certains types

³² Sources : Conférence « Big data et objets connectés, révolution technologique et anthropologique » du 9 juin 2015 par le Club Mines Informatique ; Conférence « La ville intelligente est-elle durable ? Décryptage à partir des enjeux de données » du 17 avril par le Club Ville IDDRI.

La révolution numérique désigne le bouleversement en profondeur des sociétés en particulier des Etats industrialisés provoqué par l'essor des techniques numériques (informatique et internet), qui se traduit par une mise en réseau généralisées des individus, de nouvelles formes de communication et une décentralisation dans la circulation des idées. (Source : Laurent Sorbier, « Quand la révolution numérique n'est plus virtuelle », Esprit, mai 2006). Concrètement, elle désigne l'introduction massive de la technologie numérique dans tous les domaines et moments de la vie quotidienne, de l'économie et de l'administration.

d'acteurs. Les entreprises du secteur de l'énergie par exemple, de culture ingénieur, recrutent des informaticiens, tandis que les entreprises du secteur de l'informatique recrutent des ingénieurs pouvant leur apporter des compétences sur les domaines d'application du numérique au sein de collectivités territoriales (énergie, transports, bâtiment, etc.).

Cette évolution de la perception du numérique s'accompagne d'une mise en place de moyens visant à favoriser l'innovation de rupture. Cela peut prendre la forme de pépinières ou d'incubateurs destinés à accompagner l'émergence de projets et la création d'entreprises, ou de cellules transverses directement intégrées au sein de grands groupes afin de développer rapidement et de façon opérationnelle des nouvelles idées innovantes. La méthode agile, dont l'objectif est le développement d'innovations de rupture grâce à des phases d'expérimentation nombreuses et fréquentes, se caractérise par une innovation plus forte, une interaction plus élevée avec les utilisateurs, et une meilleure capacité à respecter les délais. Les nombreux incubateurs mis en place en France participent à cette dynamique, comme par exemple la Halle Freyssinet en cours de rénovation et qui accueillera en 2016 le plus grand incubateur du monde.

En réaction au développement de services et d'outils automatisés et dématérialisés, la problématique de la place de l'homme comme acteur-décideur prend de plus en plus d'ampleur. Le *machine learning* se développe grâce à des outils qui vont répondre de manière automatique à des situations particulières, et l'enjeu est aujourd'hui de savoir jusqu'où il sera acceptable et pertinent de suivre les solutions proposées par ces outils et s'il reste de la place pour la décision individuelle et politique.

5.4.2. Les acteurs et rôles dans la gestion des données

L'évolution des pratiques en matière de gestion des données territoriales, et en particulier de l'utilisation accrue de ces données par tous types d'acteurs entraîne une reconfiguration des rôles des différents acteurs des territoires ainsi que l'émergence de rôles nouveaux.

1. Le rôle émergent de l'administrateur de données

Le concept d'administrateur de données (ADD) - ou CDO : *chief data officer* - a récemment émergé, à la fois au sein des entreprises, en particulier américaines, mais aussi dans les collectivités et au niveau de l'Etat.

Au sein d'une entreprise ou d'une collectivité, l'administrateur de données est responsable de l'état et de la gouvernance des données. Son rôle est d'avoir la connaissance des services et de leur activité ainsi qu'une vision transversale et une connaissance de la stratégie de l'organisation générale. Il doit également avoir des compétences techniques sur les méthodes et les outils et des capacités d'animation. Il peut avoir différents profils : "fédérateur et animateur", "gardien de la sémantique", "normalisateur" ou "gardien de la cohérence".

Il est possible de distinguer plusieurs types d'administrateurs de données :

- > L'ADD documentaliste (mission courante de l'ADD) ;
- > L'ADD « qualité des données » : plutôt focalisé sur le contrôle qualité, l'amélioration et le suivi de la qualité des bases de données ;
- > L'ADD animateur, plutôt focalisé sur l'assistance aux utilisateurs dans l'amélioration de l'usage de leurs outils informatiques ;
- > L'ADD diffuseur, focalisé sur les partenariats et les échanges de données ;
- > L'ADD méthode : plutôt focalisé sur l'assistance aux concepteurs et réalisateurs de systèmes informatiques ;
- > L'ADD normalisateur : plutôt focalisé sur la publication de normes et de référentiels internes.

Au sein des collectivités, le rôle d'administrateur de données peut être exercé par le directeur des systèmes d'information (DSI), dans une phase de développement de l'attention portée aux données et à leur utilisation par les collectivités.

A l'échelle d'un Etat, la notion d'Administrateur Général des Données se développe également. La France s'est dotée en septembre 2014 d'un administrateur général des données (AGD) en la personne d'Henri Verdier, directeur d'Etalab. Ce dernier est placé sous l'autorité du Premier ministre au sein du secrétariat général pour la modernisation de l'action publique (SGMAP). Il s'agit de la première fois en Europe qu'une telle fonction est créée au niveau national. L'AGD est

autorisé à connaître les données détenues par l'administration de l'Etat et ses opérateurs. Ses missions sont les suivantes :

- > Bâtir progressivement une véritable gouvernance de la donnée au sein de l'Etat en assurant l'homogénéité, la bonne circulation et la pleine utilisation des données produites par tous services et leur circulation dans l'économie (exploitation et réutilisation par les chercheurs ou les entreprises) dans le respect de la vie privée et des cadres juridiques ;
- > Veiller à la production et à l'acquisition de données essentielles ;
- > Permettre d'améliorer, en coopération avec la DISIC et les administrations, les données produites, leurs formats et leur interopérabilité, avec pour objectif d'en faciliter l'agrégation, la circulation et le traitement ;
- > Stimuler l'appropriation par l'Etat des méthodes de l'analyse prédictive et des stratégies fondées sur la donnée, lancer des expérimentations pour éclairer la décision publique et diffuser les outils, méthodes et une culture de la donnée au sein des administrations et au service de leurs objectifs respectifs.

2. Les acteurs territoriaux clés

A l'échelle territoriale, on assiste également à une reconfiguration du rôle des différents acteurs et de leurs modes d'interaction.

Premièrement, le rôle des pôles de compétitivité est central dans la mise en œuvre de projets d'expérimentation notamment liés au numérique et apportant des réponses aux enjeux environnementaux. Ils se présentent comme des acteurs médiateurs et facilitateurs de l'intégration des nouvelles possibilités du numérique et participent à la stimulation et diffusion de l'innovation. Il s'agit de permettre l'émergence effective de projets portés par des entreprises et des acteurs de la recherche dans une démarche de coopération.

En Ile-de-France par exemple, trois pôles de compétitivités portent sur le numérique au service du territoire et sont des acteurs clés de la mise en œuvre de plus de 400 projets innovants :

- > Le pôle Advancity (ville et mobilité durables) qui cible le développement durable de la ville, de l'habitat et de la mobilité en milieu urbain ;
- > Le pôle System@tic Paris-Région qui concentre son activité sur la révolution numérique au service de six marchés applicatifs, dont notamment les transports, la confiance numérique & sécurité ou la gestion intelligente de l'énergie ;
- > Le pôle Cap Digital qui vise à positionner la région parmi les références mondiales du numérique d'un point de vue industriel et stratégique.

Les pôles de compétitivité sont des écosystèmes construits à partir de la réalité du territoire (ils sont construits par les acteurs territoriaux eux-mêmes) avec la caractéristique de regrouper et de favoriser les coopérations entre des acteurs proches géographiquement et ayant un intérêt ou thème technologique commun. La démarche est appuyée par des moyens notamment financiers afin de soutenir l'innovation collaborative sous la forme de projets de recherche associant des industriels et des laboratoires publics. L'objectif est de créer des synergies et de l'innovation en associant tous les acteurs du territoire à la démarche, générer des projets et produits créateurs de richesse et d'emploi.

Deuxièmement, le rôle des collectivités se reconfigure, afin de répondre aux besoins et attentes des citoyens. Cela peut notamment se traduire par l'achat de nouveaux services aux entreprises propres à servir le territoire et à renvoyer aux citoyens le résultat du traitement des données produites en partie par eux. Les collectivités ont également un rôle de tiers de confiance dans la gestion des données territoriales, qui consiste notamment à mettre en relation non plus seulement des acteurs publics avec des acteurs privés mais aussi des acteurs privés entre eux. Il peut s'agir notamment d'assurer la sécurisation de leurs données et leur utilisation restreinte dans le cadre de projets particuliers. Ce tiers de confiance encourage le développement de nouveaux services.

Enfin, il est important de souligner que ce sont les élus qui font les choix technologiques sur les territoires, et qui déterminent la direction prise par le développement de ces nouveaux services. Des acteurs et structures sont là pour les accompagner : les pôles de compétitivité fournissent un accompagnement sur l'analyse, les industriels ont des visions plus transversales et ouvertes sur les technologies existantes et à venir et les experts rendent des avis différenciés qui nourrissent les débats.

3. La montée en puissance de l'individu acteur

Comme présenté plus haut, de nombreux services développés aujourd'hui grâce au big data font participer les citoyens au développement urbain. Des données d'utilisation des services urbains peuvent être récoltées de façon passive, mais chaque citoyen peut également participer de façon active en faisant remonter ses besoins et attentes via des applications particulières.

Le citoyen est-il contributeur malgré lui ? Est-il obligé de contribuer pour accéder au bénéfice ? Cette question est fréquemment posée, et la distinction des contributions citoyennes passives et actives permet d'y apporter une réponse. En effet, les individus connectés à internet contribuent un minimum, mais de façon passive et involontaire : une simple requête dans un moteur de recherche enrichit la base de données de ce moteur de recherche et contribue donc à en alimenter l'intelligence.

En plus de ces données caractérisant l'utilisation des services urbains (transports par exemple) ou services numériques (moteurs de recherche), l'individu peut également partager de façon passive ou active certaines de ses données personnelles nécessitant d'être protégées (position, centres d'intérêt, etc.), venant alimenter des flux de données nombreuses et en temps réel (big data). De nouveaux services proposant des bénéfices aux consommateurs doivent pouvoir être créés tout en protégeant les données sensibles de ces derniers. Les pays européens sont très attentifs à cette question de protection des services et des besoins : un règlement européen sur la protection des données personnelles devrait être publié à l'automne 2015 à ce sujet.

Il est pour finir important de noter qu'en plus de la contribution individuelle de chaque individu, la multitude de contributeurs est un acteur à part entière, capable de créer à la fois des données en quantité, de qualité, ainsi qu'un cadre de confiance. L'exemple de l'encyclopédie en ligne Wikipédia est particulièrement parlant, avec une version anglaise qui présente un pourcentage d'erreur moins important que l'encyclopédie Britannica, qui serait équivalente à une pile de livres de 120m de haut une fois imprimée, et que de nombreux citoyens consultent avec un degré de confiance assez élevé du fait que les contenus sont revus et complétés par les utilisateurs.

6. Pistes de réflexion pour la gestion des données territoriales

La proposition de pistes de réflexion pour la gestion des données territoriales vient répondre à une finalité générale qui est également celle de l'étude « Approche intégrée et multicritère de la modélisation territoriale » : **améliorer la gestion des données territoriales pour répondre aux besoins en termes d'aide à la décision des collectivités et des acteurs locaux dans le domaine de l'environnement**. Il s'agit dans cette partie de proposer des **pistes pour élargir la réflexion**.

Les premières phases de l'étude ont fait émerger dix problématiques clés entourant la gestion des données territoriales. Un exercice de priorisation des enjeux a permis de dégager trois **problématiques transversales**, à partir desquelles un travail a été mené en co-production avec l'ADEME pour définir les suites à donner à cette étude. Deux **buts**, présentés dans le tableau ci-dessous, ont été formulés dans un souci de répondre via deux approches complémentaires à chacune des trois problématiques. Le premier concerne le cadre (juridique, économique, technique et institutionnel) et le deuxième concerne les acteurs.

Finalité	Problématiques identifiées	Buts
Améliorer la gestion des données territoriales pour répondre aux besoins en terme d'aide à la décision des collectivités / des acteurs locaux dans le domaine de l'environnement	Manque de structuration des données et d'harmonisation des formats limitant l'interopérabilité des outils	Promouvoir un cadre pour assurer la qualité des données et leur valeur ajoutée potentielle
	Incertitude sur le(s) modèle(s) économique(s) pour la valorisation des données freinant les projets de développement d'outils	Faire émerger le dialogue et les projets entre les parties prenantes de la collecte, du traitement et de l'usage des données
	Absence de modèle de gouvernance « de référence » pour partager les données, qui limite l'initiative des acteurs	

Chaque but a ensuite été décliné en objectifs stratégiques pour approfondir la réflexion et les projets autour de la gestion des données territoriales.

But		Objectifs stratégiques	
1	Promouvoir un cadre pour assurer la qualité des données et leur valeur ajoutée potentielle	1.1	Contribuer à l'évolution du cadre réglementaire et juridique pour s'adapter aux nouvelles opportunités et risques liés aux données
		1.2	Capitaliser sur les nouveaux modèles de gouvernance et d'affaires pour proposer un panel de solutions, sensibiliser les élus et « acteurs de la donnée » à l'importance de la gouvernance des données
		1.3	Travailler à la convergence des formats de collecte des données
But		Objectifs stratégiques	
2	Faire émerger le dialogue et les projets entre les parties prenantes de la collecte, du traitement et de l'usage des données	2.1	Favoriser le dialogue et la compréhension entre les acteurs à l'échelle nationale (publics, privés, recherche, citoyens)
		2.2	Renforcer la compréhension des enjeux de l'usage des données et l'émergence de projets au niveau local
		2.3	S'appuyer sur et participer au développement des initiatives locales en matière de gouvernance des données territoriales, pour adapter les modèles de gouvernance aux différents besoins et niveaux d'acteurs et aux modalités de dialogue

Ces objectifs stratégiques ont ensuite eux-mêmes été déclinés en objectifs opérationnels, recensés dans le tableau suivant. Ces pistes de réflexion ont vocation à **engager le dialogue avec l'ensemble des acteurs autour de la gestion et la gouvernance des données**.

Buts		Objectifs stratégiques		Objectifs opérationnels	
1	Promouvoir un cadre pour assurer la qualité des données et leur valeur ajoutée potentielle	1	Contribuer à l'évolution du cadre réglementaire et juridique pour s'adapter aux nouvelles opportunités et risques liés aux données	1	Connaissance du cadre juridique existant
		2	Capitaliser sur les nouveaux modèles de gouvernance et d'affaire pour proposer un panel de solutions, sensibiliser les élus et « acteurs de la donnée » à l'importance de la gouvernance des données	2	Evolution du cadre juridique
		3	Travailler à la convergence des formats de collecte des données	1	Lieu d'échange national sur la question de la donnée territoriale
2	Faire émerger le dialogue et les projets entre les parties prenantes de la collecte, du traitement et de l'usage des données	1	Favoriser le dialogue et la compréhension entre les acteurs à l'échelle nationale (publics, privés, recherche, citoyens)	2	Animation territoriale sous l'égide du groupe national
				3	Nécessité d'une politique d'administration des données publiques
		2	Renforcer la compréhension des enjeux de l'usage des données et l'émergence de projets au niveau local	4	Dispositif de centralisation des initiatives existantes
				5	Mise en œuvre opérationnelle à l'échelle du territoire
		3	S'appuyer sur et participer au développement des initiatives locales en matière de gouvernance des données territoriales, pour adapter les modèles de gouvernance aux différents besoins et niveaux d'acteurs et aux modalités de dialogue	1	Valorisation des initiatives et connaissances existantes
				2	Evolution des modèles de structuration des données
2	Faire émerger le dialogue et les projets entre les parties prenantes de la collecte, du traitement et de l'usage des données	1	Favoriser le dialogue et la compréhension entre les acteurs à l'échelle nationale (publics, privés, recherche, citoyens)	1	Dialogue à l'échelle nationale
				2	Compréhension mutuelle entre utilisateurs et fournisseurs de solutions
		2	Renforcer la compréhension des enjeux de l'usage des données et l'émergence de projets au niveau local	1	Lien entre enjeux territoriaux et modélisation
				2	Emergence de projets multi-acteurs
		3	S'appuyer sur et participer au développement des initiatives locales en matière de gouvernance des données territoriales, pour adapter les modèles de gouvernance aux différents besoins et niveaux d'acteurs et aux modalités de dialogue	1	Expérimentation et soutien de l'innovation
				2	Généralisation de la mise en place de structures de gestion des données innovantes
2	Faire émerger le dialogue et les projets entre les parties prenantes de la collecte, du traitement et de l'usage des données	3	S'appuyer sur et participer au développement des initiatives locales en matière de gouvernance des données territoriales, pour adapter les modèles de gouvernance aux différents besoins et niveaux d'acteurs et aux modalités de dialogue	3	Conception et diffusion d'un panorama des solutions
				4	Accompagnement juridique des acteurs

7. Annexes

7.1. Tableau synoptique des 11 outils analysés

Secteurs ou activités génératrices d'impacts examinés



Impacts modélisés ou examinés



Estimation du caractère intégré et multicritère de l'outil

Caractère intégré

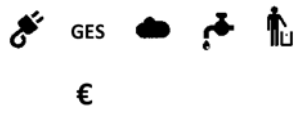
- Outils d'observation territoriale
- ● Modélisation simple: linéaire, sectorielle
- ● ● Modélisation avancée: rétroactions, multi-agents

Caractère multicritère

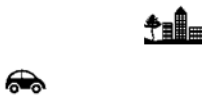
- Outils « monothématiques »
- ● Outils « multithématiques »
- ● ● Outils « multithématiques » avec un volet d'aide à la prise de décision

NB : Les informations présentées ci-après sont issues d'entretiens menées courant 2015 et validées par les porteurs des outils début 2016. Les caractéristiques et fonctionnalités des outils ont pu évoluer depuis.

OUTIL Outils des AASQA CIBLES Collectivités, Etat, industriels, associations, UE PORTEURS AASQA, ATMO France CREATION Outils de mesures: 1980 Outils d'inventaires: 2000 Modèles de simulation : 2000	DESCRIPTION Le réseau de mesures, dont les résultats sont à destination de l'Etat, des collectivités, industriels, association, de l'UE (rapport annuel), outils historiques des AASQA : le réseau comprend des stations de mesure, des camions-laboratoires, des campagnes de mesures, etc. Les outils de modélisation, à destination des collectivités et de l'Etat (PPA) : modèles régionaux ou interrégionaux à une maille de 3-4-5 km; modèles urbains sur une agglomération ou une zone industrielle avec une maille de 1m à 10m environ; modèles à l'échelle du quartier ou du bâtiment. Les outils d'inventaires, alimentant d'autres outils et pour les services de l'Etat et les collectivités (PPA, PCAET, SRCAE, PLUi, ...) : inventaires de quantification de toutes les consommations et de la production d'énergie liées aux activités humaines ainsi que des émissions de GES et de polluants sur les territoires.	SECTEURS MODELISES  IMPACTS 	CARACTERE INTEGRE Inventaires : ● Modélisation : ● ● CARACTERE MULTICRITERE Inventaires : ● ● Modélisation : ● ● ● SITE WEB http://www.atmo-france.org/fr/
OUTIL CMS Lyon Confluence CIBLES Grand Lyon Autres collectivités PORTEURS Toshiba Systems CREATION 2014	DESCRIPTION CMS Lyon Confluence, pour Community Management System, est un outil de visualisation en temps quasi-réel de la consommation et des productions d'énergie à l'échelle du quartier. Il constitue un des 4 piliers du projet démonstrateur, qui comporte par ailleurs : - La réalisation d'un bâtiment-modèle en efficacité énergétique et intégration d'EnR au sein du quartier de la Confluence à Lyon, - Le déploiement d'une flotte de véhicules électriques en auto-partage, - L'installation de compteurs (« energy box ») pour assister les habitants dans la maîtrise de leur consommation énergétique. CMS est un outil de gestion et de contrôle global des données liées à la consommation énergétique de ces 3 éléments listés ci-dessus. Les résultats sont calculés sur la base du signal transmis par les capteurs des bâtiments par pas de 10 min.	SECTEURS MODELISES  IMPACTS 	CARACTERE INTEGRE ● CARACTERE MULTICRITERE ● ● ● SITE WEB http://www.toshiba.fr/innovation/generi c/smart-community/
OUTIL Equitée CIBLES Collectivités : SCoT, PCET, SRCAE PORTEURS BURGEAP - ICE CREATION 2014 - 2015	DESCRIPTION EQUITEE est un outil conçu pour accompagner les collectivités, agences d'urbanisme, aménageurs et services de l'Etat dans l'élaboration des stratégies de transition énergétique (PCEAT, planification des réseaux) et la territorialisation dans les documents d'urbanisme (SCOT, PLU, PLH, PDU). L'outil livre un diagnostic des enjeux énergétiques et climatiques de la région à l'échelle du quartier (IRIS), avec un focus sur les vulnérabilités, et des liens entre population, habitat, aménagement, transports. La fonctionnalité prospective de l'outil permet d'intégrer des données spécifiques du territoire et tester des scénarios d'aménagement ou des programmes d'actions sectoriels.	SECTEURS MODELISES  IMPACTS 	CARACTERE INTEGRE ● ● ● CARACTERE MULTICRITERE ● ● ● SITE WEB http://www.equitee.fr/burgeap-business-geografic/

OUTIL ForCity	DESCRIPTION ForCity est à la fois une société et le nom donné à la plateforme territoriale de simulation de l'évolution urbaine développée par cette start up. La plateforme entend répondre au besoin d'approche systémique du territoire, en intégrant les composantes de la ville, et surtout leurs interactions mutuelles à différentes échelles de temps et d'espace : la population, sa composition et sa répartition sur le territoire ; les transports, l'énergie, l'immobilier, etc. L'outil ForCity propose une visualisation 3D des résultats et une utilisation collaborative entre utilisateurs. Il s'agit d'un simulateur spatial sur lequel on peut greffer différents processus linéaires (déplacements) ou réticulaires (réseaux). Les porteurs de l'outil insistent sur le caractère flexible et adaptable de la plateforme pour prendre en charge de nombreux raisonnements de modélisation et de simulation. Ils envisagent de continuer à y intégrer de nouveaux modules et fonctionnalités pour répondre aux besoins des acteurs publics et privés des territoires.	SECTEURS MODELISES  IMPACTS 	CARACTERE INTEGRE  CARACTERE MULTICRITERE  SITE WEB http://www.forcity.com/
OUTIL H+T	DESCRIPTION L'outil H+T (Housing + Transport) propose une définition de l'accessibilité économique du logement en combinant une estimation du coût du logement et du coût des déplacements. L'objectif est d'indiquer la meilleure localisation au regard du prix du logement et de la proximité des services, en y intégrant le coût des déplacements induits par cette localisation. Pour chaque zone d'une ville ou d'un territoire, les parts des coûts des déplacements d'une part, du logement d'autre part, et enfin des coûts cumulés logement + déplacements dans le budget moyen des ménages sont comparées à la part moyenne de ces coûts dans le budget des ménages du territoire considéré. L'outil propose aussi des cartes d'émissions de gaz à effet de serre permettant de comparer les émissions par zone et par ménage.	SECTEURS MODELISES  IMPACTS 	CARACTERE INTEGRE  CARACTERE MULTICRITERE  SITE WEB http://htaindex.cnt.org/
OUTIL MEU	DESCRIPTION L'outil MEU a pour objectif de fournir aux décideurs urbains et aux distributeurs d'énergie un outil d'aide à la décision permettant de planifier et de gérer des systèmes énergétiques en zones urbaines. L'outil permet plus précisément d'établir un état des lieux énergétique des zones urbaines, de construire et d'évaluer des scénarios énergétiques sur la demande et la production, et enfin de construire une base de données sur les systèmes énergétiques urbains. L'outil part du bâtiment comme entité de base, en utilisant des données réelles de consommation mesurées, ainsi que des données estimées de consommation.	SECTEURS MODELISES  IMPACTS 	CARACTERE INTEGRE  CARACTERE MULTICRITERE  SITE WEB http://meu.epfl.ch/

OUTIL Mobisim	DESCRIPTION MOBISIM est un outil de simulation individu-centré qui permet de simuler des scénarios prospectifs sur les mobilités quotidiennes et résidentielles (individus et ménages). Il s'agit d'un outil dit « LUTI » (Land-use and transport interaction) développé pour fournir une aide à la décision aux collectivités et aux aménageurs sur des questions d'urbanisation et de politiques de transports. L'outil est conçu pour fonctionner sur un serveur de calcul (il est actuellement sur le serveur du laboratoire ThéMa). Il est uniquement utilisé par des chercheurs du laboratoire, appliqué à un territoire principalement (Grand Besançon) et n'a pas fait l'objet d'une utilisation opérationnelle pour aider la décision publique.	SECTEURS MODELISES 	CARACTERE INTEGRE 
CIBLES Collectivités Projets de recherche			CARACTERE MULTICRITERE 
PORTEURS Laboratoire ThéMA		IMPACTS 	SITE WEB http://thema.univ-fcomte.fr/mobisim/
CREATION 2006			
OUTIL NEST	DESCRIPTION NEST (Neighborhood Evaluation for Sustainable Territories) est un outil d'évaluation environnementale quantitative (ACV) destiné à aider la prise de décision. NEST est destiné aux phases de conception des Opérations d'Aménagement. Il n'a donc pas vocation à être utilisé à l'échelle du bâtiment par exemple, mais il permet à l'utilisateur de maîtriser l'impact environnemental du projet par la comparaison de scénarios.	SECTEURS MODELISES 	CARACTERE INTEGRE 
CIBLES Promoteurs, constructeurs Collectivités Lycéens, étudiants			CARACTERE MULTICRITERE 
PORTEURS Nobatek		IMPACTS 	SITE WEB http://www.nobatek-nest.com/
CREATION 2011			
OUTIL NovaEQUER	DESCRIPTION NovaEQUER est un outil d'aide à l'éco-conception des quartiers via l'évaluation des impacts environnementaux par l'application de l'analyse de cycle de vie aux quartiers. L'outil résulte de l'évolution d'un outil d'ACV à l'échelle des bâtiments, EQUER. NovaEQUER fonctionne ainsi au sein d'une chaîne logicielle intégrant ce premier outil qu'il complète sur les aspects spécifiquement liés au quartier. L'outil permet de comparer différentes variantes de projets de quartier selon différentes étapes du cycle de vie et selon une série de critères sous la forme d'éco-profil. Il peut appuyer la décision publique à différents stades dans la conception d'un quartier et servir d'outil d'échange entre les parties prenantes du projet.	SECTEURS MODELISES 	CARACTERE INTEGRE 
CIBLES Collectivités Aménageurs Agences d'urbanisme			CARACTERE MULTICRITERE 
PORTEURS Mines ParisTech Izuba Energies		IMPACTS 	SITE WEB http://www.izuba.fr/logiciel/novaequer
CREATION 2014			

OUTIL PROSPER	DESCRIPTION PROSPER est un outil de prospective énergétique territoriale permettant aux communes et intercommunalités de construire des scénarios énergétiques sur leur territoire aux horizons 2020, 2030 et 2050. Sur la base d'un état des lieux initial et multi-sectoriel en 2010 et d'un scénario d'évolution tendanciel de ce bilan tiré du SRCAE, l'utilisateur simule l'impact d'actions à la maille communale ou ECPI : maîtrise de l'énergie et substitution dans les bâtiments et l'appareil industriel, sensibilisation des usagers, mobilité douce, électrique ou commune, développement des EnR. L'impact des actions diffère en fonction des caractéristiques de chaque commune. Les résultats sont comparés aux objectifs fixés dans le SRCAE et évalués en termes économiques (coûts d'investissement, d'exploitation et impact sur la facture du territoire). Ces scénarios d'action sont ensuite consolidés à l'échelle départementale via l'intervention du SIEL, deuxième utilisateur clé de l'outil. Cela permet de visualiser l'impact global des politiques locales mises en œuvre sur le territoire sur les consommations énergétiques et leur impact GES, et de vérifier leur compatibilité avec les capacités du réseau électrique et gazier.	SECTEURS MODELISES 	CARACTERE INTEGRE 
CIBLES Commune, EPCI de la Loire. A terme : autres territoires	IMPACTS  	IMPACTS  	CARACTERE MULTICRITERE 
PORTEURS SIEL Energies Demain	CREATION 2014	IMPACTS  	SITE WEB http://www.siel42.fr/media/Documents/fiches-techniques/plaquette-prosper-siel-2015.pdf
OUTIL Ultisim	DESCRIPTION ULTISIM est un projet d'outil de simulation de l'interaction occupation du sol – transport (dit « LUTI » : Land Use and Transport Interaction) ambitieux et unique en Europe : complètement intégré qui ne résulte pas du couplage a posteriori d'un logiciel de transport et d'un logiciel d'urbanisation. Le projet de l'outil a été initié dans le cadre d'une recherche financée par le MEDDTL dans le cadre du PREDIT 4. Il a été abandonné faute de financements pour la poursuite de son développement. Une première phase de conception avait été menée, qui a abouti à un rapport final décrivant l'architecture de l'outil et formulant des recommandations. L'outil ULTISIM aurait pu être utilisé dans le cadre de projets d'infrastructures ou de transport pour simuler sur le long terme les effets de ces projets.	SECTEURS MODELISES 	CARACTERE INTEGRE 
CIBLES Agences d'urbanisme, services techniques des territoires, BE, ...	IMPACTS GES 	IMPACTS GES 	CARACTERE MULTICRITERE  SITE WEB http://www.iau-idf.fr/savoir-faire/nos-travaux/edition/ultisim.html
PORTEURS IAU Ile de France	CREATION 2009 – abandonné depuis	IMPACTS GES 	SITE WEB http://www.iau-idf.fr/savoir-faire/nos-travaux/edition/ultisim.html

7.2. Grille d'analyse pour les outils de modélisation

7.2.1. Introduction

Cette grille est destinée à orienter la conception d'outils de modélisation territoriale. Elle peut être utile à d'autres acteurs participant à encourager le recours à la modélisation territoriale : financeurs, soutiens institutionnels et techniques, par exemple. Elle permet également d'accompagner les utilisateurs des outils de modélisation territoriale, dans le choix d'un outil adapté à leur besoin.

Le développement des outils de modélisation territoriale présente des enjeux de différentes natures (techniques, de gouvernance, informationnels...) qu'il s'agit de prendre en compte tout au long du cycle de vie d'un outil (conception, développement, diffusion, mise en œuvre, mise à jour voire remplacement).

Ces enjeux ont été identifiés au cours de la phase 2 de l'étude, dont l'objectif était d'analyser 11 outils de modélisation territoriale à l'aide d'entretiens auprès de porteurs et d'utilisateurs de ces outils. Le croisement des études de cas a permis de révéler cinq enjeux pour les outils :

- > facteurs clés de succès pour les **porteurs**
- > points à considérer pour les **utilisateurs**

Porteurs	Enjeux	Utilisateurs
Définir le service fourni par l'outil et le besoin auquel il répond	Pertinence	Comprendre précisément le service fourni par l'outil et le besoin auquel il répond
Sécuriser les conditions économiques et techniques de la pérennisation de l'outil	Pérennité	Examiner les conditions de pérennisation de l'outil
Documenter les hypothèses, les données d'entrée et le fonctionnement du modèle afin d'assurer la transparence de l'outil	Documentation	S'assurer de l'existence d'une documentation pour les hypothèses, les données d'entrée et le fonctionnement du modèle, qui garantit la transparence de l'outil
Consolider un protocole de collecte des données et favoriser l'utilisation de modèles génériques	Fiabilité	Vérifier la généricité des modèles utilisés et l'utilisation d'un protocole précis de collecte des données
Envisager l'interopérabilité afin de permettre l'intégration de l'outil avec son environnement	Interopérabilité	Privilégier les outils interopérables, ou dont les porteurs ou développeurs envisagent l'interopérabilité

Plus précisément, pour les porteurs d'outils :

- > Les deux premiers points - l'analyse des besoins et la sécurisation des conditions de pérennisation - sont nécessaires au développement même de l'outil puis à son utilisation et à sa diffusion ;
- > La documentation et l'usage d'un protocole de collecte de données connu pour un outil transparent, ainsi que l'utilisation de modèles génériques et évolutifs conditionnent la qualité de l'outil ;
- > Le dernier point est lié à la possibilité de diffusion et de déploiement de l'outil : l'interopérabilité est en effet essentielle à une entente et un dialogue entre les données et outils pour en faciliter l'usage.

Pour chacun des 5 facteurs clés identifiés, une fiche est proposée, organisée comme suit :

- > Le titre est formulé à la manière d'une consigne, comme par exemple « Définir le service fourni par l'outil et le besoin auquel il répond » ;
- > Une description du facteur de succès précise pourquoi ce point est particulièrement important dans le contexte actuel ;
- > Des recommandations sont ensuite formulées pour la conception d'outils.

Les utilisateurs sont invités à prendre également connaissance de ces recommandations – formulées en premier lieu pour les porteurs – pour être guidés dans le choix d'un outil répondant le mieux à leurs besoins.

7.2.2. Cinq facteurs de succès et recommandations associées pour concevoir un outil de modélisation territoriale

1. Définir le service fourni par l'outil et le besoin auquel il répond

Description du facteur de succès

Un outil de modélisation a toujours un but précis et des cibles identifiées.

Pour le porteur d'outil, il est tentant de vouloir proposer un service multifonctions, ou de mettre en avant la capacité de l'outil à répondre à de multiples besoins. Pourtant, l'identification du besoin auquel répond l'outil, puis des objectifs visés et utilisateurs cibles, est capitale et préalable à la construction efficace de l'outil. Le degré de complexité doit être adapté aux fonctionnalités développées au regard du besoin identifié, la complexité n'apportant pas toujours une plus grande précision.

Comme critère de choix d'un outil adapté à ses besoins, l'utilisateur peut s'assurer de la prise en compte effective, par les porteurs des différents outils, des mesures recommandées décrites ci-dessous.

Mesures recommandées

- > **Définir précisément le besoin** auquel l'outil répond, ainsi que les fonctionnalités associées. Par exemple, un outil permet de simuler l'évolution des mobilités intra-urbaines afin d'améliorer les réflexions des décideurs sur les orientations possibles des politiques de mobilité, besoin qui se traduit finalement par la mise à disposition d'un outil de scénarisation, et auquel correspondent des fonctionnalités d'analyse, de simulation, des modes de représentation et de transmission des résultats, des données, etc.
- > **Comprendre comment l'utilisateur se représente le besoin** et ce qu'il est prêt à faire pour y répondre. Il s'agit d'exprimer, aussi fidèlement que possible, la manière dont le besoin s'exprime pour l'utilisateur potentiel à une échelle donnée afin de préciser l'objectif (les objectifs) de l'outil pour les différents utilisateurs et périmètres d'utilisation identifiés. Par exemple, dans une collectivité, l'agent des services techniques qui mettrait en œuvre l'outil n'a pas la même représentation du besoin que le directeur qui validera la décision d'acquiescer les services de cet outil. Les **objectifs** liés aux usages possibles et aux différentes échelles d'application de l'outil peuvent ainsi être multiples.
- > **Etablir une liste détaillée des utilisateurs potentiels de l'outil** suite à l'identification du besoin auquel répond l'outil et des objectifs associés. Distinguer, si nécessaire, les utilisateurs (amenés à mettre directement en œuvre l'outil) et les bénéficiaires finaux des résultats en sortie de l'outil. Par exemple, s'il vise une aide à la décision des collectivités, l'utilisateur pourrait être un agent des services techniques, d'un bureau d'études, d'une agence d'urbanisme ou un aménageur, mais le bénéficiaire sera le décideur (élu) ou même le citoyen de la collectivité. Utilisateurs et bénéficiaires n'ont pas toujours les mêmes attentes vis-à-vis d'un outil, aussi les résultats devront être transmis de façon adaptée à ces différentes cibles.
- > **Mettre en avant les spécificités de l'outil** au sein d'un document diffusable : réponse à un besoin précis, objectifs, cibles, contexte d'utilisation, fonctionnalités, corpus des données nécessaire, niveaux d'échelles pertinents, en rappelant aux utilisateurs et/ou bénéficiaires que l'outil ne permet pas d'obtenir la solution mais doit être considéré comme une aide à la décision. Pour aller plus loin, envisager les perspectives d'**évolution** et d'adaptation possibles de l'outil en fonction de la mise à disposition de nouvelles données par exemple.
- > **Associer, dans la mesure du possible, les utilisateurs potentiels au développement de l'outil** en affectant une partie du budget à la réalisation d'entretiens prospectifs, en amont du développement, sur l'usage souhaité ou les fonctionnalités critiques par exemple, de sorte à identifier clairement les fonctionnalités. Cette dynamique a tout intérêt à être prolongée par une méthode de développement **agile**, à partir d'un prototype, par des itérations avec les utilisateurs (club utilisateurs, retours d'expérience). L'association de ces derniers permet en effet une véritable co-conception de l'outil et facilite également son appropriation par la suite. La conception d'outils sans intervention des utilisateurs est possible et parfois même nécessaire pour des raisons de

concurrence par exemple, et nécessite dans ce cas une identification d'autant plus précise des besoins et un accompagnement des utilisateurs pour une bonne appropriation de l'outil.

- > **Faciliter l'appropriation de l'outil par l'utilisateur**, grâce à des ateliers, des supports de formation, une documentation³³, la création de réseaux d'utilisateurs, une assistance par exemple.
- > **Engager un designer dans la conception de l'interface utilisateur** pour en assurer l'ergonomie (temps de saisie, données par défaut, visuel) et la facilité d'exploitation des données de sortie (cartographie, graphiques). La sollicitation d'un designer suffisamment tôt permet d'adapter l'ergonomie de l'outil aux besoins identifiés, et inversement.

³³ Voir 2.3. Documenter les hypothèses et le fonctionnement du modèle afin d'assurer la transparence de l'outil pour ses utilisateurs et bénéficiaires, et en faciliter l'utilisation

2. Sécuriser les conditions de pérennisation de l'outil

Description du facteur de succès

La viabilité et la pérennisation d'un outil repose en majeure partie sur la mise en place d'un modèle d'affaires pertinent et en adéquation avec les besoins et objectifs identifiés.

Pour les porteurs, il est ainsi important de réaliser une étude de marché pour connaître l'écosystème d'outils potentiellement concurrents ou complémentaires, et d'identifier les risques et leviers de réussite pour mettre en place les bonnes conditions de diffusion d'un outil innovant et répondant à un besoin.

Comme critère de choix d'un outil, les utilisateurs peuvent s'informer au sujet des conditions mises en place par les porteurs pour assurer la pérennité de l'outil, qui peuvent être économiques, techniques ou juridiques.

Mesures recommandées

- > **Faire une comparaison des différentes solutions** pouvant être développées pour répondre au besoin et objectifs identifiés (serveur dédié, logiciel SAAS, plug-in, etc.)
- > **Identifier les potentielles offres concurrentes** : besoin, objectifs et/ou cibles identiques.
- > **Identifier les conditions critiques** (exemples : données non disponibles, financement public non perçu) dans lesquelles le développement ou la mobilisation de l'outil ne pourrait pas se poursuivre et proposer des mesures adéquates de réponse à ces situations.
- > **Définir l'approche économique** : montage financier, partenariats (financement public, privé, entreprise par exemple), conditions d'utilisation, (licences basique, premium, etc.).
- > **Identifier les contraintes liées à la diffusion de données privées et individuelles.** Si cela est estimé nécessaire, s'associer à une véritable compétence juridique.
- > **Déterminer d'où provient la « propension à payer »** (*willingness to pay*) assurant la pérennité de l'outil - réglementation, réalisation d'économies dans la gestion d'une politique publique, besoin d'accéder à une donnée de référence - et estimer en parallèle l'importance du service rendu par une étude coût-bénéfice.
- > **Décrire de manière précise comment l'outil proposé permet aux utilisateurs des gains de performance**, d'efficacité ou d'efficacité de la politique publique, y compris au regard des moyens déployés dans le financement, l'acquisition ou le fonctionnement de l'outil ou des services associés. En d'autres termes, il s'agit de démontrer le caractère coût-efficacité de l'outil.
- > **Lister les canaux de diffusion de l'outil et dispositifs d'accompagnement des utilisateurs** (hotline, livret, etc.) intéressants à mettre en œuvre.
- > **Mettre en place un dispositif de recette** pour suivre les demandes d'amélioration, les corrections et les versionnements de l'outil, afin d'évaluer les bénéfices économiques qui en découlent.
- > **Si possible, proposer des API** avec certains modules de l'outil pouvant s'avérer intéressants pour d'autres applications, ou des services web sur les données et/ou traitements utilisés.

3. Documenter les hypothèses et le fonctionnement du modèle afin d'assurer la transparence de l'outil et la pertinence des résultats

Description du facteur de succès

La documentation est à deux niveaux :

- > La documentation analytique, gage de transparence des porteurs d'outils au sujet du fonctionnement de l'outil (hypothèses, description fine des modèles, données d'entrée), est essentielle. Elle permet de garantir la pertinence des résultats au regard de l'objectif visé par l'outil et son utilisation.
- > La production de supports à destination des utilisateurs permet d'en faciliter la prise en main.

Les utilisateurs peuvent, pour choisir un outil transparent, prendre connaissance des documents existants (supports utilisateurs et documentation analytique) qui pourront être élaborés par les porteurs sur la base des critères suivants.

Mesures recommandées

- > **Utiliser les normes existantes en matière de documentation analytique et de métadonnées**, les formats et formalismes reconnus pour documenter les données et traitements.
- > **Dédier un temps spécifique à la documentation analytique et aux supports à destination des utilisateurs** (par exemple sous la forme d'une journée par quinzaine). Faire apparaître, dans le budget de développement et de consolidation de l'outil, le temps spécifique accordé à la documentation, afin de permettre une amélioration continue et adaptée des documents.
- > **Rédiger une documentation analytique**, c'est-à-dire rendant compte des choix stratégiques faits dans la construction du modèle. Etablir un plan d'ensemble du modèle, présentant notamment l'architecture globale de l'outil, l'articulation et les fonctions des modules éventuels puis détaillant à ces différents niveaux les fonctions mobilisées, les hypothèses sur lesquelles elles s'appuient, les solutions techniques. Expliciter les raisonnements et contraintes aboutissant à la décision de représenter ou non les principales variables. Rédiger une procédure listant les opérations de contrôle qualité et de validation des résultats de l'outil. Cette documentation, différente du « guide de l'utilisateur » doit décrire comment le développement a surmonté les difficultés rencontrées.
- > **Distinguer modèle conceptuel** (qui présente les objectifs et l'articulation des différentes fonctions de l'outil) **et modèle physique** (qui associe des solutions techniques aux fonctions du modèle conceptuel) pour faciliter la réingénierie en cas de changement de technologie pour une fonction spécifique.
- > **Mettre éventuellement une partie du code ou les modèles utilisés en open source.**
- > **Concevoir des supports adaptés aux utilisateurs** pour faciliter la prise en main de l'outil.
- > **Etablir la liste des documents et livrables** – documentation destinée aux utilisateurs et/ou documentation analytique – **qui seront rendus publics** au sujet de l'outil. En permettre la consultation en accès libre.

4. Consolider un protocole de collecte de données et favoriser l'utilisation de modèles génériques

Description du facteur de succès

L'accessibilité des données est un critère fondamental dans le protocole de collecte des données.

Pour les porteurs, il convient en effet de favoriser la mise en place de modèles génériques permettant les croisements de plusieurs sources de données en entrée des outils, et plus largement de veiller à rendre les données en sortie de l'outil également aptes à être diffusées et réutilisées.

L'utilisateur pourra chercher à s'assurer du respect des mesures décrites ci-après pour le choix d'un outil de qualité.

Mesures recommandées

- > **Utiliser les normes**³⁴ en vigueur en termes de collecte de données.
- > **Décrire précisément les données et métadonnées pertinentes à collecter** pour le fonctionnement de l'outil. Estimer en particulier si certaines données ne sont significatives qu'à partir d'une masse critique et ainsi non pertinentes. Rendre publique la liste des données que l'outil souhaiterait pouvoir utiliser, de manière à encourager le développement d'initiatives de collecte fournissant ces données.
- > **Identifier les sources des données et leur statut** (public, privé, accès gratuit, libre, restreint, payant). Si nécessaire, identifier quelles seraient les meilleures données possibles même lorsque celles-ci ne sont finalement pas mobilisées (inaccessibles, trop coûteuses, etc.). Faire la prospection des données pouvant être collectés par le biais des nouvelles technologies de l'information (téléphonie mobile, utilisateur connecté) même quand celles-ci ne sont pas mobilisées.
- > **Favoriser les modèles relativement génériques** susceptibles d'intégrer de nouvelles données sans que ne soient remis en cause la structure et le fonctionnement de l'outil. Mettre en avant le caractère évolutif de l'outil et préciser les nouvelles fonctionnalités qu'il serait possible de développer grâce à l'apparition ou à la mise à disposition de nouvelles données.
- > **Pour aller plus loin, s'intéresser à l'usage qui peut être fait des données et informations issues directement de l'utilisation de l'outil** (requêtes effectuées, sorties obtenues), pour alimenter des réflexions plus globales (exemple : données obtenues à l'échelle d'une commune, réutilisées à l'échelle du département).
- > **Respecter les normes en vigueur ou standards *de facto* pour les formats des données de sortie**, pour faciliter la mise en place de structures ou plateformes de gestion et de partage des données territoriales, et ainsi permettre la réutilisation des résultats.

³⁴ Voir CNIL : <http://www.cnil.fr/nc/documentation/deliberations/normes-simplifiees>

5. *Envisager l'interopérabilité afin de permettre l'intégration de l'outil avec son environnement*

Description du facteur de succès

Pour les porteurs, il est généralement plus intéressant techniquement de s'appuyer sur les outils existants pour intégrer certaines fonctionnalités qui y sont déjà développées à l'outil en cours de développement, plutôt que de développer l'ensemble des fonctionnalités souhaitées. Par conséquent il semble également pertinent de veiller à rendre les nouveaux modules développés interopérables avec les outils potentiels futurs³⁵. Envisager l'interopérabilité de l'outil dans ces deux sens permet de fait au porteur de se positionner avec un avantage certain par rapport à l'offre concurrente.

L'utilisateur souhaitant privilégier un outil interopérable, et donc permettant un dialogue avec d'autres outils existants conçus pour répondre à des besoins différents, peut prendre connaissance des mesures ci-dessous conseillées aux porteurs.

Mesures recommandées

- > **Utiliser les normes**³⁶ **existantes** en matière d'interopérabilité de l'outil afin d'orienter le développement en ce sens.
- > **Etablir un budget interopérabilité** spécifique. Accorder une partie du temps de développement à la prise de contact avec les porteurs d'outils complémentaires pour documenter et préciser les conditions de mise en interopérabilité de l'outil du porteur.
- > **Identifier les outils complémentaires** avec lesquels l'outil du porteur doit pouvoir se mettre en interopérabilité, comme par exemple des outils travaillant à d'autres échelles. Identifier, pour chaque outil de la liste, les contraintes techniques et juridiques à respecter pour que l'interopérabilité soit respectée.
- > **Conserver cette démarche tout au long du développement et de la consolidation de l'outil**, afin de permettre une interopérabilité des versions futures de l'outil, avec des outils complémentaires encore non existants.
- > **Etablir des partenariats** ou accords avec les porteurs de ces outils pouvant être utilisés simultanément, afin de tirer profit de l'interopérabilité recherchée.

³⁵ Dans le secteur du bâtiment et de l'architecture par exemple, la Maquette Numérique est un format unique de description technique du bâtiment, devenu obligatoire, et que l'ensemble des parties prenantes doivent donc respecter pour travailler de façon collaborative.

³⁶ Voir le Référentiel Général d'Interopérabilité (RGI) conduit par la Direction Interministérielle pour la Modernisation de l'Action Publique (DIMAP) : <http://references.modernisation.gouv.fr/interopabilite>

Sigles et acronymes

AASQA	Association Agrégée de Surveillance de la Qualité de l'Air
ACV	Analyse du Cycle de Vie
ADD	Administrateur De Données
ADEME	Agence de l'Environnement et de la Maîtrise de l'Energie
AMO	Assistance à Maitrise d'Ouvrage
API	Application Programming Interface
ARENE	Agence Régionale de l'Environnement et des Nouvelles Energies (Ile de France)
ATMO	Fédération des AASQA
BRGM	Bureau de Recherches Géologiques et Minières
CADA	Commission d'Accès aux Documents Administratifs
CDO	Chief Data Officer
CG	Conseil Général – Conseil Départemental
CNIG	Conseil National de l'Information Géographique
CNIL	Commission Nationale de l'Informatique et des Libertés
CNN	Conseil National du Numérique
CR	Conseil Régional
CRIGE	Centre Régional d'Information Géographique
CSTB	Centre Scientifique et Technique du Bâtiment
DDAF	Direction Départementale de l'Agriculture et de la Forêt
DDE	Direction Départementale de l'Equipeement
DREAL	Direction Régionale de l'Environnement, de l'Aménagement et du Logement
EACEI	Enquête Annuelle sur les Consommations d'Energie dans l'Industrie
EMD	Enquête Ménage Déplacement
EnR	Energies Renouvelables
ENTD	Enquête Nationale Transport Déplacement
EPCI	Etablissement Public de Coopération Intercommunale
EPTB	Etablissement Public Territorial de Bassin
ESANE	Élaboration des statistiques annuelles d'entreprise
FING	Fondation Internet Nouvelle Génération
GES	Gaz à Effet de Serre
GIS-MU	Groupement d'Intérêt Scientifique – Modélisation Urbaine
IDDRI	Institut du Développement Durable et des Relations Internationales
IDG	Infrastructures de Données Géographiques
IDH	Indicateur de Développement Humain
IGN	Institut National de l'Information Géographique et Forestière
INPI	Institut National de la Propriété Intellectuelle
INSEE	Institut National de la Statistique et des Etudes Economiques
INSPIRE	Infrastructure for Spatial Information in the European Community (Infrastructure d'Information Géographique dans la Communauté Européenne)
IRIS	Îlots Regroupés pour l'Information Statistique
LAET	Laboratoire Aménagement Economie Transports
LUTI	Land-Use and Transport Interaction (Interaction Urbanisme et Transport)
MEDDE	Ministère de l'Ecologie, du Développement Durable et de l'Energie
MEDDTL	Ministère de l'Ecologie, du Développement Durable, des Transports et du Logement
OREGES	Observatoire de l'Energie et des Gaz à Effet de Serre

PCAET	Plan Climat Air Energie Territorial
PCET	Plan Climat Energie Territorial
PDU	Plan de Déplacement Urbain
PLH	Programme Local de l'Habitat
PLU(i)	Plan Local d'Urbanisme (intercommunal)
POS	Plan d'Occupation des Sols
PPA	Plan de Protection de l'Atmosphère
PREDIT	Programme de Recherche et d'Innovation dans les Transports Terrestres
PSI	Public Sector Information directive (Directive sur la réutilisation des informations du secteur public)
RAEE	Rhône Alpes Energie-Environnement (Agence Régionale de l'Energie et de l'Environnement)
RGD	Régie de Gestion des Données
RTM	Restauration des Terrains de Montagne
SAFER	Société d'Aménagement Foncier et d'Etablissement Rural
SDIS	Service Départemental d'Incendie et de Secours
SEDHS	Société d'Equipeement du Département de la Haute-Savoie
SGAR	Secrétariat Général pour les Affaires Régionales
SGMAP	Secrétariat Général pour la Modernisation de l'Action Publique
SIG	Système d'Information Géographique
SITADEL	Système d'Information et de Traitement Automatisé des Données Élémentaires sur les Logements et les locaux
SRCAE	Schéma Régional Climat Air Energie
SRE	Schéma Régional Eolien
TEPCV	Territoire à Energie Positive pour la Croissance Verte
TIC	Technologies de l'Information et de la Communication

L'ADEME EN BREF

L'Agence de l'Environnement et de la Maîtrise de l'Energie (ADEME) participe à la mise en œuvre des politiques publiques dans les domaines de l'environnement, de l'énergie et du développement durable. Elle met ses capacités d'expertise et de conseil à disposition des entreprises, des collectivités locales, des pouvoirs publics et du grand public, afin de leur permettre de progresser dans leur démarche environnementale. L'Agence aide en outre au financement de projets, de la recherche à la mise en œuvre et ce, dans les domaines suivants : la gestion des déchets, la préservation des sols, l'efficacité énergétique et les énergies renouvelables, la qualité de l'air et la lutte contre le bruit.

L'ADEME est un établissement public sous la tutelle conjointe du ministère de l'Ecologie, du Développement durable et de l'Energie, et du ministère de l'Éducation nationale, de l'Enseignement supérieur et de la Recherche.



ADEME
20, avenue du Grésillé
BP 90406 | 49004 Angers Cedex 01

www.ademe.fr