

**ÉCOLE DOCTORALE DES SCIENCES DE LA TERRE
ET DE L'ENVIRONNEMENT – ED 413**
Laboratoire Image Ville Environnement - UMR 7362 Unistra/CNRS

THÈSE présentée par :
Benoit RIBON

Soutenue le : **21 septembre 2020**

pour obtenir le grade de : **Docteur de l'université de Strasbourg**
Discipline/ Spécialité : Aménagement du territoire

**Le paradigme du métabolisme territorial :
Une approche par le traitement des données**

Version publique

THÈSE dirigée par :

M. BADARIOTTI Dominique Professeur, Université de Strasbourg, UMR LIVE
M. KAHN René Maître de Conférences, HDR, Université de Strasbourg, UMR BETA

RAPPORTEURS :

Mme BARLES Sabine Professeure, Université Paris 1, UMR Géographie-Cités
M. BEYER Antoine Professeur, Cergy – Paris Universités, EA MRTE

AUTRES MEMBRES DU JURY :

Mme TALANDIER Magali Professeure, Université Grenoble Alpes, UMR PACTE



Thèse financée par l'ADEME

Remerciements

Cette thèse, si elle est le fruit d'un travail individuel, n'était pas pour autant un travail solitaire. De nombreuses personnes m'ont soutenu durant cette aventure et c'est également grâce à ces personnes que ce travail a pu être mené jusqu'à son terme ; je les remercie ici !

En premier lieu, je remercie mes directeurs, Dominique Badariotti et René Kahn, qui m'ont apporté leur confiance et qui m'ont accompagné dans cette aventure depuis ses débuts. Leurs contributions furent précieuses.

Ensuite, je remercie les membres du jury qui m'ont fait l'honneur d'accepter d'étudier mon travail : Sabine Barles et Antoine Beyer comme rapporteurs, et Magali Talandier en tant que présidente du jury. J'espère que cette thèse alimentera substantiellement leurs réflexions et leurs futurs travaux.

Je remercie aussi l'Agence de l'Environnement et de la Maîtrise de l'Énergie qui, à travers son financement, m'a permis d'être dans de bonnes conditions pour mener cette thèse pendant trois ans.

Je remercie également les autres personnes qui ont pris de leur temps pour relire mes travaux et me faire part de leurs commentaires, m'aidant ainsi à rendre mes idées plus claires, et à traquer les coquilles orthographiques. Je pense en particulier à Marina et Christophe.

Du point de vue du moral, je remercie aussi mes collègues du laboratoire pour leur soutien et notamment mes collègues doctorants avec qui j'ai passé de nombreux bons moments hors du laboratoire et dont j'espère que le soutien aura été réciproque !

Pour finir, je remercie ma famille, mes amis et mon chat, ainsi que Marina qui n'a pas seulement relu, mais m'a surtout supporté au quotidien. Ce sont des personnes exceptionnelles, mais il n'y avait pas besoin de faire une thèse pour savoir cela

Table des matières

Remerciements.....	3
Table des matières	5
Glossaire.....	11
INTRODUCTION GÉNÉRALE.....	13
I. Contexte : Évaluer notre pression environnementale	13
II. Problématique et positionnement épistémologique	14
III. Plan de la thèse	17
PARTIE I ÉTAT DE L'ART.....	19
Chapitre 1 Éléments de cadrage d'une approche par le traitement des données.....	21
I. Introduction.....	21
II. La donnée et sa place dans les cycles de l'information	22
1) Définir la donnée	22
2) Chaîne et cycles de l'information.....	24
3) L'analyse et le traitement des données.....	25
III. Conception des systèmes d'information.....	33
1) Systèmes d'information.....	33
2) Structures de données	35
3) Ontologie, modèle de données et programmation orientée objet.....	43
4) Espaces centralisés de données	47
IV. Les enjeux sociétaux de l'utilisation de données	49
1) De nouvelles possibilités pour la recherche scientifique.....	50
2) Une évolution du monde économique.....	51
3) L'appui des politiques publiques.....	53
4) Un outil pour le développement durable ?	54
5) Vers une « algorithmisation » des prises de décision ?.....	55
V. Conclusion du premier chapitre	57
Chapitre 2 Fondements et applications d'une approche par l'analyse des flux physiques et le métabolisme territorial.....	59
I. Introduction.....	59
II. Le développement des analyses de flux de matières et d'énergie	60

III. La réalisation d'une AFME	65
1) À un niveau micro : AFME des organisations et de leurs processus	67
2) Au niveau d'un produit : Inventaire du Cycle de Vie	69
3) À un niveau macro : Le métabolisme des territoires.....	70
IV. Panorama des applications.....	73
1) Amélioration des procédés et de la circulation des flux.....	74
2) Évaluation des impacts environnementaux	74
3) Étude des phénomènes dispersifs et identification des gisements	75
4) Suivi de la réglementation applicable	77
5) Communication sur les progrès et les impacts environnementaux	78
6) Compréhension scientifique et vision partagée.....	78
7) Les flux de matières dans les projets de construction et d'aménagement....	79
8) Rendre l'économie circulaire, et l'industrie écologique	81
9) Des questions de résilience et d'anticipation de l'épuisement des ressources	83
V. Les limites des AFME	84
1) Des données éparses, difficiles à regrouper	84
2) La quantification des flux à travers différents cadres théoriques.....	86
3) Rendre intelligible l'AFME auprès des acteurs du territoire	88
4) L'évolution du cadre législatif	89
5) Des compétences spécifiques à mobiliser et la coordination des acteurs	89
VI. Conclusion du deuxième chapitre et perspectives.....	90

Chapitre 3 Éléments complémentaires du traitement des données pour l'étude du métabolisme territorial..... 93

I. Les usages, implications et contraintes dans la définition du périmètre.....	93
1) Le territoire étudié et le niveau de détail spatial envisagé	94
2) La période d'étude et le pas de temps	96
3) Les matières considérées et les nomenclatures utilisées pour classifier l'information.....	98
4) La ou les unités de quantification.....	101
5) Les processus spécifiques considérés.....	102
6) Le risque de double comptage.....	102
II. Autres éléments du traitement des données	103
1) La collecte des données	103
2) Le stockage des données	107
3) Méthodes d'analyse des données du métabolisme.....	109
III. La production de l'information sur le métabolisme d'un territoire	113
1) Les axes d'analyses du métabolisme.....	113
2) La représentation graphique du métabolisme.....	115
IV. Conclusion du troisième chapitre	117

PARTIE II PROPOSITION CONCEPTUELLE119

Chapitre 4 Conception d'un Système d'Information pour l'Analyse du Métabolisme des Territoires (SINAMET)..... 121

I. Analyse du sens de l'information et des besoins de traitement.....	122
1) Ontologie du métabolisme territorial	122
2) Analyse fonctionnelle.....	132
II. Conception du SINAMET	135
1) Définition du modèle de données d'une base centralisée	135
2) Usage des techniques de programmation orientée objet et abstractions complémentaires.....	137
3) Tableau d'association pour l'importation des données	139
4) Architecture modulaire.....	142
5) Information géographique et SIG.....	144
III. Prototype du SINAMET.....	145
1) Langage de programmation et SGBD	145
2) L'usage de bibliothèques	146
3) Structure du code source	147
4) État général de l'implémentation	148
5) L'implémentation des notions de transformations et de stock.....	148
6) Temps d'exécution	149
7) Interface graphique utilisateur.....	149
IV. Conclusion du quatrième chapitre	151

PARTIE III EXEMPLES D'APPLICATIONS.....153

Chapitre 5 Consommation d'énergie du patrimoine bâti de l'Eurométropole de Strasbourg 157

I. Cadrage de l'étude	158
1) Périmètre	158
2) Données sources	158
II. Utilisation du SINAMET.....	161
1) Préparation des données	161
2) Chargement du contexte et des données de consommation	164
III. Résultats obtenus.....	165
1) Vision globale de la consommation d'énergie	165
2) Recherche de spécificités	170
IV. Discussions sur l'approche mobilisée	174
V. Conclusion du cinquième chapitre	176

Chapitre 6	Transport de marchandises par les installations portuaires dans le bassin de vie de Strasbourg	177
I.	Éléments de contexte.....	177
II.	Cadrage de l'étude	179
1)	Périmètre de l'analyse de flux	179
2)	Données sources	179
3)	Utilisation du SINAMET	181
III.	Résultats.....	183
1)	Dans les départements français	183
2)	À l'échelle du Bas-Rhin	186
3)	Dans le Port Autonome de Strasbourg	189
IV.	Discussions	194
1)	Compréhension scientifique et vision partagée.....	194
2)	Amélioration des procédés et de la circulation des flux.....	194
3)	Les projets de constructions et d'aménagement.....	195
4)	Des questions de résilience et d'anticipation de l'épuisement des ressources	197
5)	Autres applications non abordées.....	198
V.	Conclusion du sixième chapitre	199

Chapitre 7	Sensibilité à l'échelle d'indicateurs de la comptabilité des flux	201
I.	Introduction.....	201
II.	Cas réel : les importations selon la structure administrative française	203
III.	Modèle mathématique	206
1)	Description du modèle	206
2)	Définitions de l'additivité et de la sensibilité à l'échelle	207
3)	Hypothèses	207
4)	Preuve de la sensibilité à l'échelle	208
5)	Discussion du modèle.....	209
IV.	Discussion et autres approches	209
1)	Dissocier l'origine d'importation	210
2)	Calculer l'importation individuelle	211
V.	Conclusion du septième chapitre	214

Chapitre 8	Le circuit de la matière alimentaire à Strasbourg.....	215
I.	Présentation de la démarche	215
1)	Partir d'un modèle du métabolisme humain	215
2)	Flux en amont : Production, transformation, distribution, préparation	216
3)	Flux en aval : Déchets alimentaires et eaux usées.....	218
4)	Synthèse	223
II.	Discussions et conclusion du huitième chapitre	225

CONCLUSION GÉNÉRALE.....	227
I. Apports et limites identifiés à travers les applications	227
II. Contributions et perspectives.....	229
BIBLIOGRAPHIE.....	233
Liste des illustrations.....	249
Liste des tableaux	252
ANNEXES.....	253
Annexe 1 Entretien sur l'étude des flux de matières, d'eau et d'énergie réalisée à l'échelle de l'Alsace	255
I. Stratégie régionale de soutien à l'économie circulaire	255
II. Méthodologie de l'AFME	256
III. Résultats.....	258
Annexe 2 Synthèse partielle des sources de données pour la comptabilité des flux de matières et d'énergie	261
Annexe 3 Modèle de données.....	265
Annexe 4 Exemples de modules d'importation (ETL).....	267
Annexe 5 Exemple d'un module de visualisation et exportation	269
Annexe 6 Cartographies autour des infrastructures portuaires.	271
I. Palimpseste portuaire	272
II. Infrastructures de fret	273
III. Filière bois.....	274
IV. Filière BTP et industrie lourde	275
Annexe 7 Métabolisme d'un humain	277
I. Respiration.....	277
II. Alimentation	277
III. Selles et urines	277
IV. Bilan hydrique	278
V. Synthèse.....	278
VI. Bibliographie de l'annexe.....	279
Résumé en français et en anglais.....	282

Glossaire

AFME	Analyse de Flux de Matières et d'Énergie
BA(67)	Banque Alimentaire (du Bas-Rhin)
BDD(R)	Base De Données (Relationnelle)
BES	Bilan Entrées/Sorties
CGDD	Commissariat Général au Développement Durable
COS	Contournement Ouest de Strasbourg (<i>aussi connu comme GCO – Grand Contournement Ouest</i>)
DBI	DataBase Interface (<i>interface avec la base de données</i>)
DJU(18)	Degrés-Jour Unifié (référence chauffage 18°C)
EDD	Entrepôt De Données
EF	Énergie Finale
EMS	Eurométropole de Strasbourg
EP	Énergie Primaire
ETL	Extract-Transform-Load (<i>Extracto-chargeur</i>)
EWMFA	Economy-Wide Material Flow Accounting (<i>Comptabilité des flux de matières pour une économie entière</i>)
GES	Gaz à Effet de Serre
ICPE	Installation Classée pour la Protection de l'Environnement
NST2007	Nomenclature pour les statistiques de transport 2007
MAS	Maison d'Arrêt de Strasbourg
MEA	Modèle Entité-Association
MS	Matière Sèche
MTES	Ministère de la Transition Écologique et Solidaire
ORM	Object-Relationnal Mapping (<i>mise en correspondance objet-relationnel</i>)
OWL	Web Ontology Language (<i>Langage web pour la description d'ontologies</i>)
PAS	Port Autonome de Strasbourg
PCS	Pouvoir Calorifique Supérieur
PIOT	Physical Input/Output Table (<i>Tableau Entrées/Sorties Physiques</i>)
RET	Rhine Europe Terminal
RPG	Registre Parcellaire Graphique
SCEE	Système de Comptabilité Économique et Environnementale

SCN	Système de Comptabilité Nationale
SGBD	Système de Gestion de Base de Données
SI(G)	Système d'Information (Géographique)
SINAMET	Système d'Information pour l'Analyse du Métabolisme des Territoires
SitraM	Système d'Information sur le Transport de Marchandises
SIE	Système d'Information sur l'Eau
SQL	Structured Query Langage (<i>Langage structuré de requêtes sur des bases de données</i>)
tep	tonne équivalent pétrole (<i>en termes d'énergie</i>)
TIC	Technologies de l'Information et de la Communication
VdS	Ville de Strasbourg
VNF	Voies Navigables de France
(k/M/G/T)Wh	(kilo/Mega/Giga/Tera)Watheure

Introduction générale

I. Contexte : Évaluer notre pression environnementale

Consommons-nous trop ? La question n'est pas nouvelle. Posée par Arrow et al. (2004), elle s'intègre dans la prise de conscience qui s'opère depuis les années 1970 concernant les enjeux environnementaux auxquels nous sommes et allons être confrontés : dérèglements climatiques, pollution de l'air, de l'eau, épuisement des ressources, réduction de la biodiversité et des services écosystémiques qui en dépendent... Aujourd'hui, de nombreux travaux permettent une réponse tranchée : nous détruisons actuellement plus de ressources naturelles qu'il ne s'en régénère par notre consommation et par le rejet de nos déchets, augmentant ainsi dangereusement notre dette écologique, et de surcroît, de plus en plus vite (IPCC 2014 ; Servigne & Stevens 2015 ; Wackernagel & Beyers 2019).

Ce contexte peu réjouissant devrait inciter les acteurs sociaux, aux premiers rangs desquels figurent les décideurs politiques, à faire preuve de prudence dans le développement de nos sociétés afin de ne pas dépasser les limites écologiques planétaires (Steffen et al. 2015). Ces limites concernent divers capitaux naturels, en termes de qualité et de quantité (eau, air, sols, biodiversité...) et posent le problème de leur définition. Et en parallèle, il faut pouvoir comparer ces limites avec la pression environnementale exercée par nos systèmes socio-économiques. Il s'agit de qualifier et de quantifier nos prélèvements et nos rejets, et de s'assurer qu'ils sont durablement compatibles avec les ressources naturelles dont nous disposons (Kahn 2015).

Pour réaliser cette évaluation, différentes méthodes ont émergé : Bilan carbone® (ADEME 2010), calcul de l'empreinte écologique (Wackernagel & Beyers 2019), inventaire du cycle de vie d'un produit (ISO 2006), comptabilité des flux de matières (Eurostat 2001)...

Bien que différentes dans leurs finalités, ces méthodes reposent sur une approche commune de comptabilisation des flux de matières et/ou d'énergie. Elles sont applicables à différentes échelles, celle d'un produit, d'une entreprise ou même d'un territoire, moyennant une complexité de réalisation accrue. Parmi ces approches comptables, un paradigme nous interpelle : celui du métabolisme des territoires. Basé sur une métaphore d'origine biologique, il permet d'embrasser de manière globale les notions de consommation de ressources en entrée d'un système territorial, et de rejets en sortie. Un nombre croissant d'études, pour la plupart récentes, cherchent ainsi à évaluer ce métabolisme pour construire une vision d'ensemble des ressources matérielles et énergétiques mobilisées par des villes, des régions ou des pays (Beloin-Saint-Pierre et al. 2016).

L'exercice d'évaluation du métabolisme d'un territoire reste toutefois délicat. D'une part, la notion même de métabolisme est sujette à interprétation, ce qui lui confère ainsi un caractère polysémique. Si le terme est conceptuellement accrocheur, il peut manquer d'opérationnalité à cause de ces contours flous. D'autre part, comptabiliser les flux de matières et d'énergie sur un territoire reste une tâche « longue et complexe » (Bahers 2012), notamment en raison de l'importante quantité et variété de données à mobiliser pour l'exercice (Pauliuk et al. 2019). Ces données décrivent en effet des flux et des stocks de matériaux aux caractéristiques économiques, sociétales et environnementales très diverses. Réussir à les agréger et à en extraire une information synthétique, pertinente et exploitable par les acteurs du territoire reste ainsi un défi systématique dans ce type d'étude.

II. Problématique et positionnement épistémologique

Notre ambition à travers cette thèse est de rendre plus opérationnel le paradigme du métabolisme territorial, en l'abordant principalement sous l'angle du traitement des données. Nous voulons savoir comment ce traitement de données est actuellement réalisé et comment il pourrait être amélioré, c'est-à-dire enrichi, facilité, accéléré et fiabilisé ? Ce positionnement a différentes implications.

Tout d'abord, nous ne considérons le métabolisme qu'à travers les données qui le décrivent. L'objectif est de transformer ces données en une information exploitable, essentiellement orientée vers la décision et l'action pour la mise en place d'un système socio-économique durable. De ce fait, nous nous inscrivons dans une logique d'« intelligence de l'information [...] qui, par sa concentration sur l'information de valeur, soutient la réalisation des objectifs ainsi que la pérennité de l'organisation » (Mallowan 2012). On parle aussi

d'intelligence économique même si le terme semble plus restrictif, ambigu sur sa réalisation et orienté sur sa finalité. Cette logique distingue données, informations et connaissances qui sont les étapes clefs de cycles de l'information qui permettent une meilleure compréhension et action sur la réalité.

Toutefois, cette approche basée sur l'intelligence de l'information ne permet d'analyser que ce qui s'inscrit dans ces cycles séquencés. Ce qui n'est pas décrit à travers des données ne nous est donc pas accessible, c'est-à-dire finalement une grande partie de la réalité complexe. Cela est d'autant plus vrai que les études de métabolisme cherchent généralement à mobiliser des données déjà existantes et non pas à en créer de nouvelles. Notre position nous rend donc complètement dépendants de ces données, sans possibilité d'aller voir au-delà. C'est une position que nous assumons : actuellement, il nous semble y avoir un nombre important de données disponibles, alors que les moyens pour leur analyse sont encore limités. Notre priorité est donc avant tout d'être capables d'analyser l'existant, plutôt que de chercher à le critiquer ou le compléter. Dans les faits, en analysant les données, nous y portons naturellement un regard critique, mais cela intervient plutôt à la fin du processus qu'en prérequis.

Dans la logique de l'intelligence de l'information, il y a également la question de l'usage de l'information pour construire une connaissance permettant d'entreprendre une action adéquate sur la réalité. Cette question va au-delà de la problématique du traitement des données, il s'agit de l'émanation cognitive de l'information. Elle dépend des dispositions sociales et psychologiques des individus à communiquer, interpréter et s'approprier l'information (Mallowan 2012). Dans nos travaux, nous n'avons toutefois pu réfléchir qu'à la marge sur la façon dont l'information sur le métabolisme d'un territoire peut devenir une connaissance exploitable par les acteurs. En effet, il était au préalable nécessaire de faciliter la construction de l'information, ce que nous avons fait mais qui a occupé l'essentiel de notre effort. Nous avons toutefois examiné comment le paradigme du métabolisme peut être mieux compris et utilisé, sans toutefois pouvoir faire valider l'approche par les acteurs d'un territoire.

Afin de rendre cette notion de métabolisme territorial plus pragmatique dans nos travaux, nous en avons également limité la définition : pour nous, le métabolisme est de l'ensemble des flux et stocks de matières et d'énergie manipulés par les acteurs d'un territoire. Cette interprétation peut être perçue comme réductrice par rapport à des considérations plus larges qui peuvent intégrer des composantes sociales et sociétales selon le positionnement de différents auteurs (Broto et al. 2012). Toutefois, elle nous permet de poser un cadrage clair et objectif et donc de rendre le paradigme du métabolisme territorial plus opérationnel selon nous.

Utiliser ce paradigme implique aussi une réflexion sur la substituabilité des matières. De nombreuses critiques sont formulées envers les approches économiques habituellement basées sur la valeur économique des biens et services. C'est ainsi qu'à leur question « consommons-nous trop ? », Arrow et al. (2004) apportent une solution économique reposant sur la possibilité de constituer un capital fabriqué venant compenser la diminution des ressources naturelles. Or, de nombreux travaux mettent en cause ce principe de substituabilité et le qualifient de durabilité *faible*, là où une durabilité *forte* exclut cette possibilité de substitution pour un développement économique soutenable (Figuières et al. 2007 ; Mardellat 2010 ; Kahn 2015). Par conséquent, l'analyse du métabolisme ne peut se limiter à l'usage d'une seule unité d'agrégation, que ce soit des euros ou des kilogrammes, et il faut être capable de manipuler une information plus complexe permettant de discriminer les matières selon des considérations multiples : impacts environnementaux, caractéristiques physiques, valeur économique, importance sociétale...

Finalement, notre approche est transdisciplinaire par notre volonté de croiser deux champs thématiques aux implications sociétales larges et complexes. D'une part, le traitement des données inclut bien d'autres aspects que des considérations techniques, notamment lorsqu'il sert au développement des territoires (Desrosières 2008). D'autre part, la notion de métabolisme s'inscrit dans le champ de l'écologie territoriale et vient interroger notre manière de gérer collectivement les ressources naturelles, avec pour finalité de changer la dynamique actuelle pour qu'elle soit soutenable dans le temps (Barles 2017).

Cette transdisciplinarité entraîne évidemment une difficulté à poser un cadre disciplinaire clair et bien délimité. Par exemple, la disponibilité des données repose en partie sur la réglementation, qui peut en permettre, ou non, l'effectivité. Examiner le processus de traitement de données nécessite ainsi de considérer la législation, ce qui ne relève pas de notre expertise disciplinaire, mais que nous ne pouvons pas non plus occulter pour en avoir une bonne compréhension. Il en est de même pour des aspects économiques et politiques qui sont des composantes incontournables de l'échelle territoriale à laquelle nous travaillons. En ce sens, les collectivités locales, par leur choix politiques ou leur fonctionnement administratif, ont un rôle primordial à jouer pour l'aménagement et la gestion de ces territoires, et des ressources qu'ils mobilisent.

Enfin, la transdisciplinarité a également un effet réflexif : en abordant l'étude du métabolisme des territoires par l'angle du traitement des données, une partie de nos observations peuvent aussi se généraliser à d'autres traitements dans d'autres études, indépendamment de la question du métabolisme. Nos travaux contribuent ainsi en partie à

renforcer les logiques de traitement de données et de l'intelligence de l'information de manière plus générale.

III. Plan de la thèse

Pour restituer nos travaux, nous les présentons en trois parties : un état de l'art, une proposition conceptuelle et des exemples d'application.

L'état de l'art vise à formaliser les différentes notions et approches à mobiliser pour traiter les données relatives au métabolisme des territoires. Dans le Chapitre 1, nous nous concentrons sur les aspects purement liés au traitement des données et donc indépendamment du thème d'étude. Il ne s'agit pas de dresser une vision générale de la thématique, mais de présenter les éléments conceptuels, techniques et sociétaux qui vont être mobilisés dans la suite du travail. Ces éléments sont toutefois isolés à part dans ce chapitre, car ils peuvent être mobilisés pour d'autres types d'études, notamment territoriales. Dans le Chapitre 2, nous abordons les analyses de flux de matières et d'énergie et leur déclinaison territoriale à travers la notion de métabolisme. La question du traitement des données y est secondaire, même si cet aspect ne peut pas être totalement occulté pour présenter les implications de ce type d'étude. Finalement, dans le Chapitre 3, nous cherchons à croiser les deux thématiques et à formaliser les autres aspects qui émergent à leur intersection. L'objectif de cet état de l'art est d'avoir une vue d'ensemble des rouages qui permettent, mais aussi rendent complexe, l'étude du métabolisme des territoires.

À partir de l'état de l'art, nous proposons dans la partie suivante, composée d'un unique Chapitre 4, la conception d'un Système d'Information pour l'Analyse du Métabolisme des Territoires (SINAMET). Cela répond à la volonté d'une plus grande automatisation du processus de traitement des données dans l'objectif de l'améliorer. Nous décrivons ainsi notre démarche de conceptualisation de ce SINAMET : la définition d'une ontologie, l'identification des fonctionnalités attendues, l'usage de différentes techniques et la réalisation d'un prototype.

Dans la dernière partie, nous appliquons notre approche sur quatre différents cas d'étude afin de mettre en évidence les apports et les limites. Nous montrons comment il est possible de construire une analyse plus systématique de la dynamique des flux physiques sur un territoire à partir de notre état de l'art, et comment le système d'information que nous proposons facilite l'analyse des jeux de données relatifs à cette dynamique. Dans la temporalité de nos travaux, il n'était pas possible d'envisager l'étude du métabolisme complet d'un territoire, nous avons

préfére évaluer nos contributions méthodologiques sur différents cas d'études avec des périmètres plus restreints. Nous verrons également à la marge comment d'autres approches comme la cartographie ou la modélisation peuvent venir compléter notre compréhension du métabolisme des territoires. Les cas retenus dans nos travaux sont ainsi : la consommation d'énergie du patrimoine bâti de l'Eurométropole de Strasbourg, les flux de marchandises transportés par voie navigable dans le Bas-Rhin et le Port Autonome de Strasbourg, la sensibilité à l'échelle des indicateurs d'importation et d'exportation, et les flux de matière alimentaire à l'échelle de l'Eurométropole de Strasbourg.

Partie I État de l'art

Chapitre 1 Éléments de cadrage d'une approche par le traitement des données

I. Introduction

Les données sont partout : elles structurent notre vie quotidienne, que nous en ayons conscience ou non. Que ce soient les prévisions météorologiques, le solde d'un compte en banque, la quantité de gaz à effet de serre relâché dans l'air chaque année, l'heure d'un rendez-vous... Tous ces éléments sont fondamentalement distincts mais ils sont tous décrits sous la forme commune de données.

Le recours à des données remonte aux fondements des civilisations humaines. La complexification de la société, et notamment le développement du commerce en Mésopotamie, dépassant les capacités de mémorisation des hommes, les a amenés à garder des traces physiques de leurs échanges. Au début, il s'agissait de jetons d'argiles, avant de prendre une forme symbolique sur des tablettes du même matériau : ce sont les débuts de l'écriture¹ au IV^e millénaire av. J.-C. (Herrenschmidt 2007). Au cours de l'histoire, les supports physiques ont évolué (argile, papyrus, papier, bande magnétique, disque dur...) ainsi que la symbologie (hiéroglyphes, alphabets, chiffres arabes, représentation binaire...) et les techniques de reproduction (manuelle, imprimerie, numérique). Une constante demeure toutefois : la construction et l'usage de données ont permis de créer de la valeur ajoutée tout au long des siècles en facilitant la gestion des stocks, en optimisant l'usage du temps et des ressources, en aidant à comprendre des phénomènes économiques, sociaux ou scientifiques, et ainsi gouverner nos sociétés (Minaud 2014).

L'usage de données a commencé à s'intensifier à la fin du XVIII^e siècle avec la révolution industrielle et le développement d'une comptabilité d'entreprise permettant de gérer cette plus grande productivité (Chapitre 2/II). En parallèle, des États commencent à mener des études statistiques sur l'évolution démographique, le commerce, la production industrielle...

¹ Ce sont aussi accessoirement les premières tablettes *numériques*.

(Nikitin 1992 ; Flichy 2013). L'apparition des technologies de l'information et leur développement vont permettre d'entretenir une dynamique exponentielle dans l'usage des données, appuyant de nouvelles pratiques économiques, scientifiques et sociales. Aujourd'hui, le volume, la variété et la vitesse de circulation des données sont devenus si importants que l'on parle désormais de données massives ou « big data » (AFNOR 2015). Les capacités de stockage, d'échange et de traitement des données continuent de croître à un rythme important : plus de 20 % par an (Hazas et al. 2016).

Cette dynamique montre que les données sont une ressource avec un potentiel très important, et pour un grand nombre d'applications. Ce qui nous intéresse en particulier dans nos travaux est de voir comment elles peuvent être mobilisées et traitées au service du développement des territoires afin d'améliorer leur gestion des ressources matérielles et énergétiques.

Dans ce premier chapitre, nous prenons toutefois le temps de formaliser plusieurs éléments généraux relatifs aux données et à leur traitement, indépendamment de leur domaine d'application. En effet, l'usage de données s'accompagne d'un certain nombre de considérations techniques, mais aussi politiques, juridiques, et même plus largement sociétales. Isoler ces considérations nous permet d'accompagner le lecteur peu familier des éléments spécifiques du sujet et aussi de structurer cette thèse en présentant séparément les concepts que nous allons mobiliser par la suite dans l'approche transdisciplinaire. Nous clarifierons le concept de « donnée » et de « traitement de données », et leur ancrage dans les cycles de l'information. Nous présenterons ensuite différents concepts techniques liés aux systèmes d'information et à la manière de structurer et manipuler concrètement les données. Finalement, nous évoquerons les enjeux sociétaux qui résultent d'un usage de plus en plus intensif des données.

II. La donnée et sa place dans les cycles de l'information

1) Définir la donnée

Il n'existe pas de définition consensuelle de ce qu'est exactement une donnée (Thessen & Patterson 2011). Le dictionnaire Larousse² en présente six et celle trouvée sur Wikipédia ne correspond parfaitement à aucune d'entre elles. Ces différentes définitions

² <http://www.larousse.fr/dictionnaires/francais/donn%C3%A9e/26436?q=donn%C3%A9e>

permettent toutefois de mettre en avant certaines caractéristiques afin de faire de la donnée un élément plus concret et opérationnel de nos réflexions.

Au premier abord sémantique, nous pouvons comprendre une donnée comme le fait qu'elle soit... donnée³ ! Dans ce sens, nous constatons en effet que les utilisateurs mobilisent plus souvent des données produites et « données » (plus exactement distribuées) par d'autres organisations (entreprises, laboratoires, institutions...), plutôt qu'ils ne les construisent eux-mêmes. Et même si certains utilisateurs en viennent à produire leurs propres jeux de données, ceux-ci reposent sur la capitalisation d'informations obtenues (« données ») par des capteurs ou à travers une enquête sur des individus. Dans cette lecture sémantique, nous y voyons donc l'idée d'une description intangible de la réalité qu'il ne s'agit pas *a priori* de remettre en question et qui place les données comme point de départ d'un raisonnement. La confiance dans leur fiabilité est donc indispensable et est étroitement liée à la crédibilité des acteurs qui les ont produites (Brackstone 1999). Cette nécessaire crédence envers les données génère des effets sociétaux particuliers que nous verrons par la suite (section IV).

D'une manière plus pragmatique, nous pouvons définir la donnée comme une description élémentaire discrète d'une partie de la réalité globale et continue. Une donnée est ainsi l'association d'une conceptualisation qui désigne ce qui est décrit, avec une valeur explicite, c'est-à-dire stockable et transmissible⁴. Cette valeur s'inscrit dans un domaine de définition nécessaire à l'interprétation de la valeur pour le concept identifié, par exemple le typage : nombre, date, texte, association avec d'autres concepts... (Noy & McGuinness 2001 ; Boydens 2012). Les données décrivent donc des caractéristiques d'éléments (la population d'un territoire, le prix ou le nom d'un objet, la couleur d'une voiture...) ou bien des liens entre ces éléments (les départements d'une région, les modèles de véhicules d'un fabricant, l'auteur d'un article...). Pour les rendre plus exploitables, les données sont généralement structurées dans un ensemble cohérent (section III.2) et contextualisées, notamment au moyen de métadonnées⁵.

³ En anglais également « data » dérive du latin « datum » que l'on peut traduire par « la chose donnée ».

⁴ Sous forme numérique ou écrite.

⁵ Il s'agit des données qui décrivent un jeu de données (producteur, méthode, temporalité...).

2) Chaîne et cycles de l'information

Les données viennent s'inscrire dans une chaîne de l'information, qui vient en fait distinguer données, informations et connaissances et comporte différentes étapes (Bedard 1986 ; Dao 2005 ; Mallowan 2012) : la mesure qui produit des données brutes à partir de la réalité à travers des capteurs, des sondages ou des plans d'expérience (Saporta 2011) ; le regroupement et la structuration de ces données brutes pour obtenir des données structurées ; leur analyse pour en extraire de l'information ; la communication qui permet de diffuser l'information et d'en faire une connaissance (voir Figure 1).

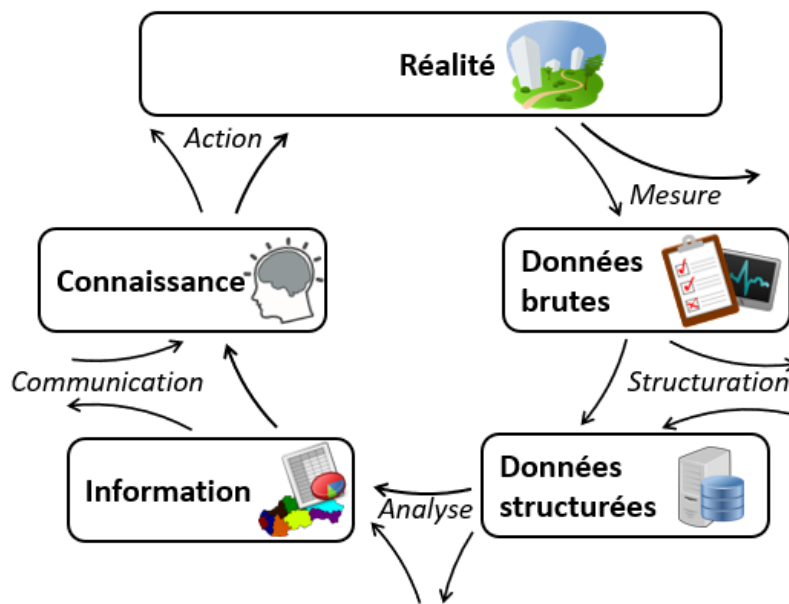


Figure 1 : Le(s) cycle(s) de l'information

Il semble pertinent de fermer cette chaîne et d'en faire un cycle : finalement, la connaissance vient influencer l'action sur la réalité (Polanco 2008 ; Mallowan 2012). Nous proposons également de croiser les cycles car l'information circule très peu en circuit fermé et différents cycles peuvent s'alimenter mutuellement. En effet, les analyses se basent souvent sur plusieurs jeux de données structurées, la connaissance résulte d'une compilation d'informations diverses et il n'est pas rare que des données servent à des applications et des analyses pour lesquelles elles n'étaient pas initialement destinées (Davis 2012), éloignant au passage l'utilisateur final de la source productrice de l'information (Boydens 2012). Nous interprétons cette complexité du cycle de l'information comme une incitation à bien dissocier les étapes du traitement de l'information. Il s'agit de faire en sorte que l'information produite puisse aussi contribuer à construire des connaissances pour d'autres acteurs ayant d'autres objectifs.

Différentes problématiques surviennent tout au long de ces cycles. Par exemple, la qualité des données structurées est évidemment un point fondamental des démarches d'analyse (Brackstone 1999). En fait, les jeux de données peuvent contenir des erreurs ; en particulier, la mesure ne peut constituer qu'une approximation de la réalité en raison de son caractère réductionniste et des aléas inévitables. L'interprétation des concepts administratifs, légaux ou scientifiques qui évoluent par nature joue un rôle prépondérant dans la définition de conventions permettant de récolter, manipuler et structurer les données. Cela peut générer des ruptures de séries ou contraindre les données à rentrer dans des cases prédéfinies, limitant ainsi la possibilité d'une compréhension plus subtile de la réalité (Desrosières 2008 ; Boydens 2012).

Par ailleurs, la question de la production de connaissances qui résulte de la communication et du partage d'informations, écrites ou non, entre plusieurs personnes est également essentielle pour agir adéquatement sur la réalité. Toutefois, les mécanismes à l'œuvre dans cette phase reposent sur des dynamiques humaines et font donc appel à d'autres champs disciplinaires, notamment la communication, la psychologie et la sociologie (Polanco 2008 ; Mallowan 2012).

Ainsi, pour améliorer notre action sur la réalité, il nous semble pertinent d'agir sur les différentes étapes du cycle de l'information. En ce sens, les dynamiques d'ouvertures de l'information (comme l'*open access* pour les publications scientifiques ou l'*open data* des administrations) nous semblent être des tendances de fond qui viennent transformer la façon dont l'information circule actuellement. Cela ne va pas sans poser de nouvelles questions ou générer de frictions avec certains modèles économiques, par exemple avec les éditeurs de revues scientifiques (Chartron 2016).

Dans nos travaux, nous avons choisi de nous focaliser sur une partie seulement du cycle de l'information, à savoir la phase d'analyse et le traitement des données qui la soutient. Ce positionnement suppose d'accepter les données structurées telles qu'elles sont fournies par les différentes organisations qui les produisent, sans juger de leur qualité. Similairement à Polanco (2008), nous distinguons l'objectif et la forme de l'analyse pour nous concentrer sur la seconde.

3) L'analyse et le traitement des données

L'analyse, telle que décrite dans le cycle de l'information, demande une attention particulière dans la manipulation des données. Il s'agit d'être capable d'exploiter des données sources structurées pour produire de l'information sous forme de modèle, de description

synthétique ou encore de mise en exergue de phénomènes... Il n'existe pas une méthode unique d'analyse, tout dépend des objectifs de l'étude, du contexte ou des données disponibles.

Malgré la diversité des analyses, nous identifions différentes étapes systématiques dans leur déroulement (Bertrand 1986 ; Polanco 2008) :

1. La définition de la problématique, du cadre théorique, des hypothèses et du périmètre de l'analyse,
2. L'identification et la collecte des données sources pertinentes,
3. Leur préparation pour l'application des méthodes d'analyse de données,
4. L'analyse des données qui peut produire différentes formes de résultats,
5. Le stockage des données entre ces différentes étapes,
6. L'interprétation visant à contextualiser les résultats pour y apporter un regard critique et répondre à la problématique en précisant les apports et les limites.

Afin d'éviter une confusion que nous pouvons trouver dans la littérature, nous distinguons bien l'« analyse » au sens général du terme qui vise à produire de l'information à partir de données sources, de l'« analyse des données » qui en est l'une des étapes, et qui vise à l'application d'une méthode sur des données préparées et produisant des résultats non interprétés, par exemple des coefficients de corrélations, des constructions graphiques ou des valeurs numériques. Une analyse peut s'appuyer sur l'application d'une ou de plusieurs méthodes d'analyse des données.

Les étapes 2 à 5 impliquent de manipuler concrètement des données⁶ et constituent le processus de traitement des données de la phase d'analyse⁷ sur lequel nous portons notre attention dans la suite de nos travaux. Nous approfondissons ici chacune de ces étapes pour mieux identifier les différents éléments sur lesquels il est possible d'agir afin de faciliter le traitement de données de manière générale. Dans le chapitre III, nous reprendrons ce processus mais en application sur la thématique spécifique de l'étude du métabolisme des territoires.

⁶ Si à une époque, le traitement de données était opéré à la main, l'usage d'outils informatiques paraît aujourd'hui incontournable.

⁷ Dans les autres étapes du cycle de l'information, il peut être également nécessaire de traiter des données, mais l'approche diffère et sort de notre focale.

a) Collecte des données sources et notion d'utilisabilité

La collecte des données est la première étape du traitement des données. Loin d'être triviale, cette étape nécessite d'explorer l'univers des données disponibles et de sélectionner les plus adaptées, selon différents critères plus ou moins conscientisés (Bellino 2013). Dans leurs travaux, Zuiderwijk et al. (2012) dressent un long panorama des difficultés qui peuvent survenir dans l'utilisation de données. Nos recherches ont été confrontées en effet à un certain nombre des problèmes mentionnés : l'absence de métadonnées, la difficulté à utiliser certains moteurs de recherche, la fragmentation des données, la diversité des formats de fichiers dont certains non adaptés à la lecture par une machine, la structure, l'encodage ou le typage des données défaillant ou non conventionnel, le coût, les restrictions d'accès, l'absence d'indexation (URL) des pages de données...

Pour mieux désigner ces difficultés, la notion d'« utilisabilité » nous paraît pertinente. Selon Bellino (2009), cette notion recoupe la facilité à trouver les données (ou repérabilité) et leur qualité, c'est-à-dire leur fiabilité, validité, exhaustivité, intelligibilité, concision, pertinence, richesse et cohérence (Batini et al. 2009 ; Bellino 2013). Évaluer la qualité des données est toutefois un exercice difficile car il nécessite finalement de connaître la réalité que sont censées décrire les données (Boydens 2012). En plus de la repérabilité et de la qualité, les questions d'accessibilité (physique, technique, droits d'utilisation) et d'interopérabilité des données par différents programmes sont aussi en lien avec leur utilisabilité (Berners-Lee 2006). Enfin, la spatio-temporalité des données, surtout dans les études territoriales, est aussi un critère que nous pouvons distinguer, même s'il pourrait être inclus dans la notion de qualité (pertinence).

Nous synthétisons ces différents critères de la notion d'utilisabilité dans le Tableau 1. Il ne s'agit pas seulement d'une approche descriptive, mais plutôt prescriptive : rendre les données utilisables est essentiel pour favoriser l'émergence de nouvelles découvertes et appuyer l'innovation (Thessen & Patterson 2011). Cela est souhaité par de nombreux chercheurs qui portent des ambitions d'ouvertures des données (Davis 2012 ; Zuiderwijk et al 2012), ou encore qui invitent à suivre des principes « FAIR », dans un sens littéral en visant à rendre les données utilisables par tous, et dans le sens de l'acronyme *Findability, Accessibility, Interoperability, Reusability* (Wilkinson et al. 2016). Il existe également la directive européenne INSPIRE citée comme référence par certains de nos interlocuteurs institutionnels et qui définit différents

principes « pour assurer l’interopérabilité entre bases de données et faciliter la diffusion, la disponibilité, l’utilisation et la réutilisation de l’information géographique en Europe »⁸.

Critère d'utilisabilité	Description	Pratiques associées pour renforcer l'utilisabilité
Repérabilité	Facilité à trouver les données.	Mise en ligne, sur des plateformes dédiées disposant de bons moteurs de recherche ; Présence de mots-clefs, d'une description riche et persistante (via les métadonnées), de liens vers d'autres données, d'un identifiant unique global et persistant (URL/URI/DOI/...) pour un accès direct.
Qualité	Fiabilité, validité, exhaustivité, intelligibilité, concision, pertinence, richesse et cohérence.	Tests statistiques ; Provenance et méthodologie de création des données explicite.
Accessibilité	Accessibilité physique (papier < ordinateur < web), technique (requête ou téléchargement), droits d'accès, droits de diffusion.	Mise en ligne ; Usage des standards de l'Open Data (plateformes, protocoles, licences...) ; Absence d'identification ou de droit d'accès pour accéder aux données ; Gratuité ; Accès aux données par téléchargement de fichier et par requête via une interface de programmation (API).
Interopérabilité	Formats et structures des données exploitables par différents programmes	Données disponibles sous différents formats dont des non propriétaires et lisibles par des machines ; Lien avec d'autres données.
Spatio-temporalité	Correspondance spatiale et temporelle entre les données disponibles et le sujet d'étude	Portails de données rassemblant et archivant les mêmes données sur différents territoires, ou des données différentes d'un même territoire ; Niveau de détail spatial et temporel fin.

Tableau 1 : Critères d'utilisabilité des données

⁸ http://cnig.gouv.fr/?page_id=8991

b) Préparation des données : restructuration, nettoyage, formatage, jointure et sélection

Les données collectées doivent souvent être retravaillées car elles peuvent contenir des erreurs, être mal structurées, segmentées ou non compatibles entre elles. Cette étape de préparation vise à permettre l'application des méthodes d'analyses de données de façon cohérente (Besse & Laurent 2016). Pour cela, différentes opérations peuvent devoir être menées :

- La restructuration visant à corriger les défauts sur la structure des données : des commentaires intégrés entre les lignes de données d'un tableur, du texte au milieu de nombres... L'utilisation d'un type de fichier de données non structuré (comme un fichier PDF ou de traitement de texte) peut également obliger à devoir retravailler manuellement les données pour les structurer (Berners-Lee 2006).
- Le nettoyage visant à détecter et à corriger les erreurs et les inconsistances dans les données (Rahm & Do 2000). Le nettoyage n'est pas toujours trivial. Même si une partie des erreurs peuvent être détectées (adresses postales erronées, dates aberrantes, codes incorrects, données redondantes...), ce n'est pas toujours le cas, notamment si elles portent sur la qualité intrinsèque des données (Boydens 2012). Par ailleurs, quand bien même une erreur est détectée, sa correction n'est pas forcément simple ni même possible. Cela exige souvent un travail manuel, ou l'exclusion de la donnée erronée, au risque d'altérer la qualité d'une étude.
- La standardisation (ou normalisation) visant à s'assurer de la bonne compatibilité des données (Batini et al. 2009). Il peut s'agir du choix du séparateur décimal (point ou virgule), de l'encodage des caractères spéciaux et accentués, de l'usage des majuscules, du format des dates et des nombres⁹, ainsi que de l'usage des mêmes conventions entre les différentes données (codes, noms, nomenclatures, unités, pas de temps...).
- La jointure visant à rassembler différents jeux de données pour n'en faire qu'un seul, généralement à l'aide d'une clef de jointure, c'est-à-dire un code d'identification commun dans différents jeux de données.

⁹ C'est un problème classique que l'on rencontre avec un numéro de téléphone inscrit dans un tableur : 0636353433 devient ainsi 636353433, mais c'est aussi le cas des codes de département Insee sur deux chiffres qui posent problèmes pour les départements 01 à 09.

- La sélection visant à ne conserver qu'une partie des données : celles destinées à l'application de l'analyse de données. Ces données figurent à l'intérieur du périmètre défini (objet d'étude, limites, période...) et ne concernent qu'un nombre restreint de variables ou caractères. Généralement, les données sélectionnées sont facilement extraites des données initiales par les fonctions de filtrage des tableurs ou des requêtes sur bases de données, ce qui permet de conserver les données initiales pour d'autres analyses de données. La sélection peut aussi s'opérer par effacement des données non considérées, mais cela les rend évidemment indisponibles pour d'autres analyses.

Ces opérations sont plus ou moins lourdes à mettre en place selon les données collectées et les objectifs de l'analyse. Dans certains cas, les données récupérées sont directement exploitables, mais en général, un travail de préparation est nécessaire. Ce travail est souvent manuel car spécifique à chaque jeu de données. L'automatisation de certaines procédures est parfois possible, mais cela nécessite quand même d'implémenter les algorithmes dédiés, qui reposent souvent sur une compilation de fonctions de bas niveau (Thomsen & Bach Pedersen 2009). Si la phase de préparation est rarement valorisée au regard du travail qu'elle engendre (Besse & Laurent 2016), elle n'est restée pas moins cruciale pour une bonne exploitation des données. Des données mal préparées peuvent empêcher l'application des méthodes d'analyse de données, ou de façon plus délétère, aboutir à des résultats incorrects pouvant engendrer de mauvaises décisions (Rahm & Do 2000 ; Boydens 2012).

c) Analyse de données

L'analyse de données est l'application d'une méthode mathématique ou algorithmique sur des données préparées afin d'en extraire une information synthétique, éventuellement visuelle. Ces données décrivent souvent une grande quantité d'éléments et un grand nombre de leurs caractères (Bouroche & Saporta 2006). L'extraction d'information peut consister à identifier des corrélations (ou non) entre différents caractères étudiés, à classifier des concepts ou individus selon leur similarité, à identifier des liens, à vérifier des modèles, à décrire les données de manière résumée... Parmi les méthodes plus courantes d'analyse de données, nous pouvons citer (Bertrand 1986 ; Bourouche & Saporta 2006) : la régression, les analyses statistiques et tests d'hypothèse, l'analyse en composantes principales, l'analyse canonique, l'analyse factorielle des correspondances, les méthodes de classification et de discrimination, et l'expression visuelle ou géométrique.

Ces méthodes vont permettre d'extraire de l'information à partir de caractéristiques de différentes natures, par exemple l'analyse spatiale interroge les localisations et proximités, l'analyse temporelle interroge les évolutions... De manière générale, nous pouvons parler d'axe d'analyse ou de dimension (Pitarch et al. 2010) pour désigner cette nature de l'information.

Les résultats produits par les méthodes d'analyse de données prennent plusieurs formes :

- des valeurs clefs ou des indicateurs,
- des modèles permettant de décrire les données,
- des données structurées : un tableau, un arbre de classification...
- des représentations graphiques.

Pour être correctement interprétés, ces résultats doivent être confrontés à l'ensemble des hypothèses ou limites qui auront pu être évoquées et sur lesquelles reposent les données, depuis la phase de problématisation jusqu'au choix des méthodes d'analyses utilisées. Cette interprétation est fondamentale, complexe, spécifique à la problématique, en partie subjective, et s'appuie plus souvent sur une argumentation littéraire que comptable. Elle n'entre ainsi pas directement dans la focale de nos travaux sur le traitement de données, même si elle est forcément mobilisée sur nos cas d'étude en Partie III pour dresser les apports et les limites de notre approche.

Par ailleurs, un de nos objectifs est de permettre à un public large d'appréhender les enjeux matériels et énergétiques de nos sociétés. Si l'interprétation des données par des experts peut apporter un éclairage sur une problématique donnée, nous voulons également permettre aux acteurs de construire leur propre interprétation des résultats. Parmi les types de résultats évoqués précédemment, les représentations graphiques semblent les plus à même d'être vecteurs d'une information complexe pour une grande variété d'acteurs. Elles permettent de construire des interprétations intuitives, là où les autres types de résultats nécessitent un examen et une compréhension plus approfondis. Cela dit, un usage mal pensé des techniques graphiques peut aussi mener à véhiculer une information déformée, voire erronée¹⁰. Ainsi, à l'instar de Chen et al. (2008), nous pensons que la construction de représentations graphiques mérite une

¹⁰ A titre d'exemple simple, les cartes géographiques sont inadaptées pour la représentation de résultats électoraux : il existe un découplage entre les surfaces des circonscriptions et leur poids électoral. Il faudrait ainsi déformer les cartes pour que les surfaces représentées des circonscriptions soient toutes identiques à travers par exemple une anamorphose.

véritable attention, à la fois pour diffuser efficacement une information mais aussi pour éviter d'éventuels écueils.

d) Stockage des données

La question du stockage des données intervient nécessairement et tout au long du processus de traitement : dès leur collecte, et même avant car il a fallu les rendre disponibles, pendant et après leur préparation, et enfin pour figer les résultats après l'analyse des données. Le stockage permet de disposer de manière pérenne les données, pour pouvoir les réutiliser ultérieurement ou les partager. Il est envisageable de manipuler les données en flux tendus dans la mémoire vive, sans stockage sur un disque dur, mais cela implique de devoir reproduire l'ensemble du processus de traitement de données à chaque fois, ce qui peut s'avérer vite chronophage.

Nous identifions deux grands paradigmes pour le stockage des données numériques : les fichiers et les bases de données. Les premiers sont associés à des hiérarchies et des programmes permettant de manipuler les données à travers une interface utilisateur. Les secondes sont associées à des interfaces de programmation permettant de formuler des requêtes pour ajouter, modifier ou sélectionner les données. Le choix du paradigme dépend de la complexité des données à manipuler, de l'adéquation des outils logiciels disponibles et des compétences de l'utilisateur. Nous noterons que l'échange de fichiers (téléchargement ou envoi par mail) constitue un facteur de risque pour l'intégrité des données : les données originales sont dupliquées et peuvent donner lieu à des versions différentes si elles sont modifiées. C'est aussi dans le rôle centralisateur des bases de données que réside une partie de leurs avantages (Inmon 2005).

Pour les études simples ou exploratoires, l'usage d'un fichier tableur est souvent privilégié. Il permet d'organiser facilement des données, et les programmes d'exploitation (comme LibreOffice Calc ou Microsoft Office Excel) permettent de réaliser de nombreuses opérations courantes (calculs, graphiques, tri...). Toutefois le format en deux dimensions d'un tableur présente des limites, et pour certains niveaux de complexité, il faut envisager d'autres structures pour les données : d'autres formats de fichiers ou plus classiquement l'utilisation d'une base de données.

III. Conception des systèmes d'information

1) Systèmes d'information

La section précédente présente les étapes habituelles dans les traitements de données. Il y a un grand nombre d'éléments à considérer et cette complexité ne peut pas être exposée pleinement à l'utilisateur. Aussi, le recours à des abstractions est nécessaire pour lui permettre de manipuler plus rapidement et facilement les données et de réduire les risques de confusion et d'altération. L'implémentation d'un Système d'Information (SI) a notamment comme objectif de rendre disponibles les abstractions nécessaires au bon déroulement d'un processus basé sur des données. Il permet d'interagir avec des utilisateurs humains ou avec d'autres SI pour « transformer des données brutes, hétérogènes, [...] en une information visible et compréhensible » (Flichy 2013, p.59). Il s'agit de manipuler facilement des données, c'est-à-dire pouvoir en ajouter, les modifier, les analyser, les synthétiser, les visualiser ou encore les partager. Les composantes d'un SI sont multiples et peuvent reposer en partie sur des formes d'organisations sociales pour gérer les flux et traitements de données. D'un point de vue technique, les SI reposent sur des composants logiciels (algorithmes, protocoles et données) et sur des composants matériels (disques durs, processeurs, canaux physiques de communication...).

L'ingénierie des SI fait partie des sciences de conception (Assar 2015). Plutôt que décrire le monde tel qu'il est, ces sciences de l'artificiel visent à imaginer comment il pourrait être et à spécifier les artefacts qui peuvent y amener, en s'appuyant notamment sur l'empirisme et l'itération (Flichy 2013). Les concepteurs doivent procéder à des choix techniques, sur la base de leurs connaissances, de leurs ressources et de leurs intuitions, forcément imparfaites. L'évaluation des artefacts conçus est ainsi indispensable pour les améliorer jusqu'à ce qu'ils répondent de manière satisfaisante aux objectifs qui leur sont assignés (Assar 2015). Le développement d'un SI est un processus long qui mobilise de nombreuses personnes et compétences à l'intersection entre plusieurs disciplines (Flichy 2013). Le travail de programmation ne représente qu'une partie du processus : celle visant à l'implémentation concrète d'un travail de conceptualisation et de formalisation des besoins des utilisateurs et du domaine d'étude.

Selon les fonctionnalités attendues par un SI, ses composantes sociales, logicielles et matérielles peuvent être plus ou moins importantes. Par exemple, dans le cas d'un système d'information pour la gestion d'un réseau ferroviaire, les trois composantes sont primordiales.

En revanche, un système d'information pour la comptabilité d'une petite entreprise peut reposer entièrement sur une approche logicielle sans nécessiter de matériel ou d'organisation en particulier. Une approche uniquement logicielle peut être rapidement dupliquée dans une autre organisation auprès d'autres utilisateurs, tandis que des composantes matérielles ou sociales spécifiques nécessitent plus de temps et de ressources pour être reproduites. Elles sont toutefois nécessaires pour maîtriser le cycle de l'information dans son intégralité (de la réalité à l'action sur la réalité), là où les approches purement logicielles ne peuvent ni mesurer ni agir physiquement sur la réalité. Au regard de nos travaux qui se focalisent sur de l'analyse de données, la composante logicielle d'un système d'information nous paraît suffisante dans un premier temps. Les composantes matérielles et humaines pourront faire l'objet de travaux ultérieurs.

Il existe différentes méthodes de conception de la partie logicielle des SI, une des plus connues étant la méthode française MERISE (Thévenot 2001). Cette méthode pose un cadre méthodologique parfois un peu rigide, mais dont nous nous sommes inspirés dans nos travaux, notamment en séparant le modèle de structuration des données et les fonctionnalités de traitement attendues. À cela s'ajoute une réflexion sur l'interface utilisateur pour la rendre ergonomique et sur la documentation, l'objectif étant de permettre aux utilisateurs d'accéder facilement à l'ensemble des fonctionnalités offertes par le SI (Valentin et al. 2010).

Le logiciel d'un SI s'appuie sur des technologies (langage de programmation, réseaux de communication, système de gestion de base de données...) et des algorithmes qui implémentent diverses fonctionnalités (socle d'applications, manipulation de données, protocole de communication, interface graphique...). Il est souvent possible de mobiliser ces algorithmes sous forme de bibliothèques externes qui peuvent être libres de droits et mises à jour régulièrement par une communauté pour améliorer leur efficacité et fiabilité. Leur utilisation permet un gain de temps considérable dans le développement. Selon les langages de programmation, la disponibilité des bibliothèques peut cependant varier : cela constitue ainsi un des critères du choix du langage à utiliser.

Afin d'accompagner le lecteur peu familier des SI, mais sans toutefois prétendre faire le tour de la thématique, il nous apparaît utile de présenter plus en détail certains éléments que nous allons mobiliser dans la suite de nos travaux. Il s'agit de présenter en particulier les structures de données, la notion d'ontologie et son articulation avec le paradigme de la programmation orientée objet, et les espaces centralisés de données.

2) Structures de données

Les structures de données sont des concepts nécessairement mobilisés pour pouvoir les manipuler. À travers ces structures, il est possible d'agréger des données, de leur donner du sens, de créer des liens entre elles, de les stocker ou de les partager avec d'autres utilisateurs humains ou machines. Elles permettent également l'utilisation d'algorithmes génériques pour mener à bien des analyses. Les algorithmes sont disponibles à travers des logiciels spécialisés, des bibliothèques de programmation, ou encore des systèmes de gestion de bases de données (Bamford & Curran 1991).

On parle parfois de données structurées et de données non structurées. En fait, pour exister numériquement, toute l'information enregistrée doit prendre la forme de données structurées. Lorsqu'on parle de données structurées, nous les entendons concises, explicites et discrètes. Cela s'oppose aux données non structurées comme le texte, le son, les images, les vidéos qui sont décrites par de longues chaînes de données, et surtout qui peuvent faire l'objet d'une interprétation globale¹¹.

Nos travaux nous ont amenés à manipuler des données à travers différentes structures. Pour appuyer notre réflexion dans la conception d'un système d'information, nous avons réalisé un panorama des structures de données disponibles : structures récursives, formats de sérialisation, tableaux, web de données, blockchains et images graphiques.

a) Structures récursives

Les structures récursives organisent les données sous forme de blocs de données (appelés nœuds ou sommets) liés les uns aux autres (Bamford & Curran 1991). Les liens (ou arêtes) entre les nœuds peuvent avoir une orientation et être également associés un bloc de donnée. Les structures récursives les plus simples (Figure 2) sont les listes chaînées qui permettent d'ordonner les données, puis les arbres qui permettent de les hiérarchiser, puis les graphes qui permettent de multiples relations entre les différents nœuds. Entre l'arbre dans lequel les nœuds sont structurés selon une notion de parent/enfant, avec au plus un parent par enfant, et le graphe qui ne limite pas les liens, nous introduisons le graphe hiérarchisé. Ce graphe repose également

¹¹ Pour illustrer cette notion d'*interprétation*, prenons une photographie numérique du nombre 42 : les données codant l'image ne contiennent pas le nombre en lui-même, mais uniquement les pixels de l'image. C'est seulement la lecture et l'interprétation de la photographie par une intelligence qui permet d'y lire le nombre 42.

sur la notion de parent/enfant, mais en autorisant plusieurs parents par enfant si cela n'enfreint pas une hiérarchie logique.

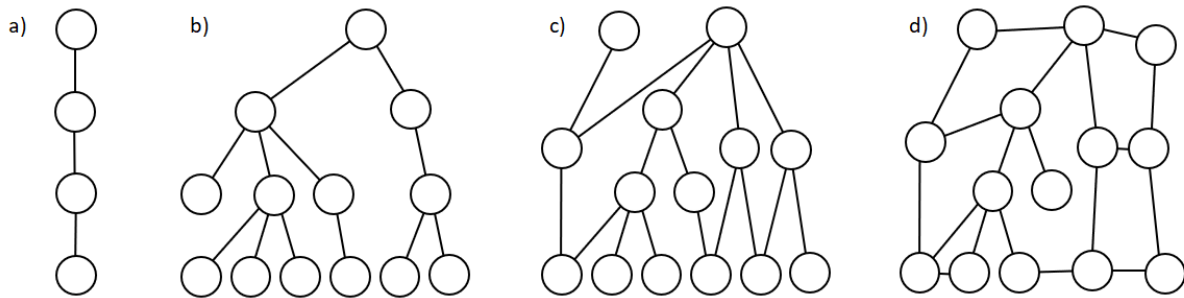


Figure 2 : Structures récursives, a) liste chaînée, b) arbre, c) graphe hiérarchisé, d) graphe

Cette structuration des données a différentes applications (Mehlhorn & Sanders 2008) : la gestion de file d'attente à travers des listes chaînées, la gestion d'arborescences (de fichiers, de classification, du découpage administratif, de généalogie...) à travers les arbres ou des graphes hiérarchisés et l'implémentation d'algorithmes efficaces issue de la théorie des graphes, domaine foisonnant en mathématiques.

La conceptualisation d'une structure récursive est un exercice de base de programmation et de nombreuses bibliothèques implémentent les fonctionnalités nécessaires pour créer et manipuler ce type de structure. En revanche, pour les utilisateurs ne maîtrisant pas la programmation, les logiciels communs ne permettent pas de manipuler l'information de manière générique à travers ce type de structure. Ce sont plutôt des fonctionnalités ponctuelles spécifiques (l'explorateur de fichiers, la hiérarchie d'un menu, des logiciels de graphes) qui permettent à l'utilisateur de naviguer dans une telle structure de manière abstraite, sans vraiment la manipuler en tant que telle.

b) Formats de sérialisation

La sérialisation consiste à organiser des données sous la forme d'une longue chaîne de caractères (sérialisation textuelle) ou de bits (sérialisation binaire). L'ordonnement des informations et l'utilisation de certains caractères particuliers (ou suite de bits) sont définis selon une convention. Cela permet d'extraire un sens complexe de cette longue chaîne et d'en lier de manière cohérente différentes parties. (Sumaray & Makki 2012)

Les formats de sérialisation textuelle sont lisibles par les humains, même si cela demande un peu d'effort et de connaître la convention pour en interpréter le sens. Les données sont séparées et structurées à l'aide de délimiteurs (, / ; - [] <> "retour à la ligne"...) pour former

des lignes et des colonnes comme avec le format CSV, ou via l'usage de balises ("`<balise>Donnée</balise>`", ou "`{Donnée}`") comme pour les formats XML ou JSON. Les trois formats cités sont parmi les plus répandus pour l'échange de données et sont particulièrement présents sur le web (Sumaray & Makki 2012).

Les formats binaires quant à eux forment une suite de bits, ce qui rend leur lecture trop complexe pour un humain, mais ne pose pas de difficulté pour une machine. Ils sont plus adaptés pour structurer des données complexes (images, son, archives compressées...) et en optimiser la taille.

Dans la pratique, tous les fichiers que nous manipulons reposent sur des formats de sérialisation binaires, même si cela est abstrait pour l'utilisateur : ce sont les différents logiciels qui permettent d'en extraire le sens et de le rendre accessible à l'utilisateur humain à travers une interface (graphique, tactile...). L'intérêt des formats textuels est qu'il suffit d'un éditeur de texte basique (par exemple le bloc-notes) pour les lire, ce qui peut permettre à un utilisateur d'avoir une vision directe sur les données. C'est un point utile pour les développeurs qui ont besoin de visualiser ce qu'ils manipulent, mais les utilisateurs classiques ont rarement besoin de manipuler des formats de sérialisation directement.

c) Tableaux

Les tableaux sont un moyen efficace de structurer l'information, à la fois très familière pour les humains, mais aussi pour les machines informatiques dont l'architecture est fondamentalement compatible avec cette approche. Nous pouvons distinguer plusieurs types de tableaux :

- Les tableaux à une dimension de lecture, que nous appellerons liste d'enregistrements, sont composés d'une succession de lignes (plus rarement de colonnes), chaque ligne formant un groupe de données dont le sens est identifié par l'intitulé de colonne.
- Les tableaux à deux dimensions, où chaque case est mise en correspondance avec son intitulé de ligne et de colonne, formant ainsi un triplet de valeur.
- En généralisant, nous pouvons conceptualiser des tableaux à 3,4... n dimensions, formant ainsi des tableaux multidimensionnels (des hypercubes), où chaque donnée est à l'intersection de n intitulés, formant ainsi un (n+1)-plet de données.

Les listes d'enregistrements et les tableaux 2D se manipulent aisément avec un logiciel tableur. Pour les tableaux de dimension supérieure, il faut faire appel à des solutions plus

avancées, par exemple les technologies OLAP (Online Analytical Processing) ou des bibliothèques de programmation. Il est également toujours possible de présenter un tableau à n dimensions sous la forme d'enregistrements avec $n+1$ colonnes¹², à condition de répéter les intitulés de colonnes initiaux dans chacune des lignes d'enregistrement.

d) Tableaux associatifs

Les tableaux associatifs (parfois appelés dictionnaires ou tables d'associations) sont des structures de données où des valeurs sont mises en correspondance avec des mots-clefs¹³, chacun d'eux faisant référence à une et une seule valeur, éventuellement nulle. Les tableaux associatifs sont simples et commodes pour repérer, manipuler et éventuellement ordonner selon les mots-clefs une série de valeurs distinctes. Ils peuvent être le support de techniques de recherche efficaces, grâce à l'utilisation de techniques de hachage ou reposants sur des structures d'arbres équilibrées (Mehlhorn & Sanders 2008). La structure est toutefois conceptuellement limitée : elle ne permet pas de mettre en relation les valeurs entre elles, sauf à définir et interpréter des conventions pour lier les mots-clefs entre eux, par exemple en utilisant des morceaux identiques au début de plusieurs mots-clefs.

e) Tableaux liés : Modèle Entité-Association

Il est possible de structurer des données sous la forme d'un ensemble de tableaux d'enregistrements liés entre eux. Pour concrétiser ce lien, il suffit que les données d'une colonne soient communes à plusieurs tableaux. On parle de clef primaire et de clef étrangère qui permettent respectivement d'identifier de manière unique un enregistrement et d'y faire référence depuis une autre série d'enregistrements. Cette approche est la traduction opérationnelle du modèle conceptuel entité-association formalisé par Chen (1976) et qui a servi de point de départ à la conceptualisation des modèles de données dans nombreuses méthodes comme Merise, UML... (Dessaigne 2005).

De manière pragmatique, le modèle entité-association vise à structurer les données relatives à un domaine à travers un ensemble de tables, chacune visant à rassembler les données d'un type d'entité particulier (appelé également *classe d'objets*) comme des clients, des produits, des villes. En fait, presque n'importe quel nom commun tangible peut constituer un

¹² Un tableau 2D de taille $\ell \times c$ peut se présenter sous la forme de $\ell \times c$ lignes et 3 colonnes.

¹³ Il s'agit plus exactement de « chaîne de caractères »-clef, c'est-à-dire pouvant comprendre des espaces et des caractères spéciaux.

type d'entité. Ces entités sont décrites par des propriétés qui peuvent être de type dit primitif (nombre, date, texte court... nous parlerons d'attributs) ou représenter une association avec un ou plusieurs autres entités du modèle (nous parlerons de relations). En raison de la grande complexité qui peut être conceptualisée à travers cette approche, elle est devenue fondamentale en informatique. Elle est en particulier mobilisée par la forme du diagramme de classe du langage UML (Unified Modeling Language) qui est un langage standard pour la formalisation des SI (Dessaigne 2005).

f) Structures de données basées sur Internet

L'apparition d'Internet a été une révolution dans la façon dont nous partageons l'information. À l'origine, elle n'a pas fondamentalement changé la façon de structurer les données et a simplement permis d'y accéder à distance, ce qui est déjà un progrès conséquent. Toutefois, nous voyons émerger deux structures dont l'essence même est d'être distribuée sur Internet : le web de données et la blockchain (dans un français peu usité « chaîne de blocs »). Bien que reprenant des structures existantes (respectivement le modèle entité-association et les listes chaînées), cette distribution impose des contraintes supplémentaires pour rendre la structure fonctionnelle, mais ouvre dans le même temps des possibilités nouvelles.

❖ Web de données

Le web de données (en anglais *linked data*) a pour objectif de structurer des données sous la forme de bases distribuées et reliées entre elles sur Internet. Pour qu'une base puisse s'intégrer avec les autres, elle doit respecter un certain nombre de conventions définies par le World Wide Web Consortium (W3C)¹⁴. En bref, à travers le web de données, les entités à décrire sont repérées par des identifiants uniques (URI – Uniform Resource Identifier) et leurs caractéristiques prennent la forme d'une liste de triplets <Sujet ; Prédicat ; Objet>¹⁵. Des liens explicites avec d'autres bases sont également souhaités, ainsi que la réutilisation d'ontologies (voir section 3). Enfin, le traitement de l'information s'effectue à l'aide de langages définis comme RDF, OWL ou SPARQL (Berners-Lee 2006). Ces conventions permettent de rendre interopérables les données et leur signification entre différentes bases de façon à permettre aux

¹⁴ Organisme international de standardisation à but non lucratif chargé de promouvoir la compatibilité des technologies du web.

¹⁵ Le Sujet est l'entité décrite (par exemple un pays), le prédicat est sa propriété (par exemple la population), l'objet est la valeur (par exemple 66 millions)

utilisateurs d'abstraire cette multitude de bases et de manipuler des données comme s'il ne s'agissait que d'une seule.

En appui à l'émergence d'un web des données, plusieurs grandes bases de données ouvertes sont devenues compatibles avec cette technologie. Nous pouvons notamment citer les projets DBpedia et Wikidata (voir section 4) qui visent à structurer les données disponibles sur Wikipédia (Ismayilov et al. 2018), ou des sites de données institutionnels comme l'IGN ou le portail des données ouvertes de l'Union européenne¹⁶. Toutefois, si le concept est séduisant, son implémentation reste complexe et demande à mobiliser des compétences et des moyens spécifiques, d'un niveau supérieur à ceux nécessaires à l'implémentation d'une base de données relationnelle (Pauliuk et al. 2019). Ainsi, la dynamique de développement du web de données nous semble encore faible malgré l'existence de plusieurs travaux encourageants (Davis 2017 ; Ismayilov et al. 2018).

❖ Blockchain

Plus souvent désignée sous son nom anglais, même en français, une *blockchain* (« chaîne de blocs ») peut être assimilée à « un très grand cahier [virtuel] que [...] tout le monde puisse lire, sur lequel chacun puisse écrire, mais qui soit impossible à modifier et indestructible » (Delahaye 2015, p.80). L'essence d'une blockchain est ainsi d'être répliquée sur Internet, sur différentes instances qui communiquent et se synchronisent entre elles à travers un protocole sécurisé et prédéfini permettant d'établir un consensus global. La décentralisation des données est gage de confiance : aucun acteur ne peut venir modifier des données validées par consensus et synchronisées sur une multitude d'instances (Waelbroeck 2017).

La première implémentation d'une blockchain remonte à 2008 comme fondement de la monnaie électronique (cryptomonnaie) « Bitcoin » et dont le développement a eu un écho certain (Pavel 2017). En effet, puisque le fonctionnement de cette monnaie repose uniquement sur son protocole décentralisé, elle ne dépend ni des États, ni d'aucun individu en particulier et permet d'assurer l'anonymat des transactions. Nous devinons assez bien que cela ouvre un grand nombre de questions sociétales (juridiques, sécuritaires...) et techniques (consommation d'énergie, système de confiance...). Pour certaines de ces questions, il peut être encore difficile d'apporter des réponses satisfaisantes (Pavel 2017).

¹⁶ IGN : <http://data.ign.fr/endpoint.html> et EU Open Data Portal : <https://data.europa.eu/euodp/en/sparql>

Si historiquement la blockchain a émergé avec la cryptomonnaie, elle n'en est pas la seule application. En fait, l'établissement d'un système de confiance robuste permet d'imaginer de nouvelles opportunités présentant de forts enjeux économiques comme la création d'outils notariaux, le suivi de produits à travers une chaîne d'acteurs, des systèmes de vote... Probablement nous n'en sommes encore qu'au début d'imaginer ce qu'il est possible d'en faire (Delahaye 2015 ; Waelbroeck 2017).

g) Les images graphiques

Les images graphiques sont en quelques sortes à la frontière de notre définition d'une structure de données. En effet, d'un côté, la possible interprétation globale d'une image ne semble a priori pas favorable à la structuration des données. Toutefois, le fait qu'elles soient graphiques, c'est-à-dire qu'elles soient obtenues à partir d'une application ou projection mathématique, permet également de rendre explicite les données à travers l'agencement des formes géométriques. Ainsi, à l'approximation graphique près, il est souvent possible de reconstruire les données originales à partir d'un graphique trouvé dans la littérature¹⁷. Il y a ainsi une forme de structure, même s'il est bien entendu que ces graphiques ne sont pas adaptés pour manipuler des données de manière opérationnelle.

C'est justement ce statut *semi*-structuré des images graphiques, c'est-à-dire avec et sans interprétation possible, qui facilite l'assimilation et la compréhension des données (section II.3) : par la structuration visuelle de chaque donnée, nous pouvons accéder à une interprétation globale de l'information contenue dans ces données. Ainsi, les graphiques sont généralement mobilisés à la fin de la chaîne de traitement des données, lorsqu'il s'agit justement d'interpréter les résultats d'une analyse.

Les images graphiques sont toutefois limitées par la quantité de données qu'elles peuvent représenter : par le nombre de points, mais aussi et surtout par le nombre de dimensions affichables. Ainsi, la majorité des graphiques permettent de représenter 2 ou 3 dimensions simultanément. Avec un peu technique, il est possible d'arriver à 4 dimensions (voir Figure 3), voire 5 en rendant le graphique dynamique¹⁸. Au-delà, il nous paraît difficile d'assurer sa

¹⁷ De notre expérience, c'est même parfois plus rapide (malheureusement approximatif) que de chercher à retrouver les données originales ayant servi à la construction du graphique.

¹⁸ L'outil Gapminder propose dans ses outils un exemple de graphique dynamique représentant les pays du monde en 5D : <https://www.gapminder.org/tools/>

lisibilité. Pour des études complexes, ce nombre maximum de dimensions est souvent dépassé : il n'est donc pas possible de représenter toute l'information en une seule image. Nous reviendrons sur cet aspect plus loin dans nos travaux (Chapitre 3/III.1).

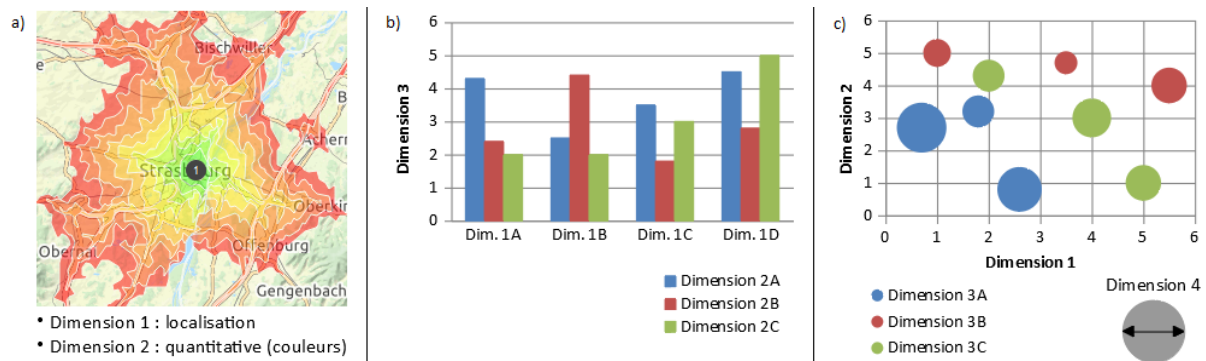


Figure 3 : Différentes dimensions des graphiques.

Exemples de graphiques à deux (a), trois (b) et quatre dimensions (c).

h) Synthèse

Nous avons ainsi identifié différentes structures à notre disposition pour gérer des ensembles de données. Il est très souvent possible de convertir des données d'une structure vers une autre, même si cela n'est pas toujours simple. Le choix d'une structure va dépendre de la complexité de l'information contenue dans les données (voir section suivante) et de l'objectif : s'agit-il de stocker les données, de les manipuler via une interface homme-machine ou via des algorithmes, ou de les représenter ? En guise de synthèse, nous présentons dans le Tableau 2 page suivante l'adéquation ressentie des différentes structures selon l'objectif. Cette synthèse vise ici à formaliser notre expérience et nos intuitions dans une démarche exploratoire ; elle n'est pas le résultat d'une analyse rigoureuse.

Type de structure	Objectif			
	Stockage & extraction	Manipulation via interface	Manipulation via algorithmes	Représentation
Structures récursives	+	×	++	=
Liste d'enregistrements et tableaux à 2 dimensions	++	++	++	++
Tableaux simples à n (> 3) dimensions	++	×	++	×
Tableaux associatifs	++	=	++	+
Tableaux liés (modèle entité-association)	++	=	+	+
Format de sérialisation textuel	+	=	=	=
Format de sérialisation binaire	+	×	=	×
Web de données	++	×	=	×
Blockchain	+	×	=	×
Images graphiques	×	×	×	++

Tableau 2 : Adéquation ressentie des structures de données pour différents usages

(++ : très adéquat ; + : adéquat ; = : peu adéquat ; × : pas adéquat).

3) Ontologie, modèle de données et programmation orientée objet

a) Notion d'ontologie

Le choix d'une structure de données dépend de l'objectif et de la complexité de l'information à structurer. Cette complexité est toutefois particulière à expliciter. Une approche courante pour cela consiste à définir l'*ontologie* de l'information. Selon les disciplines, une ontologie désigne différentes approches, mais avec une notion commune de la recherche du sens et de l'essence des choses. Du point de vue informatique, celui qui nous intéresse principalement pour permettre le traitement de données, une ontologie est la spécification d'une conceptualisation. Il s'agit de formaliser une vue abstraite et simplifiée du monde que l'on veut étudier à travers un nombre restreint de termes pour désigner les concepts mobilisés, leurs

caractéristiques, leurs propriétés, et les relations qu'ils peuvent avoir entre eux, comme les inclusions, les contraires, les voisinages... (Gruber 1993). La définition d'une ontologie permet la mise en place d'un vocabulaire commun utilisable pour partager des informations entre différentes personnes (chercheurs, gestionnaires...) et avec des machines (Noy & McGuinness 2001). Une ontologie est aussi un point de départ pour définir un modèle de données qui peut être implémenté dans un système d'information (voir section suivante).

La pertinence d'une ontologie repose sur plusieurs points (Gruber 1993) : la clarté des concepts utilisés qui doivent être aussi objectifs que possible et notamment indépendants d'un contexte ; la cohérence de l'ontologie qui ne doit pas permettre d'inférer des raisonnements contradictoires ; l'extensibilité de l'ontologie et la possibilité d'ajouter de nouveaux concepts sans avoir à toucher aux fondements ; une déformation d'encodage minimale, c'est-à-dire qui ne nécessite pas de modifier le sens initial d'un concept pour qu'il puisse s'inscrire dans une ontologie ; et un engagement ontologique minimal qui nécessite de peser l'usage de chaque terme pour limiter leur nombre (et donc la complexité), tout en permettant de décrire correctement ce que l'on souhaite étudier.

La définition d'une ontologie est souvent le fruit d'un processus itératif dans lequel le vocabulaire est supposé, testé, validé, amélioré... Dans leur guide, Noy & McGuinness (2001) dressent différents points à aborder dans ce processus. Il s'agit notamment de déterminer les concepts clefs et leurs terminologies qui sont mobilisés dans le domaine d'application pour lequel on veut définir une ontologie. Il faut ensuite définir les liens entre ces concepts, leurs propriétés et les règles qui régissent ces liens et propriétés pour garantir l'intégrité des données (par exemple des règles hiérarchiques ou d'exclusion). La réutilisation d'ontologies existantes permet également de renforcer la pertinence et le caractère partagé du vocabulaire utilisé.

Une ontologie a vocation à être générique. Aussi, elle peut être décrite dans un langage naturel et vise plutôt un objectif de description générale. Elle n'est pas destinée à être implémentée telle quelle dans un SI (Spyns et al. 2002). Il existe toutefois, différents standards, comme le OWL (Web Ontology Language) mobilisé notamment à travers le web de données, qui permettent d'explicitier formellement le vocabulaire, les propriétés et les règles logiques entre les différentes notions (Fléty & De Sède-Marceau 2009). L'usage de tels standards est toutefois assez complexe.

b) Modèles de données

Une approche plus classique pour implémenter une ontologie dans un système d'information est d'en dériver un modèle de données. On peut considérer trois niveaux de modèles de données, même si, dans la pratique, ils ne sont pas toujours explicitement distingués ou décrits comme tels. Du plus général au plus précis, nous avons : le Modèle Conceptuel de Données (MCD), le Modèle Logique (MLD) et le Modèle Physique (MPD).

Le MCD, en raison de son haut niveau d'abstraction est finalement très proche de la définition d'une ontologie (Gruber 1993). La distinction entre une ontologie, surtout lorsqu'elle est formalisée à travers un standard, et un MCD n'est d'ailleurs pas évidente, tous deux s'intéressant à décrire la sémantique d'un domaine. La différence principale relève de l'intention : l'ontologie est plutôt orientée sur la recherche d'un consensus pour partager de l'information entre différents utilisateurs, tandis que le MCD est une première étape prescriptive pour construire la BDD d'un système d'information afin de respecter un cahier des charges (Spyns et al. 2002 ; Pierra et al. 2005).

À partir du MCD, il est possible de construire le MLD qui vise à expliciter un modèle entité-association à travers une structure en tableaux liés. Enfin, le dernier modèle, le MPD, précise en plus le typage des données des différents tableaux (nombre, texte...) et les contraintes pour construire l'algorithme qui va pouvoir être interprété par un Système de Gestion de Bases de Données (SGBD) pour rendre ainsi exploitable la structure des données.

La mise en œuvre d'un modèle entité-association dans une base de données, la faisant ainsi devenir *relationnelle* (BDDR) est une approche standard dans l'industrie, mature, dominant le marché du stockage de l'information et disponible dans pratiquement tous les SGBD (Inmon 2005 ; Dessaigne 2005). Ces derniers offrent les fonctionnalités pour accéder, ajouter, modifier ou supprimer des données dans la base par l'intermédiaire de requêtes écrites dans un langage spécifique, souvent du SQL (Structured Query Language). L'exécution de ces requêtes est très efficace et permet de manipuler de grands ensembles de données (Dessaigne 2005).

c) Programmation orientée objet

La mise en œuvre d'une base de données relationnelle est donc une phase habituelle dans le déploiement des systèmes d'information. Cela permet de stocker, modifier et extraire de grandes quantités de données au sein d'un modèle qui peut être complexe (Dessaigne 2005).

Toutefois, à ce stade de notre description, les interactions entre des algorithmes de traitement et une base de données se limitent à la construction de requêtes qui vont d'extraire des données spécifiques des différentes tables, individuellement ou sous forme de listes ou de tableaux. Ces requêtes sont parfois complexes à construire, surtout lorsqu'elles interrogent de multiples liens entre les tables. C'est une limite de cette structure en tableaux liés.

Le développement de l'informatique a fait émerger un paradigme pour d'éviter de devoir manipuler des données via un ensemble de tableaux : la programmation orientée objet. Ce paradigme permet la manipulation des données sous forme de paquets appelés *objets* (O'Neil 2008). Ces objets de données disposent de propriétés qui peuvent être des attributs primitifs ou des liens avec d'autres objets. Cette approche permet d'abstraire la complexité des jointures et des héritages entre les différentes tables du modèle de données, elle offre un accès direct aux objets à travers leurs liens et donne la possibilité d'associer des fonctionnalités spécifiques à ces objets (des *méthodes*).

La programmation orientée objet offre ainsi un cadre plus intuitif pour le développement d'algorithmes et réduit la complexité conceptuelle de la manipulation de l'information. Elle permet de contourner différentes difficultés (Fonseca et al. 2002) : le partitionnement horizontal ou vertical des tableaux de données, les changements de vue, c'est-à-dire la façon d'afficher les données, le processus de zoom qui doit être discrétisé pour correspondre à différents niveaux de résolution. De plus, le paradigme orienté objet est fondamentalement compatible avec le modèle entité-association et ne demande donc pas un travail supplémentaire de formalisation pour être implémenté. Les types d'entités d'un MEA ou les concepts d'une ontologie deviennent ainsi simplement ce qu'on appelle des *classes* d'objets caractérisés par leurs propriétés, c'est-à-dire des attributs primitifs (comme un nombre, une chaîne de caractères...) ou des liens avec d'autres objets.

La POO et les BDDR restent cependant deux approches techniques séparées. Au sein d'un SI peuvent donc cohabiter un modèle objet et un modèle de données distincts, charge aux concepteurs d'implémenter la logique des liens entre les deux modèles. Pourtant, au regard de leur compatibilité, il semble pertinent de vouloir manipuler une BDDR à travers la POO sur la base d'un modèle commun. Pour cela, il est possible de mobiliser les techniques de *mise en correspondance objet-relationnel* (en anglais Object-Relational Mapping - ORM) qui sont précisément destinées à cela. Elles prennent souvent la forme d'une librairie implémentant l'abstraction qui permet la synchronisation entre les objets manipulés par les algorithmes et les données de la base. Cette solution nous permet de bénéficier à la fois des techniques de stockage

performantes des BDDR et de l'intuitivité offerte par la POO pour le développement d'algorithmes. Il n'est toutefois pas possible d'ignorer complètement la gestion d'une BDD : la construction des requêtes pour sélectionner les données de manière efficace ne peut pas être totalement abstraite. L'usage d'un ORM pose aussi des problématiques spécifiques (O'Neil 2008) : à quel moment synchroniser les données, comment gérer le multi-accès, quel impact de l'abstraction sur la vitesse de traitement... Toutefois, au stade de nos travaux, ces problématiques techniques ne sont pas particulièrement significatives au point d'être l'objet d'une attention particulière.

4) Espaces centralisés de données

a) Entrepôts de données

Le développement des Systèmes de Gestion de Bases de Données (SGBD) à partir des années 1970 a permis de faciliter et d'accélérer les traitements de données au sein des organisations, en permettant de rassembler en un lieu unique des données autrefois éparpillées dans différents fichiers (Inmon 2005). Toutefois, les capacités de calcul augmentant, la quantité de données manipulées a fait de même, menant à la multiplication des bases de données au sein des organisations. Cette multiplication, souvent menée de manière itérative sans vision à long terme, manque parfois de crédibilité. Ainsi, Inmon (2005) met en avant les problématiques engendrées par cette multiplication « naturelle » des bases de données : existence de données similaires, multiples définitions des périmètres d'étude, perte d'information à la suite de l'intégration de données extérieures, variabilité dans l'implémentation de méthodes de traitements... Tous ces facteurs peuvent amener des analyses similaires à des conclusions différentes, rendant ainsi instable la prise de décision à partir des résultats de ces analyses.

Pour rendre plus cohérentes les analyses entre elles et fiabiliser le processus de prise de décisions, Inmon a proposé en 1991 la mise en place d'*Entrepôts de Données* (EDD ; en anglais, *data warehouse*). Le principe est assez simple : il s'agit d'une base de données centralisée qui rassemble une synthèse transversale et standardisée des données situées dans les différentes bases opérationnelles existantes dans une organisation, sans toutefois les remplacer. À la différence d'une BDD classique, les données stockées dans un EDD n'ont pas vocation à être modifiées ou supprimées, éventuellement elles peuvent être synthétisées à nouveau pour l'archivage. L'EDD constitue ainsi une référence unique dans l'organisation sur laquelle peuvent se baser les analyses pour répondre à des interrogations transversales. Cela simplifie grandement la recherche de données, réduit le risque d'incohérences et facilite la réitération des

analyses dans d'autres contextes (temporels, géographiques...). En d'autres termes, cela permet une réduction des coûts liés à ces opérations.

L'implémentation d'un EDD repose sur la définition d'un modèle de données pour les différentes typologies d'information que l'on souhaite stocker (les *sujets*). Elle nécessite de définir également le niveau de détail que l'on souhaite conserver lors de l'agrégation de l'information (la *granularité*), ainsi que les standards à utiliser, par exemple pour les identifiants, les dates, les localisations géographiques... Un EDD est souvent implémenté de manière itérative, en commençant à le rendre opérationnel pour un sujet en particulier, puis en l'étoffant avec d'autres sujets. À partir d'un certain moment, la taille et l'accès à l'EDD deviennent problématiques. Il s'en suit alors généralement la création de *magasins de données* qui copient une partie restreinte des données de l'EDD selon un domaine métier particulier (finance, suivi client...) afin d'en faciliter l'accès et l'usage pour le travail d'analyse.

L'alimentation en données d'un EDD est réalisée à travers des *extracto-chargeurs*. Ces modules au nom particulier visent à : extraire des données de sources hétérogènes, généralement une base de données ou un tableur, de préparer ces données (voir section II.3), puis de les charger dans l'EDD. En anglais, on parle de modules *Extract / Transform / Load* (ETL). Ce sont des éléments cruciaux dans un EDD. Ils contribuent à assurer l'intégrité des données, ils doivent s'assurer qu'elles ne soient pas chargées plusieurs fois et que leur format et leur granularité définis sont bien respectés. L'implémentation des modules ETL peut représenter 80 % du temps de développement de l'EDD (Thomsen & Bach Pedersen 2009).

Dans les cas les plus simples, les données sources prennent la forme d'enregistrements, et l'opération à mener consiste en une *mise en correspondance* (en anglais *mapping*) des différents champs (ou noms de colonnes) avec le modèle de données de l'EDD. Si les champs sont parfaitement équivalents, ce travail de mise en correspondance peut être facilement réalisé avec une interface graphique (*Graphic User Interface – GUI*). Dans les cas les plus complexes, les données doivent être préparées pour correspondre au modèle de données avant de pouvoir être chargées dans l'EDD. L'usage d'une GUI peut alors ne pas être adapté pour répondre à la spécificité de toutes les opérations de préparation à mener. Le plus simple est souvent de mettre au point les ETL en développant directement les algorithmes adaptés, ouvrant ainsi la possibilité d'implémenter n'importe quelle approche tant qu'elle est conceptuellement formalisée (Assar 2015). De toute manière, s'il doit y avoir préparation des données, cette tâche est souvent confiée à des personnes qui ont des compétences en termes de développement informatique et qui peuvent donc se passer d'une GUI (Thomsen & Bach Pedersen 2009).

b) DBpedia et Wikidata

Dans la même philosophie de vouloir regrouper les données en un point central, les projets DBpedia et Wikidata sont particulièrement remarquables. Ces deux projets s'appuient sur la réputée encyclopédie collaborative Wikipédia pour la compléter d'une base des données qu'elle contient. Les dynamiques et les choix techniques de ces deux projets diffèrent, expliquant leur existence parallèle (Ismayilov et al. 2018).

Dans nos travaux, nous nous focalisons surtout sur les modèles de données utilisés. DBpedia repose sur l'usage strict des techniques du web de données où chaque sujet d'étude est repéré de manière unique (par un URI) et caractérisé à travers une liste de triplets <Sujet ; Prédicat ; Objet>. Wikidata en revanche, s'il est également compatible avec les techniques du web de données, présente un modèle de données qui nous semble plus facilement utilisable pour décrire la réalité¹⁹. Dans ce modèle, en plus d'avoir chaque sujet d'étude (*item*) repéré par un URI, ceux-ci sont ainsi caractérisés par une liste d'affirmations (*statements*). Une affirmation regroupe une propriété (*property*), une valeur (*value*) et des qualifications supplémentaires de la valeur (*qualifiers*) comme sa source ou sa temporalité.

Ces projets sont intéressants à examiner à la fois pour les aspects techniques implémentés (web de données, lien avec Wikipédia, mise en place d'API...), mais aussi pour le modèle de données qui doit être pensé suffisamment générique pour assurer l'ambition du projet, c'est-à-dire centraliser une très grande variété de données sur notre monde.

IV. Les enjeux sociétaux de l'utilisation de données

La précédente section présente différentes techniques, essentiellement informatiques, qui permettent de faciliter le traitement des données. Ce faisant, ces techniques et de nombreuses autres, que nous ne présenterons pas ici, ont permis aux données d'occuper une place omniprésente dans nos sociétés modernes, avec un volume de données produites et traitées qui continue de croître à rythme important. Toutefois, les données ne restent qu'un outil de représentation de la réalité, et comme beaucoup d'outils, l'usage qu'il en est fait peut ne pas être que positif.

Nous avons fait le choix de faire de la donnée notre objet de travail et il nous semble nécessaire d'avoir un regard critique sur les implications sociétales de l'utilisation croissante

¹⁹ Le modèle de données est consultable sur : <https://www.mediawiki.org/wiki/Wikibase/DataModel/Primer>

de données. À travers cette partie, nous voulons souligner les bénéfices tout autant que les risques que cela présente. En effet, si les données ont toujours tenu un rôle important pour structurer les organisations, l'explosion de leurs usages notamment grâce aux progrès de l'informatique vient exacerber les enjeux qui y sont liés, et en fait apparaître de nouveaux.

1) De nouvelles possibilités pour la recherche scientifique

Les données sont fondamentales dans la recherche scientifique : elles permettent de confronter à la réalité des hypothèses, des modèles et ainsi de construire des théories scientifiques (Kitchin & Tate 2000). Les progrès techniques ont même permis à un « quatrième paradigme de la recherche » d'émerger : celui de l'exploration de données à travers différents points de vue (Hey et al. 2009). Ce nouveau paradigme, qui permet l'émergence de nouvelles découvertes, vient s'ajouter aux autres paradigmes : (1) l'observation et la pratique empirique, (2) l'élaboration et l'usage de théories, (3) la simulation, notamment à travers des outils informatiques.

Ce paradigme ne remplace pas les précédents, il vient offrir de nouvelles possibilités scientifiques tout en ayant ses propres limites. Il a en tout cas permis de réaliser d'importantes découvertes scientifiques dans tous les domaines (Kitchin & Tate 2000 ; Hey et al. 2009). Le panel des techniques disponibles s'enrichit continuellement avec de nouvelles approches de plus en plus sophistiquées et permettant de traiter des données toujours plus nombreuses, diverses et précises, avec des résultats si impressionnants que certains algorithmes, comme l'apprentissage profond sont même qualifiés d'*intelligence* artificielle. Pour certains, ces algorithmes changent la façon de faire de la science en permettant l'émergence de modèles construits sur les données et qui n'ont plus besoin d'une conceptualisation formelle humaine (Charron et al. 2018).

Malgré ces perspectives prometteuses, nous devons toutefois être conscients des limites des approches basées sur les données. Ainsi, les découvertes liées à l'exploration de données relèvent d'une part de hasard (Hey et al. 2009) : en observant différents jeux de données, avec différentes représentations, certains phénomènes peuvent apparaître spontanément sans qu'ils aient été spécifiquement recherchés. Il est possible de multiplier les analyses de données et de manière automatisée afin d'identifier de manière plus systématique des liens entre les données. Mais cela conduit à une deuxième limite : corrélation n'étant pas causalité, il est tout à fait possible de trouver des variables corrélées mais qui n'ont aucun lien réel entre elles (Ibekwe-Sanjuan 2014). Confier uniquement à des algorithmes la recherche de corrélation pour

expliquer des phénomènes peut donc donner des résultats de qualité aléatoire. Enfin, une troisième limite plus philosophique peut être formulée : l'importante quantité de données peut donner l'impression qu'il est possible d'atteindre une connaissance parfaite de la réalité, parfois simplement en allant plus loin dans la collecte de données et les techniques d'analyse. Or, la réalité est définitivement plus complexe, et même si la connaissance s'améliore, les données collectées ne représenteront toujours qu'une partie réduite, simplifiée et biaisée de la réalité (Audard et al. 2014 ; Charron et al. 2018). Certains auteurs mentionnent d'ailleurs parfois la faible qualité des données, qui ne peut pas être compensée par son volume. En fait, la problématique vient surtout du fait que les données sont souvent produites à travers un service privé à but commercial et qui n'a rien à voir avec la volonté d'une meilleure compréhension scientifique des phénomènes, en particulier sociaux (Charron et al. 2018).

Aussi, la massification des données vient bousculer les habitudes de travail des chercheurs. Entre résultats avérés et perspectives qui semblent parfois utopiques, il est difficile de savoir à quel point être enthousiaste. Mais cela nous permet au moins de placer la donnée comme un véritable sujet d'étude. Le débat existe d'ailleurs pour savoir s'il n'existerait pas une « science des données » qui serait une discipline à part entière (Press 2013). En tout cas, si elle existe, son objectif est clair : développer les méthodes permettant de maximiser l'information utile pouvant être extraite de données complexes ou volumineuses (Cleveland 2001 ; Schutt & O'Neil 2013). Cela mobilise des techniques à la croisée entre l'informatique et les statistiques, comporte des contraintes juridiques et pose de nouvelles questions d'ordre philosophique, politique, économique, ou plus largement sociétal (Besse & Laurent 2016).

2) Une évolution du monde économique

Si le débat existe encore dans la communauté académique sur l'existence et la nature d'une science de données, celui-ci semble tranché dans le monde économique, les entreprises recrutant littéralement des « data scientists »²⁰ (Besse & Laurent 2016). En effet, les données jouent un grand rôle dans le développement économique et la montée en puissance du numérique en a fait une véritable matière d'œuvre à partir de laquelle il est possible de créer de la valeur ajoutée ; pour certains, il s'agit du « nouvel or noir du XXIe siècle » (Lemoine 2018). Pour s'en rendre compte, il suffit de regarder le succès économique des géants du numérique²¹

²⁰ Souvent en français dans le texte, la traduction « scientifique des données » n'étant pas toujours d'usage.

²¹ Qui sont souvent appelés les « GAFAM » pour Google, Apple, Facebook, Amazon et Microsoft

dont le modèle économique repose en grande partie sur leur capacité à exploiter toujours plus d'information pour créer de nouveaux services diversifiés et personnalisés.

Cet essor économique n'est pas sans poser de problèmes sociétaux. Ainsi, ces nouveaux services reposent souvent sur l'analyse de données personnelles. Dans les faits, ces données sont souvent collectées à l'insu de la personne concernée, on parle parfois de données d'échappement ou *data exhaust* (Charron et al. 2018)²². Ensuite, les sociétés (du) numériques ont tendance à accumuler plus de données sur leurs utilisateurs qu'elles n'en ont réellement besoin pour le fonctionnement de leurs services (Lemoine 2018). Potentiellement, elles peuvent donc connaître une grande partie de notre vie privée. Ces données personnelles peuvent aussi être détournées de leur usage initial, piratées, rendues publiques, ou communiquées à des tiers comme des organismes de surveillance, avec les risques de dérives que cela peut engendrer, et sans que la personne concernée ne puisse le contrôler (Besse & Laurent 2016).

En plus des risques sur la vie privée des individus, la fuite des données d'une entreprise peut mettre en difficulté celle-ci. Son image peut être impactée par son manque de fiabilité, ses clients peuvent subir des préjudices, et des irrégularités ou des avantages compétitifs peuvent être révélés (Delzangles et al. 2018 ; Lemoine 2018). Les fuites de données peuvent également mettre en péril des sites sensibles s'ils venaient à être connus par le public. Cela peut tout aussi bien concerner des écosystèmes naturels particuliers que des sites stratégiques relevant de la défense nationale. Cette fuite de données peut être d'autant plus difficile à prévenir que le stockage et le traitement de données sont de plus en plus externalisés à travers le *cloud computing*. Face à ces risques, le domaine de la cybersécurité est stratégique pour protéger les intérêts des entreprises, mais aussi ceux des États (Bwele 2013).

Malgré ces problématiques qui émergent par la massification des données, ces dernières continuent de jouer un rôle fondamental dans le développement économique. Cela est vrai pour les entreprises, mais aussi pour les acteurs publics en charge de mettre en œuvre les politiques territoriales.

²² Par exemple le ciblage publicitaire lors de votre navigation sur Internet ne repose pas sur des informations que vous avez explicitement transmis dans cette intention.

3) L'appui des politiques publiques

Les possibilités offertes par les données pour la gestion des territoires et des politiques publiques font des États, et de manière générale des acteurs publics, d'importants producteurs et utilisateurs de données (Fléty & De Sède-Marceau 2009 ; Mazaud et al. 2017). Celles-ci vont en effet permettre de définir des objectifs et de suivre leur bonne atteinte, de mieux connaître le territoire, ses acteurs, ses dynamiques, ou encore de mettre en place des solutions innovantes qui améliorent le fonctionnement sociétal. Pour certains, il s'agit de parvenir à des territoires « intelligents », comme les « smart cities » (Shahrokni et al. 2015), dans lesquels les Technologies de l'Information et de la Communication (TIC) permettent par exemple une gestion optimale des flux, un accompagnement au changement des habitudes de consommation et de mobilités, ou une aide personnalisée pour les démarches administratives...

Malgré ces possibilités intéressantes, les données jouent également un rôle plus critique dans la gouvernance des territoires et des États (Desrosières 2008 ; Alphandéry et al. 2012). Ainsi, l'apparition et l'évolution des approches comptables nationales et des statistiques publiques ont contribué à justifier l'existence d'institutions qui permettent l'accès à des données utiles au développement de nos sociétés. L'État devient ainsi un producteur de données référent pour son propre fonctionnement, mais aussi celui d'autres acteurs. Mais plus que produire, il normalise. Par le choix des statistiques récoltées et les conventions de classification, il définit le cadre de lecture de l'information publique. Indirectement, il est ainsi possible de minimiser certaines thématiques en restreignant l'information disponible (et justifier leur inaction au prétexte d'un manque de données), ou au contraire profiter de l'abondance d'informations pour d'autres thématiques pour chercher spécifiquement celles qui appuient leurs politiques. Loin d'être objectives, les données, y compris celles issues des institutions publiques, sont bel et bien une construction sociale sur laquelle il convient de prendre du recul. Un usage excessif a d'ailleurs également pour effet d'accentuer le « découplage total entre le centre de recherche et la réalité de terrain » (Charron et al. 2018, p21), alors que la bonne compréhension des dynamiques sociétales requiert une connaissance fine des acteurs, connaissance qu'il est souvent difficile d'exprimer sous forme de données (O'Neil 2016). En plus de ces éléments, l'intensification du recours aux données dans l'action publique peut exacerber la fracture numérique. Les personnes n'ayant pas accès aux TIC, et les territoires ne disposant pas des infrastructures ou des compétences numériques se retrouvent ainsi marginalisés dans leur développement (Charron et al. 2018).

Pour faire face à toute la complexité des enjeux, les pouvoirs publics tentent d'encadrer l'usage de données à travers la réglementation. Les intentions annoncées sont de protéger la vie privée et les intérêts économiques tout en favorisant l'ouverture des données non sensibles pour favoriser l'innovation et ouvrir de nouvelles opportunités économiques, politiques, scientifiques ou écologiques (Audard et al. 2014). Ainsi, la première loi établissant le secret statistique en 1951 vise justement à protéger les enquêtés pour produire des agrégats statistiques sans en dévoiler les composantes individuelles (Gaeremynck 2009). D'autres textes viendront par la suite compléter ou modifier cette loi. Récemment, les lois sur le secret des affaires²³ et le règlement général sur la protection des données (RGPD)²⁴ au niveau européen visent à accentuer la protection des données commerciales et personnelles. En parallèle, l'ouverture des données non sensibles des organisations publiques a également été inscrite à travers différentes lois favorisant leur mise à disposition sur Internet²⁵. Il existe également Etalab en France, une mission interministérielle créée en 2011 et chargée de coordonner les politiques publiques autour des données. Ces politiques se manifestent entre autres par l'émergence de plateformes de données ouvertes²⁶ facilitant l'accès à des données diverses issues des administrations (Mabi 2015 ; Paquienséguy & Dymytriva 2018).

4) Un outil pour le développement durable ?

Les données peuvent participer à la construction de réponses adaptées face à des problématiques complexes, parmi lesquelles figure la protection de l'environnement. Dans cet objectif, une bonne gestion des données et de l'information peut permettre de mener des évaluations environnementales, de gérer plus efficacement les flux, les processus et les territoires, de participer à des réunions à distance, de télétravailler ou encore d'aller vers une dématérialisation de l'économie (Audard et al., 2014 ; Flipo et al. 2016).

²³ Loi n° 2018-670 du 30 juillet 2018 relative à la protection du secret des affaires.

²⁴ Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE (règlement général sur la protection des données).

²⁵ Notamment la LOI n° 2016-1321 du 7 octobre 2016 pour une République numérique.

²⁶ Par exemple la plateforme gouvernementale (<https://www.data.gouv.fr/>) ou celles plus locales comme le géoportail du Grand Est (<https://www.geograndest.fr>) ou les portails de grandes villes (Nantes - <https://data.strasbourg.eu> ou Strasbourg - <https://data.strasbourg.eu>)

Toutefois, de manière antinomique, les moyens nécessaires au traitement des données (les TIC) contribuent à une consommation croissante d'énergie et de ressource pour leur fabrication et leur fonctionnement (Diguët et al. 2019). On estime que les TIC représentent 5 à 10 % de la consommation mondiale d'électricité (près de 15 % en France), et mobilisent des quantités importantes de métaux (Flipo et al. 2016) : or (12% de la demande mondiale), cuivre (30%), argent (30%), ruthénium (80%), indium (80%)... À cela s'ajoutent d'autres problématiques comme l'emprise sur le territoire des infrastructures, notamment les data centers, ou les nuisances liées au fonctionnement des installations (Diguët et al. 2019). En fin de vie, les TIC constituent des déchets complexes à traiter, contenant de nombreux éléments toxiques et dont le recyclage est loin d'être systématique (Flipo et al. 2016 ; Rodhain et al. 2017). Ainsi, la consommation de ressources actuelle et à venir imputée aux TIC est devenue trop significative pour être compatible avec un développement durable de nos sociétés (Bihouix 2014).

Malgré les promesses des TIC en termes de dématérialisation, nous constatons plutôt l'inverse : les gains d'efficacité permis par le numérique semblent avoir au contraire permis d'intensifier les échanges et donc la consommation de ressources par effet rebond (Rodhain et al. 2017). Les TIC peuvent ainsi tout autant apporter des solutions à certaines problématiques que contribuer aux impacts environnementaux de notre société. Cela nous interroge notamment sur la pertinence de certaines approches autour des villes intelligentes, censées être plus durables, mais dont le fonctionnement repose sur l'installation d'une multitude de capteurs électroniques communicants. Les gains d'efficacité doivent être comparés aux impacts engendrés par l'équipement supplémentaire nécessaire, c'est tout l'objectif des analyses de cycle de vie dont nous parlons plus loin (Chapitre 2).

5) Vers une « algorithmisation » des prises de décision ?

L'accélération de la puissance de calcul et l'élargissement des possibilités nous fait rentrer dans un processus sociétal extrêmement singulier : celui de « l'algorithmisation » des prises de décisions qui les fait reposer sur le résultat d'un algorithme, sans intervention humaine. Le phénomène avait déjà commencé avec l'automatisation dans l'industrie, le trading haute fréquence dans la finance, ou la sélection d'informations sur les réseaux sociaux. Toutefois, avec les progrès de l'intelligence artificielle, d'autres domaines sont également en train de se transformer en profondeur (Lemoine 2018) : la mobilité et la conduite de véhicules, la médecine, la justice...

Difficile de se prononcer sur le bien-fondé de cette dynamique. Force est de reconnaître que certaines applications présentent un fort potentiel et sont souvent plus fiables que des décisions humaines. Toutefois cela génère en même temps une désagréable impression de perte de contrôle. Thompson (2013) écrit à ce sujet : « By taking human decisionmaking out of the equation, we're slowly stripping away deliberation—moments where we reflect on the morality of our actions »²⁷.

La dynamique d'algorithmisation fait reposer des choix qui peuvent avoir des conséquences humaines sur un triptyque données / algorithmes / objectifs, tous trois parfaitement formels et donc tous trois intrinsèquement biaisés pour décrire une réalité infiniment plus complexe. Nous avons vu que les données dépendent de choix de classification en partie arbitraires et réducteurs, et de mesures comportant des parts d'aléas. Il en est de même pour les choix de programmation qui reflètent en partie la subjectivité de ceux qui les implémentent (Charron et al. 2018). Enfin, les objectifs doivent être mesurables pour être algorithmiquement appréhendés, générant un paradoxe souligné par Loi de Goodhart, formulée dans ses termes par Strathern (1997) : « When a measure becomes a target, it ceases to be a good measure »²⁸. Il est ainsi tout à fait possible de maximiser une mesure en passant à côté d'un objectif global pour lequel la mesure est réalisée et ce ne sont pas les exemples issus du réel qui manquent (O'Neil 2016 ; Muller 2018).

À cela s'ajoutent les risques de défaillance, que ce soit un « bug » du programme, une mauvaise manipulation ou mise à jour, ou une attaque informatique, qui peuvent altérer un ou plusieurs des éléments du triptyque, et donc les décisions qui en découlent. On devine que la question de la responsabilité juridique en cas de défaillance est extrêmement complexe.

Pour toutes ces raisons, il nous paraît ainsi nécessaire de ne pas idéaliser une hyper-présence de l'intelligence artificielle, notamment lorsque ses décisions peuvent avoir des impacts humains. Sans nier des réelles avancées techniques, la banalisation sociétale des algorithmes risque toutefois de rendre les impacts humains de nos décisions encore plus abstraits et de voir ainsi confier notre éthique à une machine qui en est dépourvue. Nous rejoignons ainsi la position de plusieurs auteurs qui préconisent d'implémenter des systèmes

²⁷ « En plaçant la prise de décision humaine hors de l'équation, nous nous sommes lentement débarrassés de ces moments de délibération où nous réfléchissions sur la moralité de nos actions. » Traduction Daniel Péliissier.

²⁸ « Lorsqu'une mesure devient un objectif, elle cesse d'être une bonne mesure. »

d'information transparents et ouverts sur le réel pour le questionner et s'y adapter, et non pas d'imposer au réel de s'inscrire dans les cadres descriptifs qu'ils implémentent (O'Neil 2016).

V. Conclusion du premier chapitre

Dans ce chapitre, nous posons des éléments de cadrage concernant le traitement de données en général pour la suite de nos travaux. Il s'agissait de présenter les approches techniques que nous allons mobiliser : système d'information, structures, modèles et entrepôts de données, ontologie, programmation orientée objet... Ces approches sont efficaces pour le traiter des données et sont à la disposition de ceux qui savent les implémenter.

En dehors des aspects techniques, nous plaçons notre réflexion dans un cycle de l'information. Ce cycle décrit la production de données à partir de mesures de la réalité, pour en extraire de l'information et générer de la connaissance pour rétroagir sur la réalité. Notre démarche reste modeste dans ce cycle, puisque nous ne nous focalisons que sur l'une des étapes du cycle : celle de l'analyse, et même plus précisément sur le traitement de données qui la soutient. En séquençant ce traitement de données, nous avons souhaité le rendre plus opérationnel en invitant à considérer plus attentivement différents éléments. Il s'agit notamment de rendre les données plus utilisables, mais aussi de souligner l'importance du travail de préparation qui est pourtant peu valorisé en tant que tel.

En abordant notre sujet par l'angle des données et de leur traitement, nous rejoignons l'objectif de la science des données, si tant est que le terme puisse faire consensus. Notre approche est en effet de faire des données une matière première et d'en extraire un maximum d'informations utiles, notamment en rendant leur exploration plus aisée. Nous ne cherchons pas à répondre à une problématique particulière définie a priori, nous nous basons plutôt sur quelques questionnements simples qui vont nous guider dans l'exploration de données, avec l'objectif d'identifier des phénomènes et de construire une vision globale, synthétique et aussi pragmatique de l'information contenue dans les données. Cette position, complètement ancrée sur la donnée, est évidemment critiquable, et nous la critiquons d'ailleurs : accorder une confiance excessive à ces données peut avoir des conséquences sociétales négatives, surtout lorsqu'il s'agit de décrire des concepts sociétaux par nature plus difficiles à objectiver. Malgré tout, cette approche a aussi fait ses preuves dans différents domaines et, à travers cette thèse, nous voulons la mobiliser pour l'analyse des flux de matières et d'énergie à l'échelle d'un territoire, autrement désignée comme l'étude du métabolisme du territoire. Nous prêterons toutefois attention à bien identifier autant les apports que les limites de cette approche.

Chapitre 2 Fondements et applications d'une approche par l'analyse des flux physiques et le métabolisme territorial

Ce chapitre a été publié comme article scientifique dans la revue Vertigo - la revue électronique en sciences de l'environnement : « Fondements des analyses de flux de matières et d'énergie et typologies d'applications pour la gouvernance des territoires et des organisations » par B. Ribon, D. Badariotti, R. Kahn. Vertigo, Volume 18 n° 2 | Septembre 2018, URL : <http://journals.openedition.org/vertigo/20822> ; DOI : 10.4000/vertigo.20822. Nous le présentons dans une version légèrement modifiée pour mieux l'ancrer dans cette thèse.

I. Introduction

L'analyse de flux de matières et d'énergie (AFME) est un outil d'aide à la compréhension des systèmes qui s'appuie sur une comptabilité des flux en entrées et sorties et sur l'application des principes de conservation de la masse et de l'énergie. Fruit d'un héritage historique comptable et scientifique, les AFME se sont révélées indispensables au développement industriel. Elles figurent également aujourd'hui parmi les meilleurs outils disponibles pour évaluer les impacts environnementaux des organisations (entreprises notamment, mais aussi administrations, associations...) et des territoires (villes, régions, pays...).

Dans ce chapitre, et plus globalement dans nos travaux, nous considérons les AFME dans un sens général pour désigner les études qui visent à qualifier et quantifier les flux mobilisés par un système. Cela peut tout aussi bien désigner des études empiriques que celles reposant sur des méthodes plus formalisées et scientifiques. En effet, si les approches diffèrent dans leur mise en œuvre, toutes reposent sur une comptabilisation d'entrées et de sorties de matière. Elles peuvent également se recouper dans leurs applications (études d'impacts environnementaux, optimisation de processus, suivi réglementaire, mise en place d'une économie circulaire...). En considérant l'AFME d'un point de vue général, cela nous permet de prendre un peu de recul et de l'envisager du point de vue des applications pragmatiques qui peuvent en découler pour la

gouvernance des territoires et des organisations. Un parcours de la littérature nous a en effet permis d'identifier différentes typologies d'applications théoriques possibles pour les AFME. Cela montre qu'il s'agit d'un outil finalement très transversal pouvant apporter des informations à différents corps de métier : comptables, juristes, managers, urbanistes, architectes, diplomates, communicants, commerçants, scientifiques, ingénieurs, etc.

Afin d'accompagner le lecteur qui n'est pas familier des AFME, nous débuterons notre propos en présentant les fondements généraux de la démarche, du point de vue historique et scientifique, puis du point de vue méthodologique. Nous verrons notamment, à l'aide de quelques exemples, les spécificités des AFME selon trois types de systèmes : un produit, une organisation ou un territoire.

Ces trois approches peuvent toutefois présenter des usages similaires. Nous avons ainsi recensé différentes typologies d'applications pour les AFME, indépendamment du type de système étudié. L'objectif est de montrer qu'il ne s'agit pas seulement d'un outil environnemental ou technique, mais qu'il permet d'aborder également des sujets économiques, sociaux et sociétaux. En cela, il a un fort potentiel à devenir un outil d'aide à la gouvernance du développement durable des territoires et des organisations.

Néanmoins, il y a encore plusieurs limites sociétales et techniques qui font que les applications présentées ne sont pas forcément opérationnelles. Nous verrons ces limites dans la dernière partie et proposerons quelques pistes à envisager pour les dépasser.

II. Le développement des analyses de flux de matières et d'énergie

Les AFME comprennent une partie comptable des flux en entrées et sorties du système étudié. En ce sens, elles ressemblent à la comptabilité financière par partie double qui décrit les opérations et la situation financière d'une organisation à travers différents comptes. Les AFME visent toutefois à quantifier physiquement les flux par leur masse et non économiquement par leur valeur.

Ces méthodes comptables sont le fruit d'un développement par les civilisations humaines que l'on peut remonter au moins jusqu'au XIX^e siècle av. J.-C, notamment en Mésopotamie. L'extraction, la transformation et l'échange (sous la forme de biens) de matières ont permis de fonder les civilisations qui les ont mises en œuvre. Dès l'Antiquité, comptabiliser la matière a

ainsi permis d'anticiper des flux de valeurs, d'attester de créances, de gérer des stocks, et même plus largement, de gouverner la société (Minaud 2014).

En dehors des aspects comptables, les AFME reposent également sur des découvertes scientifiques qui les ont amenés à se formaliser. Notamment au XVII^e et XVIII^e siècle, on commence à réaliser que toute la matière n'est pas visible (et donc qu'elle peut échapper à la comptabilité), mais que pour autant, elle ne disparaît pas. Ainsi, Santorio (1561–1636), médecin italien, a réussi à peser la masse des produits ingérés par certains de ses patients et celle des sorties (urine et excréments). Il nota que plus de la moitié de la masse ingérée ne ressortait pas (Brunner & Rechberger 2003). C'est Lavoisier (1789), dont les travaux ont permis d'établir la célèbre loi de conservation de la masse « rien ne se perd, rien ne se crée »²⁹, qui donnera l'explication un siècle plus tard. La matière ne disparaît pas, elle ressort sous forme gazeuse et invisible de dioxyde de carbone (CO₂) et de vapeur d'eau lors de la respiration. Au XIX^e siècle, on commence à mieux comprendre la nature de la matière, de l'énergie et du lien entre les deux, notamment avec le potentiel énergétique des matières fossiles comme le pétrole, le gaz et le charbon. Cela a ainsi du sens d'évaluer en parallèle les flux d'énergie et les flux de matières, même s'il est possible de ne s'intéresser qu'à l'un des deux. Dans la suite des travaux, les aspects concernant la matière peuvent souvent être étendus à l'énergie même si cela n'est pas explicitement mentionné. Entre 1830 et 1850, les travaux de plusieurs scientifiques aboutissent à la formulation d'une loi de conservation pour l'énergie, en plus d'autres lois physiques et notamment thermodynamiques (Kuhn 1959). Ces découvertes vont accompagner la révolution industrielle, en particulier le développement de l'industrie chimique et manufacturière. En comptabilisant et en maîtrisant mieux les flux, les industries vont pouvoir mettre au point des procédés tant logistiques que techniques, qui seraient ingérables ou non rentables en l'absence de support comptable (Nikitin 1992).

Au XX^e siècle, la comptabilité va prendre une autre échelle. Alors qu'elle était utilisée de manière libre par les entreprises qui y trouvaient un intérêt, plusieurs États vont lui donner un cadre obligatoire et réglementaire. Ils mettent également en place une comptabilité publique afin d'assurer leur gouvernance et de mettre en place des stratégies de planification, particulièrement utiles dans la reconstruction à l'issue de la Seconde Guerre mondiale

²⁹ Les réactions nucléaires qui transforment de la masse en énergie mettent en défaut la maxime dans son interprétation matérielle. Mais les variations de masses sont extrêmement faibles (~ 40 g/TWh) et localisées, et peuvent être largement négligées dans les propos qui nous intéressent.

(Desrosières 2008). Cette volonté a abouti à la création du premier Système de Comptabilité Nationale (SCN) en 1953 qui a pour objectif de décrire une économie par ses flux monétaires (Eurostat 2013b).

L'émergence des problématiques environnementales à partir des années 1960 a progressivement remis en cause la pertinence de la seule comptabilité économique pour assurer un développement durable des sociétés. Le rapport Meadows (Meadows et al. 1972) est notamment l'un des premiers à avertir des limites écologiques à une croissance économique qui intègre mal la question des externalités. Il peut s'agir des rejets qui ne sont pas comptabilisés, car dépourvus de valeur économique, mais qui sont pourtant en cause dans la dégradation de l'environnement : les émissions de gaz à effet de serre (GES) (IPCC 2014), les déchets mal gérés qui finissent par relâcher leurs composés nocifs dans l'environnement... Les questions relatives à l'épuisement des ressources et à l'efficacité dans leur utilisation sont aussi difficilement abordables sous l'angle purement économique.

Pour aborder ces questions, la Commission mondiale de l'environnement et du développement dans les années 1980, et le programme Action 21 adopté en 1992 par la Conférence des Nations Unies sur l'environnement et le développement ont préconisé la mise en place d'une comptabilité qui tienne compte des éléments écologiques. Cela a abouti à une première version en 1993 d'un Système de Comptabilité Économique et Environnementale (SCEE), qui sera actualisé en 2003 et 2012 par un consortium d'institutions internationales (Nations Unies et al. 2012). Le SCEE offre ainsi un cadre comptable pour agréger et comparer des informations à travers trois types de comptes : les flux physiques (énergie, matière, eau), les actifs (les ressources naturelles disponibles) et l'activité environnementale (activité économique destinée à la protection de l'environnement). La comptabilité des flux physiques, en particulier pour la matière, repose sur les travaux menés sous l'égide du World Resource Institute à la fin des années 1990 (Pasquier 2010) puis adaptés par la suite pour les pays européens par Eurostat (2001 ; 2013a) à travers la méthodologie « Economy-Wide Material Flows Accounts » (EWMFA).

En parallèle de l'évolution de la comptabilité nationale, les AFME connaissent un intérêt grandissant dans le monde académique, en particulier depuis que Wolman (1965) s'est intéressé aux flux d'eau, de nourriture et de combustible fossile d'une ville américaine imaginaire d'un million d'habitants. Il introduisit le terme de « métabolisme urbain », emprunté à la biologie pour désigner l'ensemble des ressources mobilisées par un organisme vivant pour son fonctionnement, et la façon dont elles sont transformées avant d'être rejetées. Selon la

métaphore, l'organisme est une ville, sa nourriture en entrée est constituée par les importations et ce qu'elle extrait sur son territoire, et les sorties ne sont pas que les rejets mais comprennent également l'exportation de produits finis (voir Figure 4).

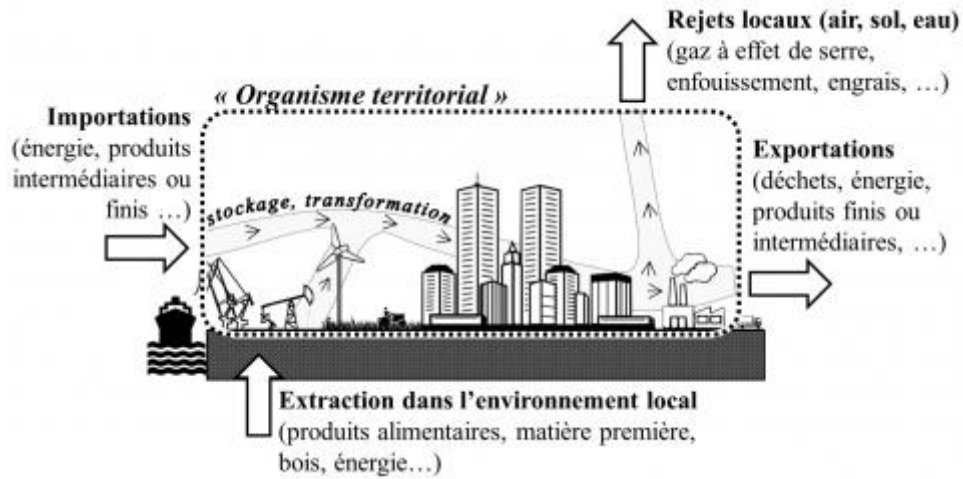


Figure 4 : « L'organisme territorial » et l'étude de son métabolisme

Cette approche par le métabolisme fut aussi féconde que discutée. Féconde, car des études sur de nombreuses autres villes à travers le monde ont suivi et ont permis d'améliorer les méthodes d'évaluation (Zhang 2013). Beloin-Saint-Pierre et al. (2016) recensent ainsi plus de 150 études de métabolisme urbain : de Bruxelles (Duvigneaud & Denaeyer-De Smet 1975) à Lisbonne (Rosado et al. 2014), en passant par Buenos Aires, Sao Paulo (Hoornweg et al. 2012) ou Paris et l'Île-de-France (Barles 2009)... Discutée, car le choix du terme « métabolisme », bien que pouvant être inspirant (Zhang 2013), induit une approche organiciste qu'il faut savoir dépasser (Barles 2010). Certains chercheurs préfèrent ainsi l'image du fonctionnement d'un écosystème (des organismes en interaction) plutôt que d'un « organisme territorial » dont le contour reste flou (Broto et al. 2012). D'autres évitent même les termes d'origine biologique les trouvant trop réducteurs pour décrire la réalité complexe d'un territoire (Veyret & Ciattoni 2011). Cependant, la formalisation du concept de métabolisme territorial l'a amené à s'affranchir de son ancrage biologique. Il désigne ainsi communément l'ensemble des processus techniques et socio-économiques par lesquels un territoire mobilise, consomme et transforme la matière et l'énergie (Kennedy et al. 2007 ; Barles 2009). Le métabolisme est ainsi souvent utilisé comme un terme général pour désigner la dynamique matérielle du territoire, et son usage n'implique pas forcément une approche organiciste.

Le métabolisme d'un territoire dépend de sa taille, de sa densité de population, de son degré de développement, de la présence de plateformes logistiques (ports, gares de fret...) ou

de sa performance écologique (taux de recyclage, d'usage de ressources locales et/ou renouvelables...). Toutefois, une grande constante que l'on retrouve dans la majorité des études est la linéarité de ce métabolisme (Barles 2010) : la matière mobilisée est d'abord extraite de l'environnement, puis transformée en un produit, avant de devenir un déchet à éliminer. Cette matière finit ainsi souvent rejetée dans l'environnement souvent en étant incinérée ou via une mise en décharge. Une meilleure efficacité dans l'usage de la matière demanderait à rendre le processus circulaire en s'assurant, entre autres que les déchets soient retransformés en matières premières. Cette linéarité est encore plus courte dans les territoires les plus urbanisés où peu de moyens de production sont présents. Les produits sont généralement importés sous forme de produits finis et, une fois consommés, ils se retrouvent sous la forme de déchets et sont évacués en dehors du territoire urbain (Ferrão & Fernández 2013). Les villes sont ainsi fortement dépendantes des territoires extérieurs qui sont moins peuplés, mais qui disposent des moyens de production pour les approvisionner et traiter de leurs déchets.

Les méthodes pour analyser les flux de matières se sont ainsi développées à la fois à l'échelle nationale, selon la méthodologie EWMFA, et à la fois à l'échelle des villes selon l'approche par le métabolisme urbain. Entre les deux, les méthodes dédiées aux échelles intermédiaires, celles des territoires³⁰, apparaissent un peu plus tard. Ce sont les travaux de Barles en Île-de-France (2009) complétés par ceux réalisés en Bourgogne (Alterre Bourgogne 2013) et en Midi-Pyrénées (Barles 2014) qui ont permis d'adapter la méthode européenne aux départements et régions français. Finalement, les auteurs des études sus-citées (Repellin et Duret pour la région Bourgogne et Barles) ont formalisé leurs travaux à travers un guide méthodologique réalisé en collaboration avec le Commissariat général au développement durable (CGDD 2014). Ce guide accompagne les territoires souhaitant se lancer dans une démarche similaire.

Ce cadrage historique montre que les AFME héritent de méthodes comptables dont les prémices apparaissent il y a plus de 3 500 ans, mais dont la formalisation par la puissance publique a commencé il y a moins d'un siècle. Elles héritent également des découvertes scientifiques qui ont permis de mieux comprendre la nature de la matière et de l'énergie et d'établir des lois de conservation. L'émergence des problématiques environnementales a rendu cet outil essentiel pour renforcer notre compréhension des interactions entre nos sociétés et

³⁰ Un territoire peut tout aussi bien désigner un pays qu'une ville. Il désigne dans ce cas une échelle intermédiaire (notamment une région ou un département).

l'environnement et pour aborder certaines questions essentielles de durabilité. Dans les parties suivantes, nous allons voir les fondements méthodologiques des AFME et comment elles peuvent être mobilisées pour aborder ces questions de durabilité.

III. La réalisation d'une AFME

Bien qu'étant le fruit d'une longue tradition comptable, l'émergence de méthodes formelles visant à quantifier et analyser les flux de matières à l'échelle des organisations et des territoires remonte à moins d'un demi-siècle. Ces méthodes ont pris plusieurs aspects avec différents objectifs (Loiseau et al. 2012a ; Zhang 2013) par exemple l'empreinte environnementale qui se focalise sur l'étude des impacts sur la biosphère, l'analyse entrées/sorties utilisée dans l'étude et la modélisation des économies ou encore l'analyse du cycle de vie des produits dont nous reparlerons un peu plus loin. L'analyse de flux de matières est ici une méthode parmi les autres, basée sur une analyse comptable des flux.

Toutefois, si ces méthodes diffèrent dans leur mise en œuvre, elles reposent toutes sur une étude de flux et la réalisation d'un bilan entrées/sorties. Elles peuvent également se recouper dans leurs applications (études d'impacts environnementaux, optimisation de processus, suivi règlementaire, mise en place d'une économie circulaire...). Nous avons fait le choix dans notre travail de ne pas considérer les AFME comme une méthodologie spécifique, mais plutôt d'utiliser ce terme pour désigner de façon générale les approches visant à qualifier et quantifier les flux mobilisés par un système. Cette vision nous permet de prendre un peu de recul et de l'examiner du point de vue des applications qui peuvent en découler. Il nous paraît cependant nécessaire, à la fois pour appuyer notre argumentaire, mais aussi pour accompagner le lecteur encore peu familier de l'approche, de présenter les fondements théoriques et pratiques des études de flux avec le support de quelques exemples.

Ainsi, pour réaliser une analyse de flux, il convient de poser avant tout le cadre de l'étude et de définir le système étudié. Les AFME peuvent être réalisées à un niveau micro pour un processus industriel ou une entreprise, ou à un niveau macro pour un territoire comme une ville ou une région (ou au-delà). Il est également possible de réaliser des AFME à niveau transversal lorsque le système étudié est un produit et que l'on s'intéresse aux flux de matières et d'énergie mobilisés tout au long de son cycle de vie. Nous parlerons des spécificités de cet Inventaire du Cycle de Vie (ICV) dans un paragraphe suivant. Ces trois approches (micro, macro, produit) sont différentes possibilités données aux acteurs pour appréhender les enjeux matériels qui les concernent, soit d'un point de vue collectif (macro), soit individuel (micro), soit au sein d'une

chaîne d'acteurs (produit). Bien que ces approches soient liées entre elles³¹ et que leurs développements méthodologiques s'influencent mutuellement (Laner & Rechberger 2016), elles ont des spécificités que nous présenterons un peu plus loin.

Après avoir défini le système, il faut également définir le cadre temporel. La période d'étude peut être plus ou moins large selon les objectifs. Les processus chimiques et les réseaux électriques ont besoin d'être supervisés en temps réel, tandis que les études territoriales se font généralement a posteriori avec un bilan des flux sur une année.

Les AFME peuvent viser l'étude de tous les flux physiques. Néanmoins, pour limiter leur complexité, certaines ne s'intéressent qu'à une partie des flux : les déchets (voir exemple de la MAS paragraphe suivant), voire spécifiquement les déchets d'équipements électriques et électroniques – DEEE (Bahers 2012) ou ceux issus de la construction (Rouvreau et al. 2013). D'autres s'intéressent aux émissions de GES (Ademe 2010), à l'eau³² (Hoekstra & Chapagain 2007), aux ressources alimentaires ou à des substances spécifiques³³ comme l'azote (Le Noé et al. 2016).

La réalisation d'une AFME comprend généralement deux composantes. La première est systématique, il s'agit du bilan (de) matières et (d') énergie qui vise à qualifier et quantifier les flux entrants et sortants du système. La seconde est plus complexe et ne figure pas dans toutes les études³⁴. Elle vise à expliquer comment les entrées deviennent des sorties à travers les différents processus de circulation, de transformation et de stockage au sein du système. Dans les versions les plus formalisées, ces deux composantes permettent de construire les tables (ou matrices) entrées/sorties physiques du système étudié (Giljum & Hubacek 2009).

L'application des principes physiques de conservation de la masse et de l'énergie aide à assurer la cohérence et l'exhaustivité du bilan, en permettant de traquer toute disparition anormale de matière. La connaissance du bilan de matière et d'énergie de certains processus génériques permet aussi d'aider à quantifier les flux. Ainsi, à l'aide du Tableau 3, il est possible

³¹ L'échelle macro est le regroupement des échelles micro, l'échelle produit est une succession d'étapes micro.

³² La masse des flux d'eau mobilisés dépasse généralement d'un ordre de grandeur celle des autres matières. L'eau est ainsi fréquemment étudiée à part pour éviter de masquer – de « noyer » – les autres flux (Fischer-Kowalski et al. 2011).

³³ On parlera d'analyse de flux de substance.

³⁴ La présence de cette seconde phase fait le distinguo entre un « bilan de matière » et une « analyse de flux de matières ».

de déduire qu'une maison individuelle moyenne chauffée au gaz³⁵ mobilise un peu plus d'une tonne de méthane et génère l'émission de près de trois tonnes de dioxyde de carbone dans l'air³⁶. Cette estimation des sorties, qui peuvent être invisibles ou non comptabilisées directement comme le CO₂, à l'aide des entrées, comme le méthane dont les quantités sont connues car facturées, peut être généralisée à beaucoup de rejets et de déchets qui sont rarement quantifiés directement. Cela implique néanmoins une inévitable approximation avec la réalité souvent plus complexe des processus : il pourrait y avoir par exemple une fuite de gaz ou une combustion incomplète faussant ainsi le bilan sans que l'on puisse s'en rendre compte juste en observant le compteur de gaz en entrée³⁷.

	Entrées		Sorties	
	Réactifs	Quantités	Produits	Quantités
Matière	Méthane (CH ₄)	1 kg	Dioxyde de carbone (CO ₂)	2,75 kg
	Dioxygène (O ₂)	4 kg	Eau (H ₂ O)	2,25 kg
Énergie			Chaleur *	15,4 kWh

* : Dans les conditions normales de température et de pression (0 °C et 1,013 bar)

Tableau 3 : Bilan de combustion complète d'un kilogramme de méthane (gaz naturel) (Haynes 2016)

Les éléments précédents montrent les aspects communs aux AFME pour les différentes échelles considérées : micro, macro, produit. Toutefois, il y a aussi des spécificités que nous allons préciser pour chacune de ces échelles.

1) À un niveau micro : AFME des organisations et de leurs processus

Le niveau micro est celui des entreprises et organisations, voire en deçà lorsque l'on considère leurs processus internes. Afin d'appuyer nos propos à cette échelle, nous prenons appui sur le bilan déchets mené à la maison d'arrêt de Strasbourg (MAS). En prévision de la mise en place d'une tarification incitative sur les déchets, la MAS a fait réaliser un bilan de

³⁵ Selon une consommation annuelle de 18 000 kWh (approximation pour une maison de 100 m²).

³⁶ À comparer avec le seuil limite d'émission recommandé par le GIEC à l'horizon 2050 évalué à 2,2 tonnes équivalent CO₂ par personne (Brunetière et al., 2013)

³⁷ Le relevé régulier du compteur peut toutefois mettre en évidence la surconsommation engendrée par une éventuelle fuite.

déchets par un groupe d'étudiants (De Boisséson et al. 2013). Occupée par environ 750 détenus pour une capacité de 444 places (!), l'étude a révélé que MAS génère environ 430 tonnes de déchets par an. Pour arriver à ce chiffre, les étudiants se sont appuyés sur plusieurs méthodes :

- Beaucoup d'entrées ont pu être facilement quantifiées à l'aide des factures, même si la masse des produits ne figure pas systématiquement et que cela nécessite donc une conversion approximative. Ainsi les quantités jetées de bouteilles en plastique (5,6 t), briques alimentaires (4,1 t) ou de barquettes alimentaires usagées (17 t) ont été estimées à partir des achats en entrée.
- Une partie des sorties est également bien connue : il s'agit des déchets traités par des prestataires spécifiques, et non par les services de ramassage municipaux. Ces prestataires indiquent sur leurs factures le poids des déchets pris en charge. En l'occurrence : 29 t de bois, 32 t de cartons, 37 t de déchets industriels banals.
- D'autres matières sont plus volatiles, notamment la nourriture dont une partie « disparaît » dans l'air ou dans les eaux usées. Seule une fraction ressort en tant que déchet solide dans les poubelles des cellules ou celles des cuisines. Pour évaluer cette part, la pesée des déchets sur 15 jours en cuisine, et la caractérisation du contenu d'une benne des déchets de cellules ont permis de donner des ordres de grandeur des quantités annuelles : 60 tonnes en sortie de cuisine, 100 tonnes en sortie des cellules.

Ainsi, l'étude a notamment permis de mettre en avant la quantité de déchets alimentaires produits : environ 200 kg par détenu par an. C'est bien plus que la moyenne des Français estimée à 79 kg (Ademe 2014) et du seuil réglementaire à partir duquel un producteur de biodéchets doit les faire traiter dans une filière adéquate³⁸. Les contraintes liées à la distribution de la nourriture en milieu carcéral expliquent en partie cet excès, mais n'exonèrent pas la MAS de cette obligation légale. La réalisation de l'analyse de flux de matières a ainsi pu permettre d'identifier l'infraction. Depuis, les pratiques de la MAS ont évolué. L'établissement composte sur place une grande partie des déchets alimentaires et surveille mieux les quantités de repas produites. Des expérimentations visent également à trouver des alternatives pour supprimer les barquettes plastiques utilisées pour la distribution des repas.

³⁸ Code de l'environnement : Article L541-21-1, R543-225 ; arrêté du 12 juillet 2011 fixant les seuils définis à l'article R. 543-225. Le seuil pour les biodéchets était de 120 t/an en 2012, et est de 10 t/an depuis 2016.

La réalisation d'une AFME à un niveau micro se base donc sur la recherche et l'interprétation de factures qui comptabilisent directement certains flux. La connaissance des processus, la caractérisation d'échantillons et des pesées ponctuelles permettent d'évaluer les quantités qui ne sont pas comptabilisées, avec néanmoins une marge d'erreur difficile à établir. Dans l'exemple, les questions d'eau et d'énergie n'ont pas non plus été abordées : l'objectif était de se concentrer sur les déchets solides. L'analyse réalisée n'est donc pas exhaustive et ne visait pas un bilan complet vérifiant la conservation de la masse à l'échelle du système. La question des déchets est toutefois très transversale et questionne la majorité des flux en sortie comme en entrée.

2) Au niveau d'un produit : Inventaire du Cycle de Vie

L'Inventaire du Cycle de Vie (ICV) vise à recenser les flux de matières et d'énergie mobilisés dans le cycle de vie d'un produit³⁹ : à partir de l'extraction des matières premières, en passant par l'utilisation, jusqu'au traitement en fin de vie (décharge, incinération, recyclage...). L'ICV est une étape indispensable à l'Analyse du Cycle de Vie (ACV). Celle-ci est définie par la norme internationale ISO 14040 (ISO, 2006) et concerne notamment l'interprétation, l'exploitation et la communication des résultats de l'ICV. Elle aide ainsi à la conception de produits respectueux de l'environnement, ou dit autrement, à l'« éco-conception » des produits (Janin 2000).

L'ICV permet de s'intéresser aux flux indirects générés pendant le cycle de vie des produits. Il s'agit de la matière mobilisée pendant les différentes phases (extraction des ressources, fabrication, usage, élimination des déchets), mais qui ne constitue pas le produit fini. Cela peut concerner le gaz utilisé pour chauffer un entrepôt, le gasoil lié au transport, les chutes de matières des processus de fabrication, l'eau du nettoyage... En ce qui concerne l'« énergie indirecte » mobilisée pendant la production, cette notion est plutôt connue sous le nom d'énergie grise.

Afin de juger convenablement des impacts d'un produit et d'effectuer un choix vertueux parmi les offres du marché, il est ainsi nécessaire de connaître les flux engendrés pendant la totalité du cycle de vie. Une erreur fréquente est de ne se concentrer que sur les performances du produit fini. C'est ainsi que l'on s'aperçoit que la construction des bâtiments basse

³⁹ Cela peut également s'appliquer à un service ou à une solution technique à la condition de définir l'« unité fonctionnelle » de service rendu pour pouvoir comparer différentes possibilités.

consommation nécessite autant d'énergie que leur fonctionnement pendant 30 à 50 ans (ARENE IdF & ICEB 2012), ou que la majorité des impacts imputables à de nombreux d'équipements électroniques survient pendant la phase de fabrication en raison de la grande quantité de matières et d'énergie mobilisées dans l'extraction des matières premières (EcoInfo 2014).

En théorie, la réalisation d'un ICV par une entreprise commence par un inventaire des ressources mobilisées en interne pour la fabrication du produit étudié. Ensuite, il s'agit d'évaluer les ressources indirectes mobilisées par les fournisseurs en amont et en aval de la chaîne logistique pour la production des ressources inventoriées.

En pratique, les entreprises connaissent très rarement les flux indirects mobilisés par leurs fournisseurs, et encore moins par les fournisseurs des fournisseurs... Elles ne peuvent pas non plus estimer précisément les ressources mobilisées pour la gestion de la fin de vie du produit, car celui-ci peut finir de différentes manières : incinéré, enfoui, recyclé... Ainsi, les ICV reposent en grande partie sur des évaluations génériques des flux indirects à l'aide de bases de données et logiciels dédiés à l'ACV. La difficulté à évaluer des flux qu'on ne maîtrise pas doit inciter à interpréter les résultats avec précaution. Tout en ayant le mérite de questionner nos moyens de production pour les rendre plus efficaces, les ACV peuvent présenter des résultats contradictoires en fonction des hypothèses réalisées et des bases de données mobilisées. Des analyses de sensibilité peuvent être ainsi mises en place pour évaluer les marges d'erreur et identifier les hypothèses et les paramètres les plus importants (Kozderka 2016).

3) À un niveau macro : Le métabolisme des territoires

Le niveau macro est celui des territoires : les villes, les pays, voire les continents... À cette échelle, les AFME se confondent souvent avec les études de métabolisme territorial, même si en pratique, ces dernières désignent plutôt la réalisation du bilan de matière que l'analyse des flux internes. Il n'y a pas de taille de territoire systématiquement pertinente pour mener une AFME, cela dépend des objectifs (Barles 2017). La démarche peut appuyer aussi bien des politiques urbaines que régionales, nationales ou même au-delà. Il est néanmoins pertinent de réaliser les AFME pour des territoires dont les contours sont communément reconnus et en usage, afin de faciliter leurs réalisations et l'appropriation des résultats. Ainsi, les territoires retenus dans les études sont généralement des territoires administratifs (région, département, ville, zone d'activité...). Dans certains cas, le contexte culturel ou géologique permet de justifier d'une autre délimitation du territoire, par exemple pour des zones transfrontalières ou

des bassins versants. Cela induit néanmoins une plus grande complexité dans la collecte des données.

Après avoir défini le territoire, il s'agit de collecter les données nécessaires à l'étude de son métabolisme. En théorie, l'AFME d'un territoire peut être réalisée en rassemblant et connectant les AFME des organisations qui le composent (entreprises, administrations, industries...). C'est l'approche dite *ascendante* ou « *bottom-up* ». Cette approche est toutefois limitée : les organisations ne réalisent souvent pas de bilan de matière formel, ou ne souhaitent pas le partager. Il n'existe pas non plus de norme pour rendre l'agrégation des bilans de matière aisée à un niveau macro. Ainsi, Brunner & Rechberger (2003) proposent d'évaluer uniquement les flux des entreprises et organisations les plus importantes pour dresser les grandes lignes du métabolisme du territoire. Cette approche a été utilisée par exemple pour l'étude du canton de Genève (Faist Emmenegger et al. 2003).

En alternative à l'approche ascendante, le CGDD (2014) suggère plutôt une approche *descendante* ou « *top-down* », qui vise à rassembler les données disponibles à l'échelle macro et fournies par différents acteurs de la statistique ou du territoire. C'est l'approche retenue par les régions françaises qui ont réalisé une comptabilité de leurs flux de matières. En pratique, cette approche se limite à la réalisation du bilan des entrées et sorties de matières et d'énergie d'un territoire. Elle ne vise pas à décrire la circulation interne des flux, à l'exception des quantités de matières recyclées ou accumulées sous forme de biens matériels durables (ou *addition nette au stock*, voir Figure 5 plus loin). C'est une des limites de la méthode qui considère le territoire comme une boîte (presque) noire. Combiner les approches ascendantes et descendantes semble être l'idéal pour combiner la connaissance des principaux flux internes apportée par la première et la vision globale obtenue par la seconde, mais encore peu d'études ont exploré cette voie plus complexe (Beloïn-Saint-Pierre et al. 2016).

Dans tous les cas, l'évaluation du métabolisme relève en grande partie d'un exercice minutieux et d'enquête à la recherche des « meilleures données disponibles » (Hoornweg et al. 2012). Le guide méthodologique du CGDD (2014) donne une liste de matières à considérer et indique les organismes qui peuvent renseigner les données. Certaines données sont gratuites et accessibles sur Internet comme la production agricole départementale⁴⁰, la production de

⁴⁰ Site Agrest : <http://agreste.agriculture.gouv.fr/> réalisé par le Service de la statistique et de la prospective au sein du Ministère de l'Agriculture, de l'Agroalimentaire et de la Forêt

déchets ménagers⁴¹, le transport de marchandises entre régions et vers/depuis l'étranger⁴², ou d'autres données sur l'environnement, l'énergie, le transport, le logement et la construction⁴³. Pour certaines matières, les données sont disponibles auprès d'organismes spécialisés. Les émissions dans l'air seront ainsi renseignées par les Agences agréées pour la surveillance de la qualité de l'air⁴⁴, les extractions minières sont connues de l'Union nationale des industries de carrières et matériaux de construction, ou de la Direction régionale de l'environnement, de l'aménagement et du logement (DREAL), etc. Les données publiques sont généralement accessibles à une échelle régionale ou départementale. Descendre à des niveaux inférieurs (villes, zones d'activités, etc.) est plus difficile. Nous y reviendrons dans la partie sur les limites de l'AFME.

Les données collectées sont ensuite classées en différentes catégories, dont les principales sont visibles en Figure 4 p. 63 : les importations, l'extraction de matières de l'environnement, les exportations et les rejets dans l'environnement. Afin d'avoir une vue complète et garantissant les principes de conservation de la masse, la méthode intègre en plus le calcul de flux d'équilibrage⁴⁵, des extractions intérieures inutilisées⁴⁶ et des flux indirects (la matière utilisée par d'autres territoires dans la fabrication des produits importés). Ainsi, l'étude menée en Alsace (Région Alsace & Ademe 2015) a abouti au bilan de matière présenté en Figure 5. De manière générale, les principales matières mobilisées dans les régions françaises en termes de masse sont les matériaux de construction, les combustibles fossiles (gaz et pétrole) et les produits alimentaires (Barles 2017).

⁴¹ SINOE@ déchets : <http://www.sinoe.org> réalisé par l'Agence de l'Environnement de la Maîtrise de l'Énergie.

⁴² Base de données SitraM : <http://www.statistiques.developpement-durable.gouv.fr/> réalisée par le service de l'observation et des statistiques du ministère de l'Environnement, de l'Énergie et de la Mer.

⁴³ Base de données Eider : <http://www.stats.environnement.developpement-durable.gouv.fr/Eider/> réalisée par le service de l'observation et des statistiques du ministère de l'Environnement, de l'Énergie et de la Mer

⁴⁴ Réunies au sein du réseau ATMO France : <http://www.atmo-france.org/fr/>

⁴⁵ Les processus de combustion ou de respiration mobilisent l'oxygène de l'air et rejettent de la vapeur d'eau et du dioxyde de carbone. Si ce dernier est pris en compte pour les combustions, dans le reste des cas, ces flux doivent être ajoutés au bilan pour respecter le principe de conservation de la masse.

⁴⁶ L'extraction intérieure de matière s'accompagne de pertes ou de déplacements de matières qui ne sont pas valorisées économiquement (restes miniers, résidus de récolte, érosion des terres agricoles...). Certains de ces matériaux peuvent avoir des effets sur l'environnement et sont donc comptabilisés. La matière étant inutilisée, la masse en entrée est identique à celle en sortie.

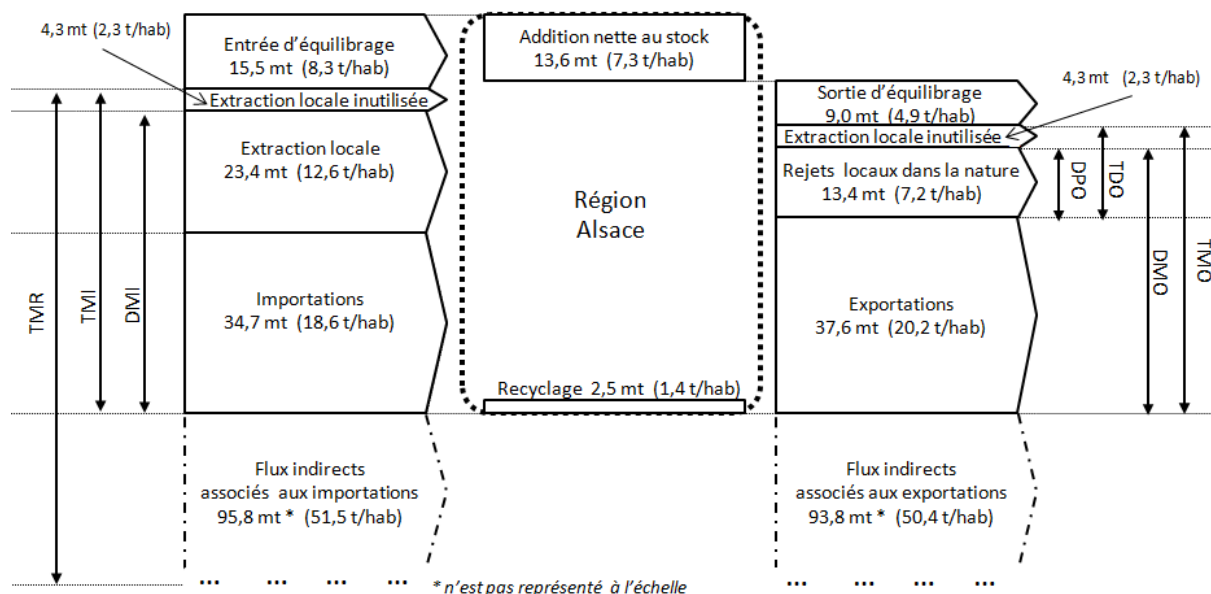


Figure 5 : Bilan de matière de la Région Alsace en 2010 à travers différents indicateurs, d'après Région Alsace et Ademe (2015).

mt : millions de tonnes, t/hab : tonnes par habitant. TMR : besoin total de matière, TMI : entrée totale de matière, DMI : entrée directe de matière, DPO : rejets locaux vers la nature, TDO : rejets locaux totaux vers la nature, DMO : sortie directe de matière, TMO : sortie totale de matière

Nous avons ainsi vu les principes régissant la réalisation des AFME, de manière générale et de manière un peu plus spécifique selon trois types de systèmes étudiés : les territoires à un niveau macro, les entreprises/organisations à un niveau micro, et les produits/services dans leur cycle de vie. Malgré leurs spécificités, ces trois approches peuvent être mobilisées dans des intentions similaires.

IV. Panorama des applications

Historiquement, les analyses de flux physiques ont permis (et permettent encore) la mise au point de procédés techniques, manufacturiers et surtout chimiques, ainsi que leur amélioration. Par la suite, l'outil s'est révélé également pertinent pour aborder différents enjeux du développement durable. Nous avons identifié dans la littérature un certain nombre d'applications possibles pour les AFME. Afin de pouvoir en dresser plus facilement le panorama, nous avons regroupé celles qui nous semblaient relever d'une même approche conceptuelle, avec probablement une certaine part de subjectivité dans la classification. Ces approches peuvent pour la plupart se décliner selon les trois types de systèmes mentionnés précédemment. Nous les décrivons toutefois de manière essentiellement transversale et

théorique, l'objectif est avant tout de montrer le potentiel et la polyvalence des AFME. Les difficultés à passer de la théorie à la pratique seront discutées dans la partie suivante.

1) Amélioration des procédés et de la circulation des flux

L'industrie repose sur la maîtrise des quantités de matières mobilisées dans les processus manufacturiers ou chimiques (Nikitin 1992). Cette maîtrise peut être renforcée à l'aide d'une AFME. Celle-ci peut en effet permettre d'identifier les processus de fabrication ou de fonctionnement qui pourraient être améliorés. Certains processus peuvent par exemple mener à des gaspillages significatifs (comme le gaspillage alimentaire de la MAS), mobiliser inutilement de la matière (surtout lorsqu'il y a des fuites), ne pas valoriser une matière qui pourrait l'être en étant mieux triée, etc. L'amélioration peut aussi concerner le transport et la circulation des flux : il s'agit d'éviter les situations paradoxales comme lorsque des territoires font venir de loin des ressources disponibles plus près, voire produites et exportées par le territoire lui-même (Buclet 2015). Ainsi, les AFME peuvent aider à améliorer les procédés et la circulation des flux et ainsi réduire les besoins matériels et énergétiques, et donc les coûts associés qu'ils soient sanitaires, économiques ou environnementaux (Brunner & Rechberger 2003).

2) Évaluation des impacts environnementaux

Certaines problématiques environnementales découlent des conséquences directes de flux trop importants entre l'environnement et nos systèmes socio-économiques, dans un sens comme dans l'autre. Le dérèglement climatique lié aux rejets de gaz à effet de serre en est un exemple (IPCC 2014). D'autres processus posent également de sérieuses questions en termes de durabilité : l'extraction de ressources naturelles qui les épuise, y compris celles qui sont renouvelables comme les ressources halieutiques (Idda 2014) ; la pollution régulière de l'eau, l'air et les sols par les rejets industriels et le transport ; l'appauvrissement des sols agricoles par la perturbation des cycles biogéochimiques du carbone, de l'azote et du phosphore (Peyraud et al. 2015), etc. Au total, la Commission européenne a ainsi recensé 14 types d'impacts

environnementaux⁴⁷. Ces impacts varient selon les caractéristiques physico-chimiques des matières prélevées ou émises dans l'environnement.

Une AFME peut ainsi permettre de quantifier les rejets des matières contribuant directement à ces problématiques environnementales et d'identifier les processus qui en sont à l'origine. L'objectif est de pouvoir mettre en place les actions adaptées pour réduire les impacts. Cela peut se faire par la réduction des quantités mobilisées ou produites, par la substitution par des matériaux plus durables et moins toxiques, par une amélioration des processus techniques et logistiques, ou encore par un meilleur traitement des déchets et rejets.

3) Étude des phénomènes dispersifs et identification des gisements

Un aspect particulier du cycle de la matière concerne les phénomènes qui la dispersent dans l'environnement la rendant difficilement récupérable. Ces phénomènes sont de cinq natures différentes :

- Les rejets liquides et gazeux : sous cette forme, la matière est intrinsèquement dispersive et s'éparpille facilement une fois émise dans l'environnement.
- Les pertes dissipatives : elles concernent essentiellement les objets mobiles qui s'usent par friction (pneus, freins...) et les infrastructures telles que les chaussées qui s'érodent sous les effets de la circulation, de la météo et du temps⁴⁸ (Landner & Lindeström 1998 ; CGDD 2014). Il est possible d'inclure également les accidents matériels qui peuvent avoir un effet dissipatif, mais difficilement quantifiable.
- L'usage de produits dissipatifs, c'est-à-dire qui ont été spécifiquement conçus pour être dispersés. Il peut s'agir d'engrais, de graines végétales ou de sel de déneigement (CGDD 2014).
- La dispersion de la matière dans les objets techniques à haute complexité. Par exemple, les circuits imprimés contiennent de nombreux métaux précieux, mais en très faible concentration.

⁴⁷ Recommandation de la Commission du 9 avril 2013 relative à l'utilisation de méthodes communes pour mesurer et indiquer la performance environnementale des produits et des organisations sur l'ensemble du cycle de vie : <http://data.europa.eu/eli/reco/2013/179/oj>

⁴⁸ L'usure des pneus représente en Alsace environ 1900 tonnes de matière dispersée dans l'environnement par an. L'usure des chaussées est quant à elle estimée à plus de 300 000 tonnes (Région Alsace et Ademe 2015).

- Les rejets de déchets solides dans la nature (hors biomasse). Bien qu'initialement consistants, ces déchets se dégradent lentement sous l'effet de la météo et du temps et relâchent des fragments ou des particules pouvant être nocifs.

Ces phénomènes dispersifs ont des dynamiques fondamentalement différentes qui doivent être étudiées séparément. Toutefois, certains points communs à ces dynamiques peuvent être soulignés. Tout d'abord, la matière finit dans l'environnement (sauf pour la partie des objets techniques qui est collectée avec les déchets) et peut donc potentiellement être source de pollution. Les pertes dissipatives sont notamment associées à des problématiques de qualité de l'air⁴⁹.

Ensuite, cette matière peut ruisseler, entraînée par le vent et les eaux des pluies ou des rivières. Une partie arrive ainsi dans les stations d'épuration, ce qui peut être considéré comme un risque de pollution car elle peut contenir des éléments métalliques toxiques (Baize et al. 2006), ou comme une ressource de plusieurs millions d'euros car certains de ces éléments sont précieux (Westerhoff et al. 2015). Une autre partie de la matière peut ruisseler jusqu'aux océans où elle s'y accumule, le résultat le plus visible étant celui des « continents de plastiques » (Garric 2012).

Enfin, une dernière caractéristique commune des phénomènes dispersifs est d'empêcher la réutilisation de la matière. Les déchets à haute complexité (DEEE principalement) ou les métaux des boues des stations d'épuration posent un même problème : les matériaux économiquement intéressants sont en faible concentration et les techniques d'extraction actuelles doivent être perfectionnées avant que l'on puisse considérer ces gisements de matières comme exploitables.

L'AFME est ainsi un outil qui permet d'aborder la question de la dispersion de la matière en évaluant les quantités mises en jeu et les conséquences environnementales. Elle peut également participer à identifier et quantifier les gisements résultants des phénomènes de ruissellements ou d'une bonne gestion des déchets, et ainsi évaluer l'intérêt économique à les exploiter.

⁴⁹ « L'usure des routes, des pneus et plaquettes de frein est responsable de 20 % des émissions parisiennes de particules » (Airparif, 2016, p15).

4) Suivi de la réglementation applicable

Afin que les matières dangereuses soient gérées avec précaution et que le développement économique se fasse dans le respect de l'environnement, la législation s'est enrichie au fil du temps d'un certain nombre de règles, par exemple :

- Le régime des Installations classées pour la protection de l'environnement (ICPE)⁵⁰ s'applique à des entreprises dont l'activité ou les substances qu'elle manipule présentent un risque pour l'environnement. Selon les volumes, les contraintes diffèrent entre une déclaration et une demande d'autorisation.
- La réglementation sur les déchets⁵¹ vise certains produits en particulier : les emballages doivent être optimisés et valorisables, les biodéchets doivent être traités de manière adéquate au-delà de 10 tonnes par an, etc. De manière générale, la loi désigne « tout producteur ou détenteur de déchets [comme] responsable de la gestion de ces déchets jusqu'à leur élimination ou valorisation finale, même lorsque le déchet est transféré à des fins de traitement à un tiers »⁵².
- L'instauration d'un marché carbone (De Perthuis & Trotignon 2017) et l'obligation de réaliser un bilan d'émission de GES pour les plus grandes entreprises et collectivités⁵³.
- La mise en place d'une fiscalité environnementale (redevance incitative, taxe, bonus-malus, appel à projets, crédit d'impôt...).

La réglementation tend ainsi à renforcer la maîtrise des flux physiques par les acteurs qui les manipulent. Elle s'appuie sur des impératifs comptables, ouvre des opportunités de marché, responsabilise les producteurs de déchets, etc. La réalisation d'un bilan de matière n'est toutefois pas légalement contrainte. C'est pourtant un outil adapté pour décrire une réalité physique à confronter avec le cadre juridique et faire le lien entre différentes obligations : comptabilisation des émissions de GES, des (bio)déchets, estimation des stocks présents dans les ICPE...

⁵⁰ Livre V – Titre Ier du Code de l'environnement

⁵¹ Livre V – Titre IV du Code de l'environnement

⁵² Code de l'environnement : Article L541-2

⁵³ Code de l'environnement - Article L229-25; Décret n° 2011-829 du 11/07/2011 relatif au bilan des émissions de gaz à effet de serre et au plan climat-énergie territorial

Combiner une analyse de flux et une analyse juridique reste toutefois complexe. Il n'est ainsi pas surprenant de voir que certains acteurs ignorent les règles qu'ils devraient respecter : c'était par exemple le cas de la MAS avec ses déchets alimentaires, ou de certains restaurateurs ne faisant pas traiter leurs huiles usagées (Georgeault 2015).

5) Communication sur les progrès et les impacts environnementaux

Les AFME permettent de suivre l'évolution des flux et d'évaluer les bénéfices environnementaux d'un projet. Ces informations peuvent ensuite être détournées de leur objectif initial de suivi et d'aide à la décision en devenant un outil d'aide à la communication environnementale, mettant en avant les progrès réalisés en termes de réduction d'émission de GES, de déchets, de consommation d'eau...

Dans un cadre plus formel, les résultats d'une AFME peuvent aussi prendre la forme d'un affichage environnemental⁵⁴. L'objectif est de porter à la connaissance du consommateur les impacts liés à la production de différents biens et services (Francois-Lecompte et al. 2013). Ces impacts peuvent notamment être évalués à partir d'un inventaire du cycle de vie. D'ailleurs, de la même manière que pour les ICV, des règles communes de comptabilité doivent être établies pour pouvoir effectivement comparer différents produits d'une même catégorie. Parfois, l'affichage prend la forme plus simple d'un label. Il ne s'agit alors plus de quantifier des impacts, mais cela peut donner tout de même une information sur la présence, l'absence ou les caractéristiques de certains flux mobilisés dans le cycle de vie du produit⁵⁵.

6) Compréhension scientifique et vision partagée

Les AFME aident à la compréhension des dynamiques d'une organisation ou d'un territoire. D'un point de vue scientifique, cette connaissance est intéressante en soi, même si elle n'est pas suivie d'applications pratiques. Elle peut porter sur les aspects géographiques, économiques, sociaux, historiques des flux mobilisés par l'organisation ou le territoire. Le champ émergent de l'écologie territoriale vise ainsi, à travers des analyses interdisciplinaires, à décrire les trajectoires socio-économico-écologiques des territoires (Buclet 2015).

⁵⁴ L'ADEME propose à cet effet une base de données pour l'affichage environnemental des produits de grande consommation : <http://www.base-impacts.ademe.fr/>

⁵⁵ Il peut s'agir par exemple de l'absence d'utilisation des produits chimiques dans les produits étiquetés « Agriculture Biologique », de l'origine des produits fabriqués à travers les marques régionales...

Mais au-delà des scientifiques, les AFME doivent permettre à l'ensemble des acteurs territoriaux concernés de se rendre compte des moyens matériels et énergétiques mobilisés par leur territoire, leur entreprise et les produits qu'ils consomment... À l'échelle régionale, l'entrée directe de matière (DMI – c'est-à-dire les importations et l'extraction locale) est de plusieurs dizaines de tonnes par habitant et par an : 31 tonnes pour un Alsacien (Région Alsace & Ademe 2015 ; voir également Figure 2), 21 tonnes pour un parisien (Barles 2014). Cette mise en perspective peut ainsi participer à la construction d'une « vision partagée ». En tentant de rendre compte de manière objective (malgré les biais) des forces et faiblesses matérielles et énergétiques d'un territoire auprès de ses acteurs, c'est bien un langage commun qui est mis en place, ou une « entrée en intelligibilité » selon Cerceau et al. (2014). Celle-ci est un préalable essentiel à toute démarche collective comme celles d'économie circulaire ou d'écologie industrielle dont nous reparlons par la suite. Savoir ce que le territoire produit, importe, consomme, jette ou recycle devient alors une connaissance partagée, augmentant ainsi la capacité des acteurs à s'inscrire dans une dynamique de coopération (Debuisson 2014). Cette connaissance partagée peut également servir de base à l'élaboration et le suivi des politiques publiques en matière de développement durable (Barles 2010).

Afin de rendre plus accessibles et compréhensibles les résultats d'une AFME, ceux-ci peuvent prendre différentes formes, que ce soient des indicateurs (Barles 2014 ; Chrysoulakis et al. 2015), des cartes (Silvestre et al. 2012 ; Le Noé et al. 2016 ; Pincetl et al. 2016), des bilans entrées/sorties, des diagrammes de flux ou diagrammes de Sankey (Barles 2014 ; CGDD 2014) et bien d'autres formes de représentations graphiques (histogrammes, circulaires...) ou structurées (voir Chapitre 3/II.3). Néanmoins, dans la partie sur les limites des AFME, nous parlerons de la difficulté à rendre cette vision effectivement partagée. L'AFME doit en effet faire l'objet d'un effort pédagogique important pour que ses tenants et aboutissants soient compris par le plus grand nombre.

7) Les flux de matières dans les projets de construction et d'aménagement

Le secteur du bâtiment et des travaux publics (BTP) est concerné à deux niveaux par la question des flux physiques. D'une part, il s'agit du secteur qui mobilise le plus de matières en termes de masse (hors eau). Le rythme de développement d'une ville peut d'ailleurs se lire à travers la quantité de matériaux de construction mobilisés (Ferrão & Fernández 2013). Encore peu de ces matériaux sont recyclés lors des opérations de démolition (Rouvreau et al. 2013). En

France, la loi relative à la transition énergétique pour la croissance verte⁵⁶ fixe un objectif de 70 % de valorisation à horizon 2020 pour les déchets non dangereux du BTP. La valorisation de la matière peut aussi concerner des stocks abandonnés : par exemple de vieilles conduites d'eau laissées sous terre, d'anciennes voies ferrées ou les matériaux de construction des friches industrielles... Ces réserves de matières représentent un gisement important qui pourrait être exploité dans le futur. C'est sur cette idée que reposent les recherches sur les « mines urbaines » (Hendriks et al. 2000 ; Barles 2017).

D'autre part, une fois en place, l'architecture structure le cadre dans lequel les flux vont circuler. Elle doit ainsi tenir compte des réseaux vitaux (eau et énergie), mais aussi des autres flux. Les travaux menés à la MAS (De Boisséson et al. 2013) montrent que les flux de déchets et leur spatialisation n'ont pas été suffisamment anticipés. Ainsi, la configuration des lieux fait que les poubelles sont entreposées et remplies à l'arrière du bâtiment, mais elles sont ramassées par les services municipaux à l'avant. Pour les amener d'un côté à l'autre du bâtiment, des agents font tous les jours parcourir 300 mètres (dans un sens) à 5/6 poubelles d'un volume de mille litres sur un chemin de terre battue peu adapté. Une configuration permettant l'entreposage des poubelles du même côté que leur collecte aurait permis d'économiser ces efforts quotidiens.

À l'échelle d'une ville, une mauvaise gestion des infrastructures et des réseaux permettant la circulation des flux peut avoir des conséquences sanitaires et écologiques importantes (UNEP 2013). Cela est particulièrement vrai pour les villes à forte croissance des pays émergents (Cohen 2004). Dans ces villes, le développement des réseaux logistiques, d'énergie et d'eau (potable et assainissement) n'arrive pas à suivre le développement urbain. Il en résulte des constructions anarchiques et des habitats densément peuplés, voire des bidonvilles (Kennedy et al. 2014). Ce développement incontrôlé génère un surcroît de pollutions qui peuvent avoir un impact sur la disponibilité en eau potable déjà parfois très réduite (Shao et al. 2006). Dans les villes des pays développés, ces problématiques sont moins critiques. Toutefois, nous sommes encore loin d'une organisation du transport des personnes et des marchandises que l'on peut qualifier de durable. Par exemple, l'étalement urbain permis par l'essor des véhicules motorisés engendre une dépendance accrue au pétrole (Ferrão & Fernández 2013). Cela pose des questions de résilience en vue de l'épuisement des ressources que nous verrons plus loin. Mais cela questionne aussi l'aménagement et la logistique des villes,

⁵⁶ Loi n° 2015-992 du 17 août 2015 relative à la transition énergétique pour la croissance verte

et plus largement des territoires (Kennedy et al. 2011). Il s'agit d'évaluer les besoins, de déterminer où et comment s'approvisionner, de faciliter la circulation des flux, le ramassage et le recyclage des déchets, ou encore de réduire le transport qui dégrade la qualité de l'air. Des travaux du programme des Nations Unies pour l'environnement ont mis en évidence l'importance des infrastructures et des réseaux pour la circulation de la matière (UNEP 2013). Ils ont un impact direct sur l'efficacité dans l'utilisation des ressources et sur la santé et le bien-être des habitants. Ces travaux ont ainsi identifié différentes approches pour adapter les infrastructures des villes dans une optique de développement durable. Toutefois, intégrer les questions relatives à la matière dans les outils d'aménagement du territoire reste encore très complexe (Chrysoulakis et al. 2015).

Réaliser une AFME en amont des projets de construction et d'urbanisation présente ainsi plusieurs intérêts. Il s'agit d'évaluer la nature et la quantité des flux qui seront mobilisées pendant les phases de chantier puis d'usage de l'espace. Durant la phase de construction, ces informations peuvent aider à coordonner un chantier plus vertueux et économique en optimisant l'usage de la matière, par exemple en exploitant la synergie entre les opérations de déconstruction et de construction, ou en s'assurant d'un meilleur tri pour favoriser le recyclage. En prévision de la phase d'usage, l'AFME peut permettre aux architectes et urbanistes de mieux anticiper la circulation des flux (par exemple celle des déchets) afin de la rendre plus ergonomique.

8) Rendre l'économie circulaire, et l'industrie écologique

Une des stratégies des politiques publiques pour assurer un développement durable vise à découpler la production de richesse de l'usage des ressources (UNEP 2013 ; Kahn 2015) ou, dit autrement, d'aller vers une *dématérialisation*. Il s'agit d'assurer un même niveau de performance économique en utilisant moins de matières et d'énergie. Pour cela, une des actions recherchées est de « boucler les flux » manipulés par les différents acteurs du territoire. Cette approche est inspirée du fonctionnement des écosystèmes naturels dans lesquels les rejets de certains organismes servent de ressources à d'autres. La matière circule ainsi dans des cycles biogéochimiques.

C'est sur ces bases que l'écologie industrielle est apparue dans les années 1970, avant d'évoluer par la suite vers les champs plus larges de l'écologie territoriale (Barles 2010) et de l'économie circulaire (voir paragraphe suivant). Dans les trois terminologies, les fondements sont les mêmes : réduire l'usage des ressources et réutiliser ou recycler au maximum la matière

auparavant considérée comme un déchet à éliminer. Cette idée se traduit concrètement par trois types de synergies entre les acteurs économiques d'un territoire (Adoue 2004 ; Georgeault 2015) : (1) La mutualisation dans la fourniture d'un service, l'approvisionnement d'une ressource ou le traitement d'un déchet ; (2) La valorisation, les déchets des uns devenant les matières premières des autres ; (3) La substitution d'une ressource par une autre plus locale et/ou renouvelable. La réalisation de bilans de matière par les acteurs permet ainsi d'identifier les synergies potentielles. Celles-ci sont plus simples à mettre en œuvre lorsque ces acteurs sont proches géographiquement et ont l'habitude de travailler ensemble (Debuisson 2014 ; Georgeault 2015). Mais même dans les zones industrielles propices, les démarches d'écologie territoriale ne sont encore qu'à un stade émergent. Ainsi, la part de matière effectivement recyclée en France reste actuellement faible et concerne moins de 10 % de la matière en entrée (DMI)⁵⁷, dont essentiellement des déchets du BTP. De plus, une grande partie de la matière utilisée par nos sociétés est impliquée dans les processus dispersifs mentionnés précédemment (concernant les émissions gazeuses, l'usure des routes, les produits complexes...) et elle ne peut donc pas toujours être réutilisée. Ainsi, en plus des synergies industrielles, il est nécessaire de trouver également des solutions pour réduire les impacts et la perte de matières résultant des processus dispersifs. Cela concerne en particulier le secteur du transport et les appareils de haute technologie (Bihoux 2014).

Le concept d'économie circulaire est apparu plus tard, dans les années 2000, avant d'être intégré dans la législation française en 2015 avec la loi relative à la transition énergétique pour la croissance verte. Il vise à rassembler au sein d'un terme unique plusieurs dynamiques économiques et écologiques en plus de la seule question des synergies industrielles. L'économie circulaire comprend ainsi également l'éco-conception (qui consiste à concevoir des produits durables et recyclables) ou la consommation responsable (qui est possible à travers l'affichage environnemental). Il y a également l'économie de la fonctionnalité dans laquelle la fourniture d'un service se substitue à l'achat d'un objet (Maillefert & Robert 2014). Le consommateur paye non plus l'objet qu'il utilise, mais sa location et le service rendu. L'entretien étant compris, le fabricant (devenu, de fait, prestataire de service) a alors tout intérêt à rendre ses produits durables et à en réduire l'obsolescence. L'intérêt environnemental d'une

⁵⁷ D'après les études menées dans plusieurs territoires français (Barles 2009 ; Alterre Bourgogne 2013 ; Barles, 2014 ; Région Alsace et Ademe, 2015)

offre de service par rapport à la vente d'un produit peut être évalué notamment grâce aux analyses de cycle de vie.

9) Des questions de résilience et d'anticipation de l'épuisement des ressources

Le fonctionnement de nos sociétés mobilise des ressources naturelles qui s'épuisent : les énergies fossiles, l'eau, les métaux, les terres agricoles, les poissons... Si nous ne parvenons pas à développer une économie plus sobre, celle-ci risque de connaître d'importants dysfonctionnements lorsque ces stocks seront épuisés, à l'horizon d'une à quelques décennies selon les ressources (Servigne & Stevens 2015). Dans ce contexte, la Commission européenne a d'ailleurs identifié vingt-sept « matières premières critiques » en fonction de leur importance économique et du risque de pénurie (Commission européenne 2017). À cela s'ajoutent les risques de conflits, de crises économiques ou de catastrophes naturelles qui peuvent également survenir et rompre l'approvisionnement d'une ressource essentielle à un territoire (Ferrão & Fernández 2013), par exemple le gaz russe pour l'Europe (Locatelli 2012).

Les risques matériels ne se limitent pas à la question de l'approvisionnement, mais concernent l'ensemble de la chaîne de production et d'élimination, notamment les infrastructures les plus stratégiques étant donné leur fonction et leur dimensionnement⁵⁸ (usine d'incinération, station d'épuration, centrales énergétiques électriques ou thermiques). D'une part les ressources que ces installations mobilisent pour fonctionner sont conséquentes, d'autre part, en cas de non-fonctionnement⁵⁹, les conséquences sanitaires, écologiques ou économiques peuvent être coûteuses, voire critiques.

Ces risques logistiques peuvent se présenter à court, moyen et long termes, et peuvent avoir des conséquences temporaires ou définitives. Ils questionnent les territoires et les organisations dans leur capacité de résilience : sont-ils capables de continuer à fonctionner avec une pénurie de ressources ? Les AFME peuvent apporter un éclairage en quantifiant les besoins et en décrivant la circulation des flux. Elles peuvent ainsi servir de base à la construction de

⁵⁸ Certaines chaufferies à bois peuvent être surdimensionnées dans le sens où elles mobilisent un volume trop important de bois. L'aire d'approvisionnement peut ainsi devenir trop grande (avec un besoin plus important de transport) ou présenter une menace pour les forêts locales (Sawerysyn et al. 2012).

⁵⁹ Par exemple l'arrêt forcé d'une usine d'incinération (Gérard 2017) ou le dysfonctionnement d'une station d'épuration (Lagain 2016).

scénarios envisageant certaines ruptures logistiques. Mais face à la multitude de scénarios possibles, il est complexe de définir les stratégies d'adaptation qui tiennent compte du contexte économique, social et géopolitique et qui permettent de surmonter d'éventuels futurs aléas. C'est toutefois l'ambition des recherches sur la résilience territoriale (Dauphiné and Provitolo 2007 ; Toubin et al. 2012).

V. Les limites des AFME

Malgré son large panel d'applications, réaliser une analyse de flux ne permet pas d'aborder tous les aspects de nos sociétés. Des questions économiques (échanges de services, valeur ajoutée...), sociales (emploi, éducation...) ou culturelles semblent difficilement abordables par le seul angle de la matière ou de l'énergie. D'autre part, quantifier les flux physiques est avant tout une approche descriptive et non explicative. Ainsi, les AFME ne permettront pas d'expliquer le comportement des acteurs du territoire (pourquoi un produit est-il jeté ? Pourquoi un acteur n'utilise-t-il pas des ressources locales ? ...), mais seulement d'en décrire la résultante en termes de flux.

Outre ces limitations fonctionnelles intrinsèques, la réalisation concrète d'une AFME pose encore des difficultés. Ce sont des difficultés d'ordre technique (pour manipuler de nombreuses données), méthodologique (pour analyser les résultats à travers différents cadres théoriques), pédagogique (pour restituer les informations à un public large), réglementaire (pour disposer d'un cadrage de la comptabilité matière) et organisationnel (pour rassembler les compétences nécessaires à l'analyse).

1) Des données éparses, difficiles à regrouper

Des facteurs essentiels pour la réalisation d'une AFME sont la disponibilité et l'utilisabilité des données (Chapitre 1/II.3). Cela est particulièrement vrai à l'échelle des territoires et les situations sont extrêmement différentes selon les pays, les cultures et les moyens des services de statistiques. En France, ces services fournissent de nombreuses informations publiques sur les flux de matière, mais les rassembler prend parfois du temps. Elles ne sont pas non plus toujours disponibles en dessous de l'échelle départementale (Barles 2014) même si des méthodes visent à évaluer les flux à des échelles plus fines (Courtonne 2016). Dans certains cas, l'information est payante⁶⁰, restreinte ou soumise au secret

⁶⁰ Par exemple, les données de la base SitraM sont payantes pour un niveau détaillé à l'échelle départementale.

statistique⁶¹. Toutefois, l'accessibilité des données va certainement s'améliorer avec la récente loi pour une République numérique⁶² qui vise une meilleure ouverture et accessibilité des données publiques. La question de la qualité des données sources est également importante. La transcription de la réalité vers l'information statistique, notamment à l'aide d'enquêtes, ne peut être que partielle et approximative (Desrosières 2008). Une grande partie des flux de matières reste ainsi non mesurée ou à la discrétion de ceux qui les manipulent.

Tous ces obstacles pour collecter et standardiser les données rendent les études territoriales longues et complexes⁶³ (Bahers 2012 ; Pauliuk et al. 2019). Une plus grande automatisation des traitements et un meilleur partage des informations permettraient de rendre la réalisation des bilans de matière plus faciles. Pour cela, plusieurs points sont à considérer :

- La généralisation des normes comptables pour la comptabilité matière. Les travaux menés par les institutions internationales (Nations Unies 2016 ; Eurostat 2013) sur la comptabilité environnementale ont ensuite été adaptés dans des études à différentes échelles (villes, départements, régions, pays). Toutefois, à l'échelle des entreprises, cette comptabilité n'est pas encore suffisamment formalisée au point de pouvoir être généralisée. Il s'agit de définir les méthodes et nomenclatures adaptées à la gestion de la matière, conçues pour faciliter les comparaisons et les agrégations et en lien avec la comptabilité économique (Ribon 2017).
- Le développement des infrastructures de données, matérielles et logicielles (Kennedy & Hoornweg 2012). L'objectif est de permettre aux acteurs de partager leurs données pour les études territoriales (CGDD 2014), ou pour parfaire les inventaires de cycle de vie des produits en réduisant les inconnues. Vu les impératifs de confidentialité, le développement et l'exploitation de ces infrastructures devraient certainement relever d'une maîtrise d'ouvrage publique.
- La sensibilisation et la formation des acteurs pour les inciter à mieux collecter les données sur la matière qu'ils manipulent, et à les partager dans une optique de coopération territoriale. Ce partage résulte néanmoins d'un équilibre entre le respect des

⁶¹ Ces règles définissent pour les entreprises un minimum d'agrégation de trois unités pour la diffusion d'une donnée statistique, et aucune de ces unités ne doit contribuer à plus de 85 % du total (Insee 2010).

⁶² Loi n° 2016-1321 du 7 octobre 2016 pour une République numérique

⁶³ Environ 85 à 105 jours × hommes pour une région (CGDD 2014 ; Région Alsace & Ademe 2015)

données privées et leur nécessaire partage pour évaluer les impacts et déployer les synergies entre acteurs.

Les données ont une place importante dans les analyses de flux territoriales. Leur agrégation et leur croisement offrent un large éventail d'applications théoriques comme nous l'avons dans notre panorama. Les outils numériques sont ainsi de nouvelles possibilités pour mieux comprendre nos sociétés et faciliter la transition écologique. Ils posent toutefois également des questions en termes de sécurité, de confidentialité, d'incompatibilité des formats et des méthodes de comptabilisation, ou encore de propriété intellectuelle (voir Chapitre 1/IV). Les pouvoirs publics doivent s'emparer de cette question en proposant des solutions techniques pérennes s'ils veulent mettre à profit les nouvelles possibilités de ces outils dans la gouvernance territoriale et sortir du monopole des grands acteurs privés du numérique (AFNOR 2015).

2) La quantification des flux à travers différents cadres théoriques

Les acteurs économiques expriment systématiquement les flux et les stocks en unité monétaire. C'est en effet un dénominateur commun pour évaluer des opérations de nature très diverse (Eurostat 2013b). De manière similaire, pour la comptabilité matière, les flux et les stocks sont généralement exprimés en unité de masse. Le choix de l'unité servant de dénominateur commun n'est pas neutre et privilégie certains processus au détriment d'autres. Par exemple, l'unité monétaire intègre mal les externalités (Buclet 2015). Quant à la masse, elle n'est pas directement corrélée aux impacts environnementaux : quelques kilogrammes d'une matière très toxique (par exemple les métaux lourds) peuvent avoir des effets bien plus ravageurs sur l'environnement que quelques tonnes de graviers de construction.

Finalement, aucune unité ne peut aborder simultanément toutes les thématiques. Mais la recherche scientifique a mis au point un large panel d'unités pour les aborder séparément. Par exemple pour la question du réchauffement climatique, il s'agit de quantifier les flux en kilogramme équivalent CO₂⁶⁴. Dans ses travaux sur la performance environnementale⁶⁵, la Commission européenne recense onze autres unités pour évaluer les impacts environnementaux

⁶⁴ Quelques exemples de conversion : 1 kg de matière solide ou liquide = 0 kg eq. CO₂; 1 kg de dioxygène (O₂) = 0 kg eq. CO₂; 1kg de CO₂ = 1 kg eq. CO₂; 1 kg de méthane (CH₄) = 25 kg eq. CO₂.

⁶⁵ Recommandation de la Commission du 9 avril 2013 relative à l'utilisation de méthodes communes pour mesurer et indiquer la performance environnementale des produits et des organisations sur l'ensemble du cycle de vie (2013/179/UE) : <http://data.europa.eu/eli/reco/2013/179/oj>

comme l'appauvrissement de la couche d'ozone, l'acidification et l'eutrophisation des milieux, l'épuisement des ressources, la transformation des sols, etc. Certaines sont reprises dans le Tableau 4 et sont accompagnées d'autres unités disponibles dans la littérature scientifique. Notons que la matière et les biens peuvent également être quantifiés par les flux indirects qu'ils mobilisent pendant leur cycle de vie. Ainsi, un objet n'a généralement pas d'impact climatique en lui-même, mais il peut être quantifié par les émissions dont il est responsable pour sa fabrication, son fonctionnement et/ou son élimination.

Thématique ou discipline	Type d'unité	Exemples d'unité(s) disponible(s)
Économie	Monétaire	Euro (€), Dollar (\$)...
Physique	Masse Volume	Gramme (g), Tonne (t)... Mètre cube (m ³), Litre (L)...
Énergie	Énergie primaire ou finale	Kilowattheure (kWh EP ou EF)
Eau	Empreinte aquatique (Hoekstra & Chapagain, 2007)	Mètre cube d'eau (m ³)
Biophysique	Émergie (Odum, 2002)	Solar Energy Joule (sej)
Impact environnemental global	Empreinte environnementale (Kitzes et al., 2008)	Mètre carré (m ²) EE
Changement climatique*	Potentiel de réchauffement planétaire (Ademe 2010)	Kilogramme équivalent CO ₂ (kg eq. CO ₂)
Substances affectant les voies respiratoires*	Modèle RiskPoll (Humbert, 2009)	Kilogramme équivalent PM _{2,5}
Acidification*/ Eutrophisation terrestre*	Modèle d'accumulation d'excédents (Seppälä et al, 2006)	Mole eq H ⁺

* Inclus dans les recommandations de la Commission européenne pour mesurer et indiquer la performance environnementale des produits et des organisations sur l'ensemble du cycle de vie

Tableau 4 : Quelques exemples d'unités disponibles pour l'étude de différentes thématiques

La matière peut donc être quantifiée par ses propres caractéristiques économiques ou physico-chimiques ou celles de ses flux indirects. Cela permet aux AFME de s'inscrire dans différents cadres théoriques qui mobilisent leurs propres méthodologies, critères et indicateurs (Figuières et al. 2007). Il en résulte fréquemment des résultats contradictoires sur l'importance de certaines matières selon les critères retenus. Une substance peut avoir un fort impact sur le réchauffement climatique, mais aucune toxicité (le CO₂) et vice-versa (le mercure). Ainsi, des recherches s'intéressent aux outils permettant la prise de décision sur la base de plusieurs critères économiques, écologiques ou sociaux (Chrysoulakis et al. 2015 ; Courtonne 2016). Il faut toutefois être capable d'analyser et quantifier les flux selon différents cadres théoriques et cela relève souvent d'approximations qui obligent à manipuler avec précaution les résultats obtenus, tant dans leur fiabilité que dans leur signification (Figuières et al. 2007).

3) Rendre intelligible l'AFME auprès des acteurs du territoire

Les acteurs du territoire sont de plusieurs natures : décideurs politiques (élus, administration), économiques (entreprises) ou techniques (ingénieurs, architectes...), acteurs professionnels (salariés, syndicats, fédérations) ou de la société civile (associations, citoyens), chercheurs et autres praticiens, etc. Leurs compétences, intentions et poids sur les décisions diffèrent, tout autant que les informations qu'ils pourront exploiter d'une AFME. Aussi, les praticiens ayant réalisé ces études doivent présenter leurs conclusions de différentes manières, tant sur la forme que sur le fond, en fonction des acteurs à qui ils sont destinés. L'objectif est de « produire de la connaissance dont peuvent se saisir les acteurs concernés, tant individuellement que collectivement » (Buclet 2015, p.40).

Pour faciliter l'interprétation des résultats par un plus large public, le guide méthodologique du CGDD (2014) comprend une partie dédiée aux modes d'analyse, permettant d'aller au-delà de la simple quantification des masses mises en jeu. Les travaux de Chrysoulakis et al. (2015) ont également abouti à l'élaboration d'un système informatique d'aide à la décision, permettant à des élus, urbanistes ou architectes de comprendre et d'évaluer les conséquences sur le métabolisme urbain de différents projets d'aménagement à travers différentes approches : construction de grille d'indicateurs personnalisés, représentations cartographiques... Toutefois, encore peu d'études sont effectivement rendues accessibles à un public diversifié (Beloin-Saint-Pierre et al. 2016) et les travaux dans ce sens méritent d'être poursuivis.

4) L'évolution du cadre législatif

Nous avons vu que la législation imposait des contraintes pour la gestion de certaines matières. Le lien entre le cadre législatif et la matière est d'ailleurs en constante évolution. Progressivement, la législation a intégré de nouvelles règles à travers par exemple la mise en place du régime des ICPE (1976), de filières de collectes nationales et règlementées des déchets (depuis 1992 selon les filières), d'un bilan GES obligatoire (2011), etc. La comptabilité matière est devenue également incontournable pour les États européens.

Ainsi, la réglementation devient de plus en plus contraignante et incitative, et demande aux entreprises et territoires d'évaluer leurs impacts environnementaux et de mieux maîtriser leurs flux selon des règles de plus en plus formelles. Récemment encore, la loi sur la transition énergétique pour la croissance verte a donné un signal fort en faveur de l'économie circulaire. Les territoires qui disposent de plus d'autonomie avec la loi NOTRe⁶⁶ doivent ainsi s'approprier cette question des ressources et en optimiser les usages.

Avec l'évolution des textes de loi, les bilans de matière deviennent de plus en plus nécessaires pour appréhender certaines réalités matérielles, respecter la réglementation et s'adapter à son évolution. Ils ne sont pas encore obligatoires, mais plusieurs travaux préconisent de les rendre systématiques pour les entreprises et les territoires (Kennedy & Hoornweg 2012 ; Georgeault 2015). Cela nécessite toutefois de nouvelles évolutions du cadre législatif pour accompagner cette généralisation de la comptabilité matière.

5) Des compétences spécifiques à mobiliser et la coordination des acteurs

Pour pouvoir réaliser et exploiter une analyse de flux de matières, il faut mobiliser un certain nombre de compétences : une expertise méthodologique, la connaissance des thématiques environnementales, une aisance dans la manipulation des données, des capacités pédagogiques et graphiques pour restituer l'information et la connaissance des acteurs territoriaux. Ces compétences sont rarement toutes disponibles au sein des administrations publiques, ou même des entreprises. Les compétences ne pouvant être mobilisées en interne, les administrations sollicitent souvent d'autres organismes à travers des marchés publics pour

⁶⁶ Loi n° 2015-991 du 7 août 2015 portant nouvelle organisation territoriale de la République

la réalisation de leur bilan de matière, comme en Bourgogne (Alterre Bourgogne 2013), en Alsace ou en Lorraine (voir Annexe 1).

La question (politique) est de savoir si les collectivités et services de l'État doivent intégrer en interne les compétences liées à la comptabilité matière. Certaines municipalités ont déjà franchi le pas sur la question de l'énergie, en intégrant des économistes de flux dans les services techniques, dans le but de réduire les consommations. Élargir cette question à la matière en général reste cependant complexe et les acteurs concernés sont nombreux. Est-ce aux régions, aux départements, aux zones d'activités, aux communes, aux DREAL, à l'ADEME, aux CCI ou à d'autres de prendre la charge de la structuration d'un pôle d'expertise sur la question matérielle ? Actuellement, la réponse n'est pas encore formelle et les acteurs impliqués dans les démarches d'économie circulaire diffèrent selon les territoires⁶⁷.

VI. Conclusion du deuxième chapitre et perspectives

En réalisant un panorama des applications possibles des analyses de flux de matières et d'énergie, nous avons vu qu'il s'agissait d'un outil particulièrement polyvalent qui peut être utilisé pour aborder différentes questions de durabilité. En aidant à appréhender certaines réalités matérielles, les AFME peuvent contribuer à définir les stratégies pour s'adapter aux évolutions présentes et à venir (juridiques, économiques, logistiques, exigences des consommateurs, disponibilité des ressources...). Elles constituent ainsi un véritable outil pour la gouvernance des organisations et des territoires. Toutefois des contraintes organisationnelles, techniques et réglementaires restent à lever avant que les AFME ne soient généralisées. Plusieurs chantiers nous semblent nécessaires pour cela :

- Poursuivre le travail de normalisation des méthodes et des règles comptables à la fois pour l'échelle macro, à partir du travail d'Eurostat pour les États, mais décliné à des niveaux inférieurs (régions, villes, zones d'activité...) jusqu'au niveau des entreprises et des produits.
- Favoriser le partage des données pour suivre la matière tout au long de son cycle de vie. Développer les infrastructures permettant de collecter, croiser et analyser plus rapidement les données sur les flux et leurs impacts, tout en considérant pleinement les exigences de confidentialité.

⁶⁷ Au niveau régional, la démarche d'économie circulaire était pilotée à l'origine par l'ADEME et le Conseil Régional en Alsace, et par la DREAL et le Conseil Régional en Lorraine (Région Alsace et Ademe, 2015).

- Poursuivre les recherches sur la restitution des résultats des études, sur le fond et sur la forme. L'objectif est de rendre accessibles les informations permettant aux acteurs du territoire d'appréhender les enjeux matériels et énergétiques qui les concernent pour les inclure dans leur stratégie opérationnelle.
- Développer des pôles d'expertise « matière » rassemblant les compétences requises pour accompagner les acteurs du territoire dans leurs études de flux et dans le déploiement des synergies territoriales.

La multiplication récente des démarches d'économie circulaire et les exigences environnementales de plus en plus contraignantes permettent de penser que les AFME vont continuer à se développer. Toutefois, elles restent avant tout descriptives et sont utiles pour élaborer ou suivre une stratégie visant à un meilleur usage des ressources. Mais la mise en œuvre effective d'une telle stratégie nécessite en plus la volonté et la collaboration des acteurs du territoire, ce qui pose des difficultés d'une tout autre nature et fait l'objet de recherches spécifiques.

Chapitre 3 Éléments complémentaires du traitement des données pour l'étude du métabolisme territorial

Dans les chapitres précédents, nous avons présenté distinctement des éléments liés au traitement de données et d'autres liés à l'analyse des flux physiques, entre autres à l'échelle des territoires à travers le paradigme du métabolisme territorial. Dans ce chapitre, nous présentons les éléments plus spécifiques qui émergent à l'intersection des deux thématiques.

Ainsi, nous allons examiner comment différents aspects des analyses de métabolisme dépendent des données et de leur traitement. Nous verrons les implications de la définition du périmètre de l'étude, puis d'autres éléments spécifiques au traitement des données. Nous terminerons en proposant quelques techniques qui nous paraissent propices à la production de l'information.

Nous nous plaçons dans une approche essentiellement comptable, dans le prolongement des travaux du CGDD (2014). Ces travaux sont complémentaires à d'autres approches plus modélisatrices qui cherchent à faire le lien avec les analyses de cycle de vie ou les tableaux entrées/sorties (Myers et al. 2019), mais au prix d'un besoin de formalisme des données sources qui n'est pas celui des données habituellement disponibles. Notre approche peut cependant faciliter le traitement des données comptables initiales pour les rendre utilisables par ces autres approches. Par ailleurs, le guide du CGDD est destiné à des territoires français, et cela influence significativement nos travaux. En effet, la disponibilité des données et leur utilisabilité peuvent fortement varier d'un pays, voire d'une région à l'autre.

I. Les usages, implications et contraintes dans la définition du périmètre

Le périmètre d'une étude de métabolisme définit habituellement l'espace géographique, le temps et les matériaux considérés (Brunner & Rechberger 2003 ; Myers et al. 2019). Pour être plus précis et plus complet, il s'agit en fait de définir (Pauliuk et al. 2019) :

- Le territoire étudié et le niveau de détail spatial envisagé
- La période d'étude et le pas de temps (annuel, mensuel...)
- Les matières (ou énergies) considérées et les nomenclatures de classification
- La ou les unités de quantification
- Le cas échéant, les processus spécifiques considérés

La définition du périmètre a des incidences sur l'ensemble de la chaîne de traitement de données (Chapitre 1/I.3) : le choix des sources de données, la façon de les préparer, les méthodes d'analyses de données qu'il est possible de mobiliser... Afin de compléter notre état de l'art, nous allons examiner les usages, les implications et les contraintes de la définition de ces différents éléments du périmètre au regard du traitement des données.

1) Le territoire étudié et le niveau de détail spatial envisagé

a) Sur la disponibilité des données et la complétude de l'information

Les études de métabolisme peuvent être menées à différentes échelles : entreprise, ville, région... (Chapitre 2/III.3). Toutefois, la disponibilité des données sources est cruciale pour en assurer la faisabilité. En France, les données territoriales qui nous intéressent sont généralement disponibles à un niveau départemental ou régional (CGDD 2014). Étudier d'autres territoires est donc de fait bien plus complexe. Certains travaux tentent tout de même d'étudier des territoires plus petits comme des villes, par exemple Montpellier (Montero 2013) ou Rennes (Bahers & Giacchè 2018) ou définis autrement que par un découpage administratif (Herbelin 2018). Cela nécessite toutefois de devoir naviguer et composer entre les différents niveaux administratifs et leur hiérarchisation, et de mobiliser des techniques de réduction d'échelle⁶⁸, en anglais *down-scaling* (Zhang et al. 2015 ; Courtonne 2016).

Dans l'approche du bilan de matière entrée/sortie, définir le territoire global est suffisant. Toutefois, lorsqu'on veut pousser l'analyse pour une meilleure compréhension des dynamiques internes, il convient également de définir le niveau de détail envisagé. Par exemple, il est

⁶⁸ La réduction d'échelle consiste à utiliser des données disponibles à un niveau fin pour en estimer d'autres à la même échelle à partir de données qui ne seraient disponibles qu'à de plus grandes échelles. Par exemple, il est possible d'estimer la production agricole d'une commune à partir de la production départementale (les données n'étant pas disponibles à un niveau plus fin) et de la surface agricole (disponible pour chaque commune dans le Registre parcellaire graphique).

possible d'étudier les dynamiques internes d'une région en s'intéressant aux flux entre départements, ou bien entre communes. À nouveau, la faisabilité de l'étude à un niveau de détail désiré dépend fortement de la disponibilité des données à ce niveau.

De manière un peu particulière, les méthodes ascendantes se basent sur des données très fines concernant les principaux sites (principales industries, bâtiments ou équipements) pour évaluer le métabolisme global d'un territoire (Brunner & Rechberger 2003). La difficulté dans ce cas est de définir ce que l'on entend par les *principaux* sites d'un territoire, les données fines n'étant ici que partielles et dans des proportions difficiles à évaluer au regard du territoire entier.

b) Sur l'évolution des territoires

Outre la question de la disponibilité des données pour le territoire étudié, son évolution peut également venir complexifier l'étude de son métabolisme. En effet, les limites administratives ont tendance à changer avec le temps. L'application de la loi NOTRe qui redéfinit le découpage régional en France depuis 2015 en est un exemple concret, avec toutefois la particularité de n'être ici qu'un regroupement de certaines régions. Dans ce cas, les anciennes données statistiques peuvent toujours être agrégées pour correspondre au découpage contemporain. L'inverse, c'est-à-dire désagréger les données contemporaines pour correspondre à l'ancien découpage, n'est cependant pas toujours possible. Ainsi, l'étude réalisée par Région Alsace & Ademe (2015) risque d'être difficile à réactualiser dans sa totalité, maintenant que l'Alsace a fusionné avec la Lorraine et la Champagne-Ardenne pour former la nouvelle région Grand Est⁶⁹.

L'application de la loi NOTRe est un cas simple : au lieu de regroupements, des redécoupages peuvent avoir lieu, que ce soit à la suite de conflits militaires ou des choix politiques. En France par exemple, la construction des Établissements Publics de Coopération Intercommunale (EPCI) semble être en constante évolution. L'Eurométropole de Strasbourg, qui est un de nos terrains d'étude, a par exemple intégré cinq nouvelles communes début 2017.

⁶⁹ Une conséquence supplémentaire est qu'il n'est plus possible de retrouver en ligne le document de communication portant sur cette étude de métabolisme, probablement à cause de la mutualisation des services d'hébergement internet qui a entraîné une réorganisation de l'information.

En travaillant à un niveau de détail encore plus fin, les aménagements urbains viennent régulièrement ajouter ou modifier des rues ou leur nom, ce qui peut poser des problèmes dans le traitement des données (voir Chapitre 5)

Suivre des territoires qui évoluent peut ainsi devenir un véritable casse-tête, à la fois pour définir un périmètre que l'on peut comparer à différents moments (étant entendu qu'il faut comparer ce qui est comparable...), mais aussi pour identifier et rassembler les données qui correspondent à ce périmètre.

c) Sur l'agrégation spatiale des données

En plus des problèmes d'évolution, la définition d'un système par ses limites géographiques génère une typologie de problème bien particulier de biais statistique : le problème d'agrégation spatiale (*Modifiable Areal Unit Problem* - MAUP). Il souligne que la manière de découper un territoire peut avoir une influence sur les données statistiques et l'interprétation que l'on en fait. Il peut rendre inappropriée la comparaison entre des territoires sur certains aspects. Nous approfondissons bien plus en détail cette question de l'agrégation spatiale dans le Chapitre 7 à travers l'exemple concret des flux de marchandises.

2) La période d'étude et le pas de temps

a) Sur la disponibilité des données

La plupart des études identifiées quantifient le métabolisme d'un territoire sur une année donnée. Dans l'étude menée par Région Alsace & Ademe (2015), la faible disponibilité des données a même contraint à rassembler des sources d'années différentes. Quelques travaux étudient l'évolution des flux sur plusieurs années (Courtonne 2016), éventuellement dans une démarche prospective (Rosado et al. 2014). Cela permet entre autres de consolider les données d'une année donnée. Ces études pluriannuelles sont potentiellement exposées à la problématique de l'évolution des territoires mentionnée précédemment.

Le pas de temps utilisé est généralement d'un an, notamment en raison des nombreuses données statistiques qui utilisent ce référentiel. Ce pas de temps est donc souvent confondu avec la période d'étude. Toutefois, il est souvent utile de pouvoir utiliser des pas de temps plus réduits (mois ou jour) pour mettre en évidence des phénomènes saisonniers, si tant est que les données sources le permettent. Nous illustrons cela dans le Chapitre 5.

b) Sur les données en temps réel et prospectives

Dans nos travaux, nous nous inscrivons plutôt dans l'analyse *a posteriori* des données de flux, sur une approche essentiellement comptable. Il est toutefois possible de vouloir étudier non pas les flux passés, mais aussi les flux présents ou futurs. Cela demande toutefois de mobiliser d'autres types d'approches : la supervision en temps réel ou les modèles prospectifs.

La technologie rend possible la connaissance en temps réel de certains flux (Shahrokni et al. 2015). Cela permet de suivre des consommations énergétiques de bâtiments, de détecter des dérives de consommation, de piloter les moyens de production électrique ou d'optimiser les flux de transport de personnes ou de marchandises⁷⁰.

Les démarches prospectives demandent quant à elles de formuler des hypothèses et d'élaborer des modèles, avec un résultat évidemment incertain en raison de son caractère prédictif. La démarche n'en reste pas moins intéressante, notamment pour aider à mieux identifier les facteurs influant sur le métabolisme. Les travaux de Rosado et al. (2014) tentent ainsi de modéliser l'évolution de la ville de Lisbonne dans les années futures.

c) Sur l'agrégation temporelle des données

Nous l'avons mentionné pour l'agrégation spatiale, mais nous pouvons également le constater sur l'agrégation temporelle des données : la manière de découper le temps peut avoir une influence sur les données statistiques et l'interprétation que l'on en fait. Nous illustrons cela Figure 6 : contrairement à l'apparente différence, ce sont les mêmes données qui sont représentées et seul le découpage temporel (par jour ou par mois) de l'information diffère. Cela est bien sûr lié au plus court mois de février qui vient modifier les perspectives. Ce *Modifiable Temporal Unit Problem* peut s'avérer également être source de mécompréhension, mais certainement dans une moindre mesure comparée au MAUP.

⁷⁰ On retrouve même sur Internet des informations en temps réel sur le trafic aérien : <https://flight-radar.eu/fr/suivi-des-vols-en-temps-reel/> ou le trafic maritime : <http://www.pilotage-maritime.nc/ais.php>.

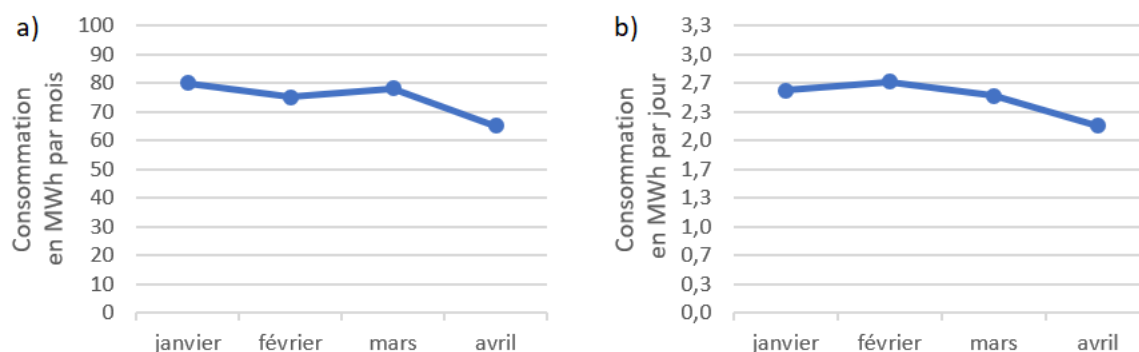


Figure 6 : Illustration du Modifiable Temporal Unit Problem à travers la consommation d'un bâtiment fictif : les données représentées en a) et b) sont les mêmes malgré l'apparente différence.

3) Les matières considérées et les nomenclatures utilisées pour classifier l'information

a) Sur la diversité des nomenclatures et leur conversion

Le choix des matières étudiées est fixé par le périmètre. Dans la méthodologie du CGDD (2014), l'objectif est d'avoir une vue d'ensemble sur toutes les matières mobilisées par une économie, mais cela n'est pas toujours le cas. Certaines études se concentrent en effet sur certaines matières spécifiques (Myers et al. 2019 ; Chapitre 2/III.3).

Souvent, une nomenclature est utilisée pour décrire les matières avec plusieurs niveaux de détail. Pour les analyses territoriales, la nomenclature proposée par la méthodologie *Economy-Wide Material Flow Accounting* d'Eurostat (2001) sert souvent comme référence, nous l'appellerons *Nomenclature EWMFA*. Elle est d'ailleurs reprise (et traduite) dans le guide du CGDD⁷¹. Elle classe hiérarchiquement les flux selon leur composition matérielle en quatre grandes catégories : la biomasse, les métaux, les minerais non métalliques et les ressources d'origine fossile. D'autres nomenclatures pour l'étude de flux sont également parfois utilisées comme la *Classification des Flux en Écologie Industrielle et Territoriale* (CFEIT)⁷² proposée par le Réseau des Chambres de Commerces et d'Industries de France.

⁷¹ Cette nomenclature est mise à disposition sous forme de tableur par le CGDD : <https://www.statistiques.developpement-durable.gouv.fr/comptabilite-des-flux-de-matieres-dans-les-regions-et-les-departements-guide-methodologique>, également disponible en version originale <https://ec.europa.eu/eurostat/web/environment/methodology> (Economy-wide material flow accounts 2018 questionnaire)

⁷² http://www.actif.cci.fr/sites/default/files/Documents/CCI_Nomenclature_EIT_public_v1.pdf

Toutefois, les données mobilisées sont rarement exprimées dans la nomenclature de destination. Par exemple, les statistiques de transport de marchandises sont exprimées dans la *Nomenclature uniforme des marchandises pour les statistiques de transport 2007* (NST2007)⁷³ qui classe les produits en lien avec l'activité économique productive : produits de l'agriculture, produits chimiques, meubles.... Les statistiques agricoles françaises, que l'on peut retrouver sur le site Agreste⁷⁴ reposent quant à elles sur une autre nomenclature : la *Classification des Produits Française* (CPF). Nous pourrions lister d'autres nomenclatures qui peuvent figurer dans les données sources mobilisées comme la *Classification statistique des produits associée aux activités* (CPA) ou la *Nomenclature Combinée* (CN) dont une nouvelle version apparaît chaque année (Eurostat 2013a).

En raison de la diversité des nomenclatures, il est très souvent nécessaire de convertir les produits d'une nomenclature vers une autre, notamment celle qui sert de référence pour l'étude (habituellement EWMFA). Il est possible de trouver des tables de conversion entre différentes nomenclatures dans la littérature. Le CGDD et Eurostat en proposent certaines dans leurs guides méthodologiques, par exemple celle permettant de passer de la NST2007 à la EWMFA. Si certains intitulés se retrouvent dans plusieurs nomenclatures, souvent la conversion d'une nomenclature doit faire l'objet d'une approximation. Il est même parfois impossible de réaliser correctement la conversion, la classification d'origine utilisée ne correspondant pas aux objectifs de l'étude. Par exemple, l'item 11.44 de la NST2007 « Parties d'appareils d'éclairage » n'est pas suffisamment explicite pour permettre une conversion selon la composition matérielle dans l'esprit de la nomenclature EWMFA. L'existence d'une catégorie « Autres/Divers/Indiscernables » est ainsi difficilement évitable. Dans certains cas, il est également approprié de convertir un intitulé dans une nomenclature vers plusieurs intitulés d'une autre nomenclature, avec l'application de différents coefficients de passage. Par exemple, dans les études des cycles biogéochimiques, les masses de produits courants (ex : 1 kg de maïs) sont converties en azote (~ 15 g d'azote ; Le Noé et al. 2016) ou en phosphore (~ 0,5 g).

⁷³ Conformément au Règlement (CE) No 1304/2007 de la Commission du 7 novembre 2007, qui établit la NST2007 comme « nomenclature unique des marchandises transportées dans tous les modes de transport concernés » (mer, voie navigable, rail, route). Il s'agit d'une mise à jour de la précédente version appelée NST-R (Nomenclature pour les statistiques de transport *Révisée*), que l'on a pu également retrouver dans d'anciennes statistiques.

⁷⁴ Agreste, la statistique agricole est le site du Service de la statistique et de la prospective (SSP) du Ministère de l'agriculture et de l'alimentation

b) Sur la qualité de l'information et son ancrage dans les choix politiques

Le travail de classification et de conversion peut engendrer une perte significative dans la qualité de l'information (Desrosières 2008)⁷⁵. Le CGDD (2014, p.73) insiste sur le fait que la méthodologie « repose sur l'exploitation de nomenclatures préétablies qui n'ont pas été conçues initialement pour la réalisation de bilans de matières ». Une étude exploratoire menée sur l'analyse du métabolisme à travers le cycle de vie des produits s'est d'ailleurs heurtée à cette limitation sans réussir à la dépasser (Ribon 2017). Les statistiques sur le transport de marchandises, pourtant essentielles dans les études de métabolisme, sont ainsi intrinsèquement limitées pour appuyer pleinement les démarches d'écologie territoriale qui nécessitent une information spécifique.

Le changement de nomenclature vers un nouveau référentiel qui tienne compte de ces spécificités permettrait d'élargir les possibilités d'usage des données statistiques. Mais cette nomenclature étant harmonisée au niveau européen, ce changement, s'il a lieu, risque de prendre du temps. Une alternative serait d'adapter les enquêtes statistiques (au niveau national donc) pour produire les données à la fois dans le référentiel de la NST2007, et dans celui d'une écologie territoriale. Une troisième voie serait que les régions, qui disposent de compétences importantes pour la transition écologique, créent leur propre système d'information sur l'utilisation des flux de ressources (CGDD 2014) et l'alimentent en partie. Notons toutefois que dans ces trois cas envisagés, l'amélioration des statistiques pour soutenir l'écologie territoriale repose sur des choix règlementaires et donc politiques.

c) Sur la désignation libre d'un produit

Dans certains cas, la désignation d'un produit dans des jeux de données ne repose pas sur une nomenclature définie, elle est alors librement définie⁷⁶. Cela permet de désigner aussi finement que nécessaire un produit et donc de l'identifier sans ambiguïté parmi d'autres. Par exemple, un produit pourrait être désigné comme « Acier non allié » ou « Acier X2CrNi18-9 », alors qu'il aurait pu devoir être dénommé « Acier » ou « Métal ferreux » dans une nomenclature préétablie. Si une désignation peut être fonctionnelle dans une entreprise particulière qui a sa manière d'identifier les produits, elle peut toutefois ne pas être exploitable dans un autre

⁷⁵ L'existence fréquente d'une catégorie « Autre/Divers/Indiscernable » dans les nomenclatures l'illustre bien.

⁷⁶ Dans le cas où c'est un nombre restreint d'intitulés qui reviennent, on peut considérer qu'ils forment une nomenclature spécifique au jeu de données.

contexte. Les intitulés libres ne permettent pas non plus de réaliser facilement et automatiquement des agrégats statistiques, il faut user d'algorithmes plus évolués pour les exploiter (Desrosières 2008 ; Davis 2017).

4) La ou les unités de quantification

a) Sur la pluralité des cadres d'analyse

Le choix de l'unité est essentiel pour s'inscrire dans différents cadres théoriques et aborder diverses problématiques qu'elles soient économiques, environnementales ou sociales (Chapitre 2/V.2). La plupart des études de métabolisme quantifient les flux en kilogramme (kg), ou en kilowattheure (kWh) pour la partie énergétique, mais ce choix peut être facilement critiqué (Van der Voet et al. 2004). Aussi, il est pertinent de ne pas se limiter à une seule unité pour la quantification des flux et des stocks et de pouvoir en étudier plusieurs (Myers et al. 2019). Évidemment, pouvoir manipuler différentes unités implique qu'elles soient disponibles dans les données sources ou qu'il soit possible de les évaluer à l'aide de coefficients ou de modèles. Cela implique également une structuration des données plus complexe pour naviguer entre les différentes unités.

b) Sur les nécessaires conversions entre unités

Quelles que soient la ou les unités retenues, il est usuel que les données sources ne soient pas exprimées dans ces unités. Par exemple, les consommations de gaz et de pétrole sont souvent exprimées en tonnes équivalent pétrole (tep). Afin de standardiser les données, des coefficients de conversion sont utilisés pour obtenir l'équivalent en kg ou kWh. Ces coefficients peuvent être soit des constantes absolues (le passage de la tonne au kg ou de la tep au kWh), soit dépendre de caractéristiques physiques (tep de gaz vers des kg) ou sociétales (kWh de gaz en euros) de l'élément étudié qui peuvent varier selon les contextes. Le changement d'unité à travers des coefficients de passage implique ainsi souvent une approximation, et donc une perte de qualité de l'information difficilement évitable, si ce n'est en évaluant les incertitudes ou en utilisant des modèles prenant en compte les conditions variables mais qui sont plus complexes à mettre en œuvre. En plus des approximations, l'application de coefficients est une démarche souvent manuelle, peu intéressante, spécifique à chaque type de produit, et qui prend du temps pour repérer ces coefficients dans la littérature et pour les appliquer adéquatement sur les données.

5) Les processus spécifiques considérés

Lors d'une étude de métabolisme territorial, l'ambition est généralement pouvoir tenir compte de tous les processus existant sur un territoire. Toutefois, et entre autres parce qu'il est souvent difficile de regrouper assez de données pour atteindre cette ambition, il est possible de ne concentrer l'étude que sur certains processus particuliers. C'est le cas de notre étude des consommations d'énergie du patrimoine bâti de l'Eurométropole de Strasbourg (Chapitre 5) : nous ne nous intéressons qu'aux bâtiments détenus par la collectivité, ignorant de fait la majeure partie du patrimoine total de la métropole. Dans un autre exemple, nous avons également étudié les marchandises transportées spécifiquement par voie fluviale (Chapitre 6).

L'étude de processus spécifiques permet d'aller plus loin dans les détails et d'exploiter des jeux de données qui peuvent être précis mais qui restent partiels. L'idéal reste de pouvoir assembler plusieurs études spécifiques pour former une image plus globale du territoire, sur une approche *bottom-up*. Cela sous-entend néanmoins que la structure des données et des résultats soit compatible d'une étude à l'autre, ce qui est rarement le cas étant donné qu'il n'existe encore pas de convention largement consensuelle pour cette structuration. Rendre les études compatibles et complémentaires peut donc être également un défi en termes de traitement de données.

6) Le risque de double comptage

Le double comptage se traduit par la prise en compte deux fois d'un flux qui ne devrait être recensé qu'une seule fois (Myers et al. 2019). Il peut être voulu, dans l'objectif de vérifier des données en les obtenant par différents moyens. Toutefois, il est plus souvent une erreur involontaire qui mène effectivement à un double comptage du flux dans les bilans. Cela peut notamment se produire lors de l'intégration de données avec des périmètres différents mais qui se recoupent en partie, ou alors lorsque la hiérarchisation des données est ambiguë⁷⁷.

En suivant la méthode du CGDD (2014), le risque de double comptage est faible. Ce risque devient plus significatif lorsqu'il s'agit de croiser des données supplémentaires, par exemple pour mener à bien des analyses de cycle de vie (Loiseau et al. 2012b ; Myers et al. 2019) ou pour explorer certaines dynamiques plus spécifiquement. Il ne semble

⁷⁷ Par exemple, des données de production agricole annoncent pour un territoire fictif : 900 t de céréales, 300 t de riz. Sans autre information, il n'est pas possible de savoir s'il s'agit de 900 t ET 300 t, ou de 900 t DONT 300 t.

pas y avoir de méthode systématique permettant d'éviter le double comptage, il convient surtout d'être précautionneux lors de l'intégration de différents jeux de données ayant des périmètres qui se recoupent.

II. Autres éléments du traitement des données

En plus des implications liées à la définition du périmètre, nous présentons dans cette section d'autres spécificités du traitement des données (Chapitre 1/II.3) appliquées aux études de métabolisme. Il s'agit des éléments liés à la collecte des données, à leur stockage et aux méthodes d'analyses de données disponibles.

1) La collecte des données

La collecte de données pour l'étude du métabolisme d'un territoire est un exercice minutieux et d'enquête à la recherche des « meilleures données disponibles » et c'est souvent cette disponibilité qui restreint les possibilités d'analyses des flux (Myers et al. 2019). La multitude des sources pose également d'autres difficultés (Chapitre 2/V) : qualité variable, difficulté d'accès, temps de préparation... Le travail à mener décrit par le CGDD (2014) est long : il nécessite de réunir des données émanant d'une cinquantaine de sources, chacune ayant des besoins spécifiques notamment pour la préparation des données. De notre expérience, très peu de ces sources présentent un haut niveau d'utilisabilité (Chapitre 2/II.3). Le niveau de détail peut être insatisfaisant, les données peuvent être mal structurées, ou l'effort à fournir pour les trouver être important. Pour les données qui ne sont pas disponibles sur Internet, la prise de contact avec les organismes qui disposent des données n'est pas toujours simple. Ceux-ci peuvent refuser de communiquer leurs données, facturer leur mise à disposition ou demander l'établissement d'une convention. Pour les entreprises privées, l'attitude habituelle est de refuser le partage des données, c'est une des difficultés principales des approches dites « *bottom-up* » (Shahrokni et al. 2015 ; Chapitre 2/III.3).

Afin de faciliter les études de métabolisme et notamment le regroupement des données, nous évoquons plusieurs possibilités (Chapitre 2/V) : généraliser les normes de la comptabilité matière, développer les infrastructures de données, et former les acteurs à mieux collecter et partager les données sur la matière qu'ils manipulent. Nous soulignons également le rôle nécessairement moteur des institutions publiques dans ce processus, rôle qui ne nous semble pas encore assez affirmé et opérationnalisé.

Pour compléter ces éléments, nous nous attardons ici sur deux autres aspects de la collecte de données : le travail de recensement des sources et l'influence de la réglementation sur la disponibilité des données.

a) Le recensement des sources de données

Le recensement des sources de données disponibles dans le guide du CGDD (2014) est un travail salubre au regard de leur nombre, près de cinquante. Nous identifions toutefois quelques limites à ce travail :

- Les sources ne sont pas présentées de manière synthétique, ce qui masque l'ampleur de la tâche et ne permet pas un accès rapide : il faut en effet entrer dans le détail du texte pour retrouver ces sources. Les liens URL permettant un accès direct aux données ne sont pas non plus mentionnés pour la plupart.
- Comme tout document finalisé, il n'est pas dynamique. Or, de nouvelles sources peuvent apparaître ou d'autres disparaître, les organismes qui produisent ces données et/ou leurs systèmes d'information peuvent changer, et avec eux la façon d'accéder aux données. Ainsi certaines références du document sont déjà obsolètes, et une série de détails dans le guide mériteraient d'être corrigés⁷⁸.
- L'échelle nationale du document ne permet pas de faire état des particularités locales. Il existe notamment des organismes à plus petite échelle (régionale, voire municipale) qui disposent également de données qui peuvent être mobilisées pour une étude de métabolisme⁷⁹.

Afin de faciliter le travail d'identification des sources de données, nous proposons un tableau synthétique accessible en ligne (repris partiellement en Annexe 1) qui reprend les sources identifiées par Barles (2014) et le CGDD (2014) avec certaines de leurs caractéristiques (producteur, contenu, URL, condition d'utilisation...)⁸⁰. Nous proposons aux contributeurs volontaires de venir participer à ce travail afin de permettre son amélioration et sa mise à jour régulière. Il vise à (1) donner une meilleure vue d'ensemble des jeux de données à mobiliser et

⁷⁸ Nous en avons recensé une partie des détails qu'il faudrait corriger à cette adresse : https://frama.link/CGDD_guide

⁷⁹ C'est par exemple le cas des associations agréées de surveillance de la qualité de l'air (la plupart connue sous le nom ATMO) qui produisent des données à l'échelle régionale et en deçà.

⁸⁰ <https://frama.link/DonneesMetabolismeTerritorial>

y faciliter l'accès notamment par un lien URL le cas échéant, (2) être actualisable pour recenser de nouveaux jeux de données ou en modifier les caractéristiques, (3) permettre de référencer des sources plus locales (régionales, départementales, voire communales).

b) L'influence du cadre réglementaire

Nous avons déjà évoqué le rôle de la réglementation dans le développement des analyses de flux de matières et d'énergie (Chapitre 2/V.4). Dans cette partie, nous examinons plus spécifiquement son influence sur la disponibilité des données qui nous intéressent. Nous identifions notamment trois types d'influences : (1) La réglementation peut obliger des acteurs à communiquer certaines informations à des organismes publics, (2) elle peut protéger et restreindre la diffusion publique de certaines données statistiques, (3) elle peut au contraire obliger la diffusion publique d'autres données. Nous retrouvons en (2) et (3) la recherche d'un difficile compromis entre la protection des intérêts privés et stratégiques par la confidentialité des données, et le développement de la connaissance et de l'innovation territoriale par leur ouverture (Chapitre 1/IV.3). Certains auteurs, dont nous rejoignons la position, soulignent que malgré les contraintes sur les intérêts économiques privés, l'impératif écologique nécessite certainement un meilleur partage des ressources, et par conséquent un meilleur partage des données (Georgeault 2015).

Pour illustrer le premier point, voici quelques exemples réglementaires d'obligations de certains acteurs à communiquer une partie de leurs informations :

- L'Autorité de régulation des transports (anciennement Arafer) dispose de prérogatives lui permettant de recueillir diverses informations utiles auprès des entreprises et organismes liés au transport de personnes et de marchandises, notamment pour alimenter un Observatoire des Transports⁸¹.
- Les propriétaires de bâtiments à activités tertiaires de plus de 1000 m² doivent renseigner leur consommation d'énergie via une plateforme numérique⁸².

⁸¹ Articles L1264-2 et L2132-7 du code des transports

⁸² Décret n° 2019-771 du 23 juillet 2019 relatif aux obligations d'actions de réduction de la consommation d'énergie finale dans des bâtiments à usage tertiaire

- Les gestionnaires de réseau énergétiques doivent transmettre leurs données de consommation et production à l'échelle des IRIS au service statistique du ministère chargé de l'énergie via une plateforme numérique⁸³.

Il ne s'agit là que de quelques exemples, il en existe en fait une multitude et cela peut d'ailleurs être assez contraignant pour ceux qui doivent produire et communiquer les données⁸⁴.

Sur le deuxième aspect lié aux restrictions dans la diffusion de données, les textes relatifs au secret statistique et au secret des affaires sont certainement ceux qui ont le plus d'impacts sur nos travaux (Chapitre 2/IV.3). De par ces textes, toutes les données qui permettent d'identifier un acteur en particulier sont par défaut protégées, c'est-à-dire, au mieux, difficiles d'accès et d'usage, mais plus souvent elles ne sont tout simplement pas communiquées. Les études de métabolisme doivent ainsi souvent se passer de ces données qui pourraient pourtant apporter une grande finesse et précision, surtout sur des petits territoires comme des villes. Une autre restriction réglementaire qui a un impact important dans nos travaux concerne la diffusion des données statistiques relatives au transport de marchandises par voie ferrée. En raison de l'ouverture à la concurrence du marché du transport ferroviaire, la diffusion des données a connu des interruptions, et il existe désormais une procédure contraignante pour accéder à ces données si l'on souhaite un niveau de détail satisfaisant, par exemple à l'échelle départementale⁸⁵.

Malgré certaines restrictions, nous constatons en même temps, et en troisièmement, des obligations réglementaires pour la diffusion publique de données. C'est par exemple le cas pour les données locales d'énergie qui sont depuis 2016 mises à disposition par le ministère de la

⁸³ Arrêté du 18 juillet 2016 fixant les modalités de transmission des données de transport, distribution et production d'électricité, de gaz naturel et de biométhane, de produits pétroliers et de chaleur et de froid.

⁸⁴ Cela fait notamment écho au témoignage de l'un de nos interlocuteurs travaillant dans le service énergie d'une municipalité : « on a l'impression que les couches du mille-feuille ne mutualisent pas leurs données ».

⁸⁵ Les dispositions des articles R1211-1 à R1211-15 du code des transports définissent ainsi l'accès aux informations relatives au trafic ferroviaire, il s'agit anciennement du décret n° 2012-555 du 23 avril 2012 relatif à l'accès de l'État, des collectivités territoriales et de leurs établissements publics à certaines informations et données sur le transport ferroviaire. Ce décret a été transposé dans les dispositions communes et abrogé en tant que tel en mai 2014. Il est malgré tout encore cité dans différents documents récents émanant des services ministériels, par exemple : <https://www.statistiques.developpement-durable.gouv.fr/sites/default/files/2018-11/datalab-essentiel-136-transport-ferroviaire-2016-fevrier2018.pdf>

transition écologique et solidaire (MTES 2018)⁸⁶. D'autres données comme les émissions polluantes des principaux sites industriels en France sont aussi rendues publiques⁸⁷.

Ces différents exemples nous permettent de souligner l'influence significative de la réglementation (et donc du politique qui l'élabore) dans la faisabilité des études de métabolisme. Cette influence peut être perçue aussi bien contraignante (obligeant à un travail de compilation et de communication de données pour certains acteurs, limitant la connaissance accessible pour d'autres, ou encore fragilisant une position compétitive d'une troisième catégorie), que favorable au développement territorial.

2) Le stockage des données

Les questions liées au stockage des données se posent à différentes étapes du traitement des données (Chapitre 1/II.3). Ainsi, pour les études de métabolisme, comme pour d'autres, le stockage des données sources (leur mise à disposition et leur structuration), intermédiaires (collectées et préparées), et finales (la mise à disposition des résultats de l'étude) est un aspect important du traitement des données.

Comme nous le mentionnions précédemment, les données sources sont dispersées. Une grande partie des données sources disponibles sur Internet le sont sous forme de tableurs, en particulier sous forme de liste d'enregistrements de flux pour différents territoires et produits (production agricole, transport de marchandises, émissions polluantes...). Certaines données sont également accessibles via une interface de programmation (API). Dans certains cas toutefois, les données ne sont pas correctement structurées, obligeant par exemple à devoir analyser des fichiers PDF ou à rassembler manuellement des données isolées. Dans ce contexte, le renforcement de l'utilisabilité des données (Chapitre 1/II.3) et le développement de système d'information régionaux (Chapitre 2/V.1) et des plateformes de données ouvertes semblent être des approches pertinentes pour structurer et favoriser l'utilisation de ces données sources (Morris 2016).

⁸⁶ Données disponibles sur : <https://www.data.gouv.fr/fr/datasets/donnees-locales-denergie/>

⁸⁷ Arrêté du 31 janvier 2008 relatif au registre et à la déclaration annuelle des émissions et de transferts de polluants et des déchets. Données disponibles dans le registre des émissions polluantes : <https://www.georisques.gouv.fr/dossiers/irep-registre-des-emissions-polluantes>

Dans une perspective de développement de plateformes de données autour du métabolisme des territoires, nous citons celle déployée par le groupe « *Metabolism of cities* » visant à regrouper les connaissances autour de la thématique⁸⁸. Elle comprend notamment un espace centralisé pour regrouper de manière standardisée les données émanant des différentes études de métabolisme à travers le monde. Nous citons également les travaux de Davis (2012) qui ont permis le développement d'Enipedia, une plateforme pour rassembler, structurer et mettre à disposition les données de systèmes énergétiques territoriaux à l'aide des techniques du web de données.

Ces plateformes de données apportent également des solutions opérationnelles pour la mise à disposition des résultats finaux. Leur usage est notamment renforcé par la dynamique de la science ouverte qui vise à améliorer l'accessibilité des études, des données et de la littérature grise. Toutefois, cette mise à disposition des données et des résultats peut aussi être perçue comme problématique dans les études de métabolisme où l'incertitude est élevée. Le choix a par exemple été fait lors de l'étude Région Alsace & Ademe (2015) de ne pas diffuser les résultats quantitatifs détaillés au public pour en éviter une interprétation stricte (voir Annexe 1).

En dehors des aspects de disponibilités des données sources et de mise à disposition des résultats, le traitement des données nécessite le stockage des données intermédiaires, souvent sur un ordinateur local et/ou sous forme de tableurs : les données sources téléchargées, les données préparées (en parfois plusieurs étapes), puis les données finales obtenues après leur analyse. La méthodologie du CGDD (2014) propose surtout, mais pas uniquement, d'analyser les données via une synthèse globale. Ainsi, un tableur préformaté accompagne le guide méthodologique et vise à faciliter la structuration de cette synthèse des flux sur le territoire selon la nomenclature EWMFA. Toutefois, si un tableur est commode pour télécharger, visualiser, manipuler des données (Chapitre 1/III.2), il reste limité. D'une part, la complexité de la réalité ne rentre pas facilement dans des tableaux en deux dimensions. Il faut ainsi opérer des choix de structure qui réduisent la complexité et donc la subtilité des informations retenues. D'autre part, la rétro-analyse de la synthèse des données, c'est-à-dire comprendre comment les données sources ont permis de produire les données finales, peut être fastidieuse. Dans la suite de nos travaux (Chapitre 4), nous proposerons un modèle pour le stockage de données afin de dépasser ces limites.

⁸⁸ <https://metabolismofcities.org/resources/data/datasets>

3) Méthodes d'analyse des données du métabolisme

Il existe un grand nombre de méthodes d'analyses de données, y compris pour l'analyse du métabolisme territorial. Sans prétendre être exhaustif, il nous semble important de présenter celles qui nous apparaissent les plus communes : la production d'indicateurs à partir du bilan entrées/sorties, les analyses spatiales et les approches graphiques. Ces méthodes ne sont pas forcément exclusives entre elles, les analyses spatiales et les indicateurs pouvant prendre une forme graphique. Elles ont toutefois des spécificités qu'il peut être intéressant de mobiliser selon l'application et le public visés. Ces différentes approches peuvent également être intégrées dans des outils d'aide à la décision, à l'image des travaux de Chrysoulakis et al. (2015) ou de Courtonne (2016), mais ces outils sont plutôt destinés à un public initié et de décideurs (élus, architectes, ingénieurs...) plutôt qu'à un public large.

a) La construction d'indicateurs autour du bilan entrées/sorties

La construction d'indicateurs est une approche fréquente pour le suivi de politiques territoriales. Au regard du métabolisme, les indicateurs courants sont ceux construits à partir du Bilan Entrées/Sorties (BES) (voir Chapitre 2/III.3 ; voir également Figure 5 p. 73 et Tableau 17 p. 202), même si selon le périmètre de l'étude et la compréhension de la notion de métabolisme, un grand nombre d'autres indicateurs peuvent s'y retrouver, par exemple la qualité de l'air, les émissions de GES... (Chrysoulakis et al. 2015)

Le guide du CGDD (2014) explique comment construire 17 indicateurs du BES et comment les interpréter. Il s'agit essentiellement d'une approche comptable des flux, c'est-à-dire procédant par leur classification et l'addition des quantités en tonnes. Éventuellement certains indicateurs peuvent être combinés entre eux pour en produire de nouveaux, comme l'*Intensité Matières* qui rapporte les quantités de matières mobilisées à la création de valeur ajoutée engendrée (PIB). Ces usages et combinaisons de différents indicateurs sont un vaste champ des possibles et posent à chaque fois la question de l'interprétation que l'on peut tirer du résultat. En simplifiant la réalité, les indicateurs la rendent plus accessible mais évidemment plus approximative. Dans une certaine mesure, ils permettent également une vision subjective du territoire, en retenant et combinant les indicateurs selon différentes priorités. Cette subjectivité qui peut être également assumée justement pour faciliter l'accès à une compréhension personnalisée du territoire (Chrysoulakis et al. 2015).

Dans le cas du métabolisme, les indicateurs permettent de présenter le bilan global d'un territoire et donner un ordre de grandeur des quantités de matières mobilisées pour le fonctionnement d'une économie. Ils peuvent également servir à suivre l'évolution de territoires dans le temps, ou à comparer différents territoires. Dans ces deux cas cependant, la méthode de traitement des données doit être identique d'une période ou d'un territoire à l'autre pour obtenir des résultats comparables. C'est difficilement le cas, tant les hypothèses et les données à mobiliser sont nombreuses, et les statistiques institutionnelles différentes, en particulier d'un pays à l'autre. D'autres phénomènes statistiques, en particulier le MAUP (voir Chapitre 7), viennent également complexifier cette comparaison entre différents territoires. Aussi, malgré la forme commode des indicateurs pour présenter une vision synthétique du métabolisme d'un territoire, leur construction, leur comparaison et leur interprétation sont exposées à différents biais. Il convient ainsi d'être attentif dans leur utilisation et notamment de ne pas faire de ces indicateurs une fin en soi (Chapitre 1/IV.5).

b) Les analyses spatiales

Un type d'analyse spatiale du métabolisme consiste à identifier les zones d'approvisionnement et la destination des exportations : d'où /de quelle distance viennent les matières que nous mobilisons ? Où exportons-nous ? Ces interrogations permettent de mettre en exergue les spécificités de certains territoires exportateurs en raison par exemple du modèle agricole, de leur capacité d'extraction de matières premières, de la production d'énergie... Nous pouvons ainsi étudier les relations matérielles entre différents territoires et leurs interdépendances. Elles peuvent aussi permettre d'aborder la question de la résilience d'un territoire vis-à-vis de l'approvisionnement de certaines matières (Chapitre 2/IV.9). La connaissance de l'origine des flux peut également appuyer le développement d'une économie plus locale. Notons toutefois que la distance, si elle fait sens dans le développement d'une économie locale, n'est pas forcément corrélée aux impacts environnementaux : les impacts liés au transport sont souvent faibles au regard de ceux imputables à la production des marchandises et dépendent également de leurs modes et conditions de transport.

Une grande difficulté des analyses spatiales est que les statistiques dont nous disposons reflètent mal l'origine réelle des marchandises. Nous avons bien des informations sur les lieux de chargement et de déchargement, mais les marchandises sont fréquemment chargées et déchargées plusieurs fois, pour optimiser la logistique notamment avec différents moyens de transport. L'information sur l'origine et la destination réelle des flux est alors perdue. Pour tenter d'aller vers une meilleure connaissance de cette origine ou destination réelle,

Silvestre et al. (2012) proposent notamment de calculer l'origine de rang 2 des marchandises (pour celles circulant en France), à travers un calcul matriciel incluant des données de production locale avec celles sur le transport de marchandises. Mais même des informations de rang 2 ne reflètent pas vraiment la complexe réalité des chaînes logistiques.

c) La visualisation de données

Les techniques de visualisation de données peuvent permettre de diffuser efficacement une information complexe, grâce à un assemblage structuré d'un grand nombre de données sous une forme graphique (Chapitre 1/II.3). Les études de métabolisme peuvent ainsi bénéficier de ces techniques pour faciliter l'assimilation et l'interprétation des données (Chapitre 2/IV.6). Pour favoriser cette approche, le groupe de travail Metabolism of Cities (2016) a notamment collecté près d'une centaine de propositions graphiques pour représenter le métabolisme d'un territoire, ce qui nous permet d'identifier les représentations les plus courantes.

En premier figure sans conteste le diagramme de Sankey⁸⁹. Il consiste en des flèches de largeurs proportionnelles aux flux à représenter. Cette approche graphique permet de visualiser simultanément différents types de flux, leurs quantités et leur origine et destination. Elle peut aussi servir à vérifier visuellement que la loi de conservation de la masse est bien respectée : les quantités entrantes et sortantes d'un processus peuvent être facilement comparées par l'alignement de la représentation des flux. Ces raisons rendent les diagrammes de Sankey très adaptés pour la représentation de flux, même si leur construction peut s'avérer un peu complexe pour qu'ils restent lisibles.

Une version très simplifiée du diagramme de Sankey est la représentation graphique du BES qui permet la lecture directe de quelques indicateurs (Figure 5 p. 73). Le décalage entre les entrées et sorties représente l'addition nette au stock qui équilibre le bilan. Certains travaux proposent également une représentation dynamique du BES qui donne accès interactivement à une décomposition des différents flux (Elioth 2014).

Pour d'autres indicateurs qui ne seraient pas lisibles à travers le BES, il est possible d'utiliser des diagrammes en (é)toile, ou diagramme de Kiviat. Ils peuvent servir à représenter simultanément plusieurs indicateurs (Figure 7.a), même s'il nous semble plus judicieux d'utiliser sa forme discrète (Figure 7.b) qui évite de lire une forme de continuité entre des

⁸⁹ Quelques exemples supplémentaires de diagrammes de Sankey : <http://graphicalmemes.com/gallery.html>

indicateurs bien distincts. Il est possible de superposer plusieurs de ces diagrammes pour comparer différents contextes, territoire ou année notamment.

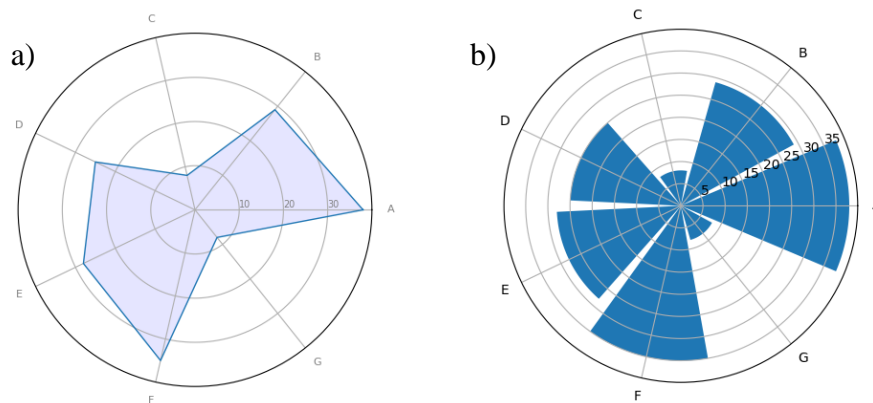


Figure 7 : Exemple de représentations d'indicateurs :
diagramme en étoile (a) et son équivalent discret (b)

Concernant les analyses spatiales, celles-ci s'appuient souvent sur des représentations cartographiques. Cela permet de mettre en évidence certains territoires par leurs spécificités ou par les flux qu'ils mobilisent (Le Noé et al. 2016 ; Figure 30 p. 183). D'autres représentations, par exemple un diagramme de Sankey pour les aires d'approvisionnement (Figure 8), ou un tableau matières/origines (Figure 34 p. 188) peuvent aussi venir questionner la distance parcourue par certaines marchandises.

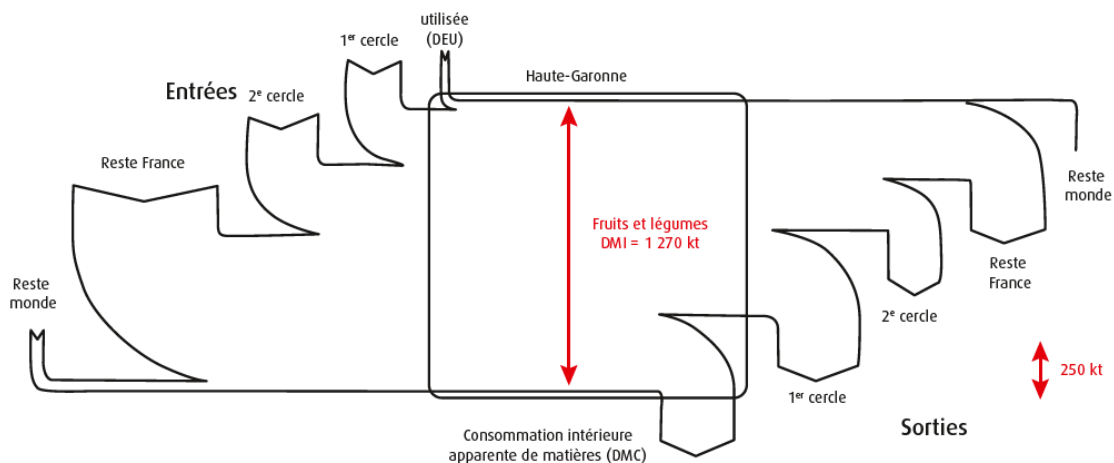


Figure 8 : Exemple de représentation de l'aire d'approvisionnement et d'exportation des fruits et légumes de Haute-Garonne (extrait de CGDD 2014)

Les quelques techniques de visualisation présentées précédemment sont des exemples avec une certaine spécificité pour les études de métabolisme. Les techniques plus classiques de représentation (nuage de points, histogramme, diagramme circulaire...) sont évidemment également disponibles pour appuyer cette représentation.

III. La production de l'information sur le métabolisme d'un territoire

1) Les axes d'analyses du métabolisme

Nous avons vu qu'il existe un large panel de méthodes d'analyse de données et de techniques de visualisation qui peuvent être mobilisées pour mieux comprendre le métabolisme d'un territoire. Cela laisse penser que cette notion peut s'aborder sous plusieurs perspectives : spatiale, temporelle, quantitative, qualitative... Il s'agit de répondre à plusieurs natures d'interrogation : où, quand, combien, quoi ? En poursuivant le fil de cette réflexion, nous reprenons naturellement les interrogations de la méthode empirique « QQQQCCP » pour Qui, Quoi, Où, Quand, Comment, Combien, Pourquoi. Cette méthode, héritée de la Rome antique à travers l'« hexamètre⁹⁰ de Quintilien » reste très utilisée aujourd'hui. Elle permet de tendre vers une complétude de l'information en précisant des circonstances (Valentin et al. 2010). Appliquées à notre objet d'étude, ces questions nous permettent en effet d'interroger de manière un peu plus systématique le métabolisme d'un territoire :

- Qui : Quels sont les acteurs qui échangent et transforment la matière ou l'énergie ?
- Quoi : Quelles sont les matières ou énergies mobilisées ?
- Où : Dans quels territoires / à quels endroits du territoire sont-elles mobilisées ?
- Quand : À quel moment / à quelle période ?
- Comment : Par quels moyens de transport ?
- Combien : En quelle quantité ?
- Pourquoi : Dans quel but ou intention ?

Étant donné nos réflexions sur l'unité de quantification (Chapitre 2/V.2), il nous semble nécessaire d'ajouter toutefois une huitième question :

- Combien de quoi : dans quelle unité les quantités sont-elles exprimées ?

Cela constitue ainsi huit dimensions ou axes pour l'analyse du métabolisme d'un territoire qui sont illustrés Figure 9. D'autres travaux proposent également différentes dimensions pour l'analyse des flux, en ajoutant des dimensions « processus » et « scénario », mais sans mentionner les acteurs, les modes de transport et les usages (Pauliuk et al. 2019). Il

⁹⁰ La question combien n'y figurait pas à l'origine.

nous semble toutefois que ces deux dimensions proposées ne sont pas aussi explicites que les autres et donc plus difficiles à conceptualiser.

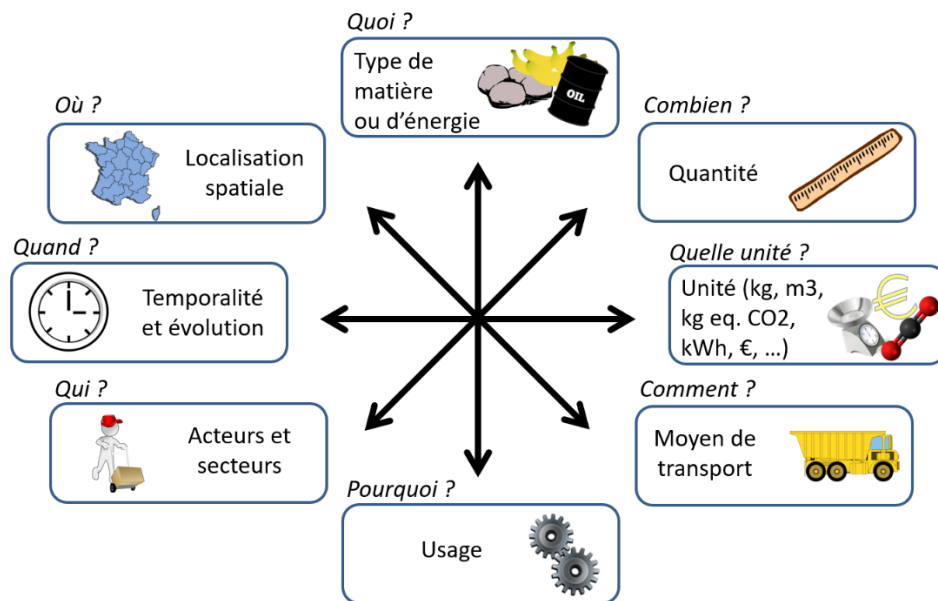


Figure 9 : Huit dimensions pour l'analyse du métabolisme d'un territoire

Les différentes dimensions peuvent être renseignées avec des degrés de finesse différents : on peut parler de métaux, ou de fer, d'argent et de cuivre ; on peut étudier les flux sur un an, sur un mois ou une semaine ; on peut les localiser dans une région, dans une ville ou selon des coordonnées GPS. Il est également possible de distinguer deux dimensions spatiales en considérant séparément l'origine et la destination des flux. Nous noterons que vouloir analyser le métabolisme sous différentes dimensions ne signifie pas disposer des données pour le faire. En fait, il est très fréquent de ne pas pouvoir obtenir de réponses à certaines de ces questions à travers les données, notamment « pourquoi ? » ou « comment ? ». La connaissance du terrain peut être ainsi indispensable pour compléter l'information.

Ces différentes dimensions peuvent être utilisées pour imaginer et structurer des problématiques, ou au contraire décortiquer des questions pour les rendre plus explicites en précisant les différentes dimensions. Par exemple, nous nous intéressons à la question des déchets alimentaires dans le Chapitre 8. De façon générale, il faut trouver les moyens de les réduire et de favoriser leur retour à la terre (Peyraud et al. 2015). La lecture de cette thématique à travers notre « octomètre » nous permet d'esquisser assez facilement et de façon systématique un certain nombre de problématiques (voir Tableau 5). Cela ne signifie pas que son usage permet de faire le tour d'une thématique, mais il aide à élargir le point de vue et à aller vers une meilleure formalisation des problématiques.

Axe d'analyse	Problématiques esquissées
Qui	Quels sont les acteurs qui produisent les déchets alimentaires, les collectent ou en permettent le retour à la terre ? (<i>ex : Les restaurateurs, les particuliers, les industries agroalimentaires, les agriculteurs...</i>)
Quoi	Quelle est la composition des déchets alimentaires ? (<i>ex : Pain, huile, épluchures, denrées emballées périmées...</i>)
Où	Où collecte-t-on les déchets alimentaires ? Ces lieux de collecte sont-ils proches des lieux de production ? Où peut-on valoriser la matière ?
Quand	Comment évolue la production de déchets valorisés ? Quel(s) jour(s) les déchets sont-ils collectés ?
Comment	Comment sont transportés ces déchets ? (<i>ex : par camion, par vélo-cargo...</i>)
Pourquoi	Pourquoi ces déchets ont été produits ? (<i>ex : en raison d'une surproduction, parce que les denrées étaient périmées...</i>)
Combien	Quelles sont les quantités produites, collectées et valorisées ?
Combien de quoi ?	Quels sont les volumes en masse ? Combien cela coûte-t-il ? Quel impact climatique par l'émission de GES ?

Tableau 5 : Exemple d'application de l' « octomètre » du métabolisme territorial pour esquisser différentes problématiques autour de la thématique des déchets alimentaires.

2) La représentation graphique du métabolisme

Considérer le métabolisme d'un territoire comme des données structurées selon huit dimensions aide également à sa représentation. En effet, les techniques graphiques permettent de visualiser les données selon plusieurs dimensions (Chapitre 1/III.2). Connaître les dimensions à explorer pour répondre à une problématique permet d'identifier plus facilement les techniques adaptées. Nous formalisons cette approche dans le Tableau 6 qui reprend différentes techniques graphiques et leur adéquation ressentie avec les différentes dimensions du métabolisme. Il s'agit toutefois d'une synthèse exploratoire et subjective qui pourrait faire l'objet d'une analyse plus approfondie.

Techniques graphiques \ Dimensions du métabolisme	Quantitative	Qualitative	Spatiale	Temporelle	Relative aux acteurs	Relative aux unités	Relative aux modes de transport	Relative à l'intention
Axe gradué (continu)	+	x	=	+	x	x	x	x
Axe non gradué (discret)	+*	+	+*	+*	+	+	+	+
Diagramme circulaire	+	+	+*	+*	+	x	+	+
Dégradé (continu) de couleurs	+	x	x	=	x	x	x	x
Palette de couleurs (discrète)	=	+	+*	+*	+	=	+	+
Cartographie	x	x	+	x	x	x	x	x
Taille des formes	+	x	x	x	x	x	x	x
Type de forme	x	+	=	=	+	=	+	=
Flèche (Diagramme de Sankey)	+	+	+*	x	+	x	+	+
Répétition de graphiques	x	+	+*	+*	+	+	+	+
Graphique dynamique	x	x	x	+	x	x	x	x

Tableau 6 : Adéquation ressentie de différentes techniques graphiques pour la représentation des différentes dimensions du métabolisme territorial.

+ : technique adaptée (+* si discrétisation) / = : technique adaptable sous conditions /
x : technique non adaptée

Si l'examen du métabolisme à travers différentes dimensions peut aider à la construction graphique, il n'en révèle pas moins la limite intrinsèque : le nombre de dimensions du métabolisme (huit) est trop élevé pour faire l'objet d'un graphique unique qui ne peut représenter qu'un nombre limité de dimensions (deux à quatre, difficilement plus). Celles qui ne peuvent pas être représentées doivent prendre des valeurs fixées. En plus du nombre de dimensions limité, chacune d'entre elles peut avoir un grand niveau de détail, obligeant à ne représenter qu'une partie des données ou à les synthétiser pour que cela puisse rester lisible sur le graphique.

Ainsi, pour représenter graphiquement le métabolisme d'un territoire, il nous semble nécessaire de combiner deux approches : restreindre la complexité du graphique en limitant le périmètre et le niveau de détail, et multiplier les représentations pour examiner le métabolisme sous différents angles. L'idée du second point est de construire une image cognitive d'un concept en 8D à partir d'un ensemble de représentation entre 2 et 4D⁹¹. En plus de construire cette image, la multiplication des visualisations peut aussi aider à révéler des erreurs ou des incohérences dans les données (Davis 2012).

Un autre point important est de remarquer que la disponibilité des représentations graphiques reste souvent dépendante de choix éditoriaux. Ces choix sont faits par les personnes qui produisent ou commandent les graphiques. Leur travail reste essentiel, ne serait-ce que pour sélectionner les représentations les plus pertinentes parmi la multitude des possibles. Elles imposent cependant le périmètre, le niveau de détail et les représentations qui correspondent à leur focale. Dans la plupart des cas, il n'est pas facilement possible de changer ces paramètres, sauf à recommencer l'étude.

Pour éviter d'imposer des choix graphiques, il faut pouvoir rendre ceux-ci interactifs : le périmètre, le niveau de détail et les représentations deviennent des choix paramétrables, laissant le lecteur (le « *visualiseur* ») libre d'explorer les données à sa manière, selon sa propre focale et ses propres considérations métier (Flichy 2013). C'est par exemple la proposition d'Elioth (2014) à travers son graphique interactif du BES de Paris et l'Île-de-France, qui permet d'accéder à différents niveaux de détails sur les produits. L'exemple reste toutefois limité dans ses fonctionnalités et les représentations graphiques interactives sur les flux de matières et d'énergie restent encore rares.

IV. Conclusion du troisième chapitre

Dans ce chapitre, nous attirons l'attention sur les différents détails et difficultés du traitement des données spécifiquement à l'étude du métabolisme des territoires. Ces aspects viennent s'ajouter aux considérations plus générales du traitement des données énoncées dans le premier chapitre.

Un des principaux éléments qui ressortent de nos réflexions est l'imbrication entre la définition du périmètre de l'étude et l'utilisabilité des données. Dans une étude de métabolisme,

⁹¹ Un peu à l'image d'une multitude de photographies 2D qui permettent de construire une représentation en 3D.

où il n'est habituellement pas envisagé de construire de nouvelles données, c'est bien l'utilisabilité des données existantes qui est le facteur limitant dans cette définition du périmètre. D'autres aspects qui s'intègrent dans les étapes du traitement des données sont également évoqués : la collecte, le stockage et les méthodes usuelles d'analyse de données. En plus de la formalisation de ces éléments, nous proposons également de renforcer la facilité à trouver les données en listant de manière synthétique les sources mentionnées dans le guide méthodologique du CGDD, travail qui pourra être complété par la suite pour intégrer d'autres sources.

Ce chapitre a également permis de s'interroger sur la nature des informations que nous voudrions pouvoir produire. Ainsi, nous proposons d'inscrire l'analyse du métabolisme selon huit dimensions dérivées des questionnements de bases (quoi, qui, combien, quand...). Cette approche permet d'interroger plus systématiquement le métabolisme d'un territoire ou certaines de ses composantes. De plus, la structuration selon différentes dimensions vient aussi soutenir sa représentation graphique qui nous paraît une approche pertinente et adaptée pour une diffusion large d'une information complexe. Cela nécessite toutefois de multiplier les graphiques accessibles, notamment de manière interactive, pour s'adapter aux diverses préoccupations des acteurs du territoire.

Nous clôturons ici un état de l'art en trois chapitres autour de la question du traitement des données dans l'étude du métabolisme des territoires. Cet état de l'art montre la pluridisciplinarité de la thématique, mais également des aspects très pragmatiques dans sa mise en œuvre opérationnelle. Ces aspects ne sont pas forcément complexes, mais ils sont nombreux et multipliés par le nombre de sources de données à traiter. Dans ce contexte, il nous semble possible et pertinent d'envisager une plus grande automatisation du traitement de ces données en grand nombre à travers un système d'information. Nous présentons cette approche dans la partie suivante.

*Partie II Proposition
conceptuelle*

Chapitre 4 Conception d'un Système d'Information pour l'Analyse du Métabolisme des Territoires (SINAMET)

Malgré le nombre croissant d'études du métabolisme de différents territoires, le traitement des données reste encore une opération délicate. Les méthodes proposées requièrent souvent un traitement manuel des données ; le CGDD (2014, p.21) indique d'ailleurs dans son guide sur la comptabilité des flux de matières qu'« il n'existe pas, à ce jour, d'outil permettant de structurer les données de façon standardisée et d'en automatiser le traitement ». D'autres travaux internationaux font également le même constat (Pauliuk et al. 2019). Plusieurs solutions logicielles dédiées à l'étude de flux existent (Harpet & Gully 2013), comme Presteo (Georgeault 2015) ou Actif⁹² pour l'écologie industrielle, OpenLCA pour l'analyse du cycle de vie des produits (Ciroth et al. 2014) ou encore STAN pour la modélisation de processus (Cencic & Rechberger 2008). Toutefois, aucune de ces solutions ne permet d'aborder la comptabilité des flux de matières et d'énergie de façon multidimensionnelle à l'échelle d'un territoire. Ainsi, faute d'outil adéquat, l'étude du métabolisme d'un territoire est un exercice long et complexe (Bahers 2012 ; Morris 2016), présente des risques d'erreur et oblige à s'interroger systématiquement sur la façon de manipuler les données.

Une plus grande automatisation des processus de traitement des données pourrait ainsi contribuer à améliorer les études de métabolisme territorial sur différents plans :

- en facilitant le développement de nouveaux processus de traitements pour diversifier l'information accessible,
- en réduisant le temps consacré au traitement des données, pour mieux l'allouer à la compréhension des informations et à l'action,

⁹² <http://www.actif.cci.fr/>

- en rendant les traitements reproductibles dans d'autres contextes, notamment temporels et spatiaux,
- en rendant les traitements transparents, à condition d'en révéler le code source,
- en réduisant le risque d'erreur lié à la manipulation des données.

Nos travaux visent ainsi la conception d'un Système d'INformation pour l'Analyse du MÉtabolisme des Territoires (SINAMET) afin d'automatiser le traitement des données pour ce type d'étude. Sa conception, comme pour d'autres systèmes d'information (SI), est envisagée comme un processus long et itératif, qui devra mobiliser de nombreuses personnes et compétences à l'intersection entre plusieurs disciplines (Chapitre 1/III.1). Il n'est ainsi pas possible d'envisager dans le cadre de cette seule thèse la réalisation d'un système pleinement fonctionnel et opérationnel, mais c'était bien notre objectif d'en formaliser les fondements et de réaliser un prototype pour valider l'approche.

Dans ce processus de conception, l'état de l'art des chapitres précédents nous permet d'explicitier un certain nombre de problématiques dans les opérations de traitement des données. Sur cette base, nous allons procéder à une analyse ontologique et fonctionnelle pour formaliser le sens de l'information que nous devons manipuler, et les fonctionnalités qui vont permettre de la manipuler. Nous justifierons ensuite les différentes approches que nous mettons en œuvre pour concevoir le SINAMET. Enfin, nous concluons en présentant le prototype qui a été réalisé, quelques résultats obtenus et ses limites.

I. Analyse du sens de l'information et des besoins de traitement

1) Ontologie du métabolisme territorial

La mise au point d'une ontologie (Chapitre 1/III.3) du métabolisme territorial va nous permettre de définir une structure pour les données qui en permettent l'étude. La notion de métabolisme est en effet trop abstraite pour être exploitable en tant que telle et il faut pouvoir la rendre opérationnelle dans un système d'information.

Une première étape dans la définition de l'ontologie consiste à identifier les concepts sous-jacents, clairs et concrets qui sont mobilisés à travers la notion générale. Pour identifier ces concepts, nous partons d'une définition que l'on retrouve assez largement dans la littérature scientifique : Le métabolisme territorial est l'ensemble des processus techniques et socio-économiques par lesquels les territoires mobilisent, consomment et transforment la matière et

l'énergie (Kennedy et al. 2007 ; Barles 2008). Bien que certains termes soient déjà plus concrets (territoires, matière, énergie), la notion de « processus techniques et socio-économiques » mérite plus de précisions. Ces processus sont ainsi les échanges, qui ne sont pas forcément matériels et peuvent être des services, et les manipulations des acteurs du territoire qui se manifestent sous la forme de flux (physiques, financiers ou d'information) ou de stocks (Erkman 2004).

Cette première analyse met ainsi en exergue onze concepts : Territoire, Mobilisation, Consommation, Transformation, Matière, Énergie, Échanges, Manipulation, Acteurs, Flux, Stocks. Les notions de « Mobilisation » et de « Manipulation » peuvent être considérées comme redondantes avec celles de « Consommation » et de « Transformation », aussi nous les écartons. La notion de consommation est rattachée à celle de flux, la consommation d'un produit pouvant être confondue avec un flux entrant. Les échanges sont également confondus avec des flux. Enfin, la matière et l'énergie sont regroupées et associées avec les concepts supplémentaires d'objets et de service sous le terme de « Produit ». Un produit peut ainsi désigner tout ce qui peut faire l'échange entre deux acteurs, sans se limiter à la seule considération matérielle.

Finalement, après regroupement, six concepts ressortent comme étant fondamentaux pour donner du sens à la notion de métabolisme territorial : Territoire, Acteur, Produit, Flux, Stock, Transformation. Nous allons poursuivre la définition de l'ontologie en identifiant pour chacun de ces concepts, leurs propriétés et les relations qu'ils entretiennent avec les autres concepts. Ces propriétés sont synthétisées dans l'Annexe 3.

a) Territoire

Un territoire est une notion assez large qui peut recouvrir différents concepts selon les disciplines (Paquot 2011 ; Laganier et al. 2002). Il peut faire référence à des dimensions matérielles (les caractéristiques géo et biophysiques et l'aménagement de l'espace par les infrastructures), organisationnelles (les acteurs sociaux et institutionnels eux-mêmes caractérisés par leurs rapports relationnels), et identitaires (l'identité du territoire, à la fois objective – limites, histoire, patrimoine culturel – et subjective dans la représentation que s'en font les acteurs). Ces aspects du territoire sont dynamiques et changent avec le temps, les territoires pouvant même apparaître ou disparaître selon la conception que l'on se fait de cette notion.

Dans l'analyse du métabolisme, la notion de territoire est nécessairement mobilisée pour trois raisons. Il faut tout d'abord pouvoir (1) interpréter des données statistiques déclinées selon

un découpage administratif. Ce découpage établit des limites géographiques légalement objectives (mais parfois contestées) qui permettent de (2) spatialiser l'information. Enfin, il est possible d'identifier les acteurs qui évoluent dans un territoire donné. Cela va ainsi permettre de (3) sélectionner et contextualiser l'information pour la rapprocher de l'environnement parfois très local de ces acteurs.

Pour rendre explicite le territoire au sein d'un système d'information, nous le définissons comme une aire géographique (une partie de la surface terrestre) communément désignée par différents acteurs, généralement par un nom. Tant que l'on ne cherche pas à cartographier l'information, et on ne cherche pas systématiquement à le faire, il n'est pas nécessaire de connaître précisément les limites de ces aires. L'imaginaire renvoyé par le nom d'un territoire et qui résulte d'une construction sociale est généralement suffisant pour identifier ce territoire et éventuellement s'en faire une idée spatiale. Cette définition permet de conceptualiser des territoires arbitraires ; il est toutefois plus simple et utile de faire référence à des territoires largement reconnus, notamment ceux formés par le découpage administratif : pays, régions, villes... jusqu'à la parcelle cadastrale et les bâtiments. En deçà, nous pouvons encore imaginer de diviser les différentes parties d'un bâtiment, même si nous ne sommes plus dans le cadre d'un découpage purement géographique mais aussi spatial.

À cette notion de territoire, il convient de définir les propriétés qui y sont associées. Pour cela, nous nous inspirons très fortement des travaux de l'Organisation des Nations Unies pour l'alimentation et l'agriculture (*Food and Agriculture Organization of the United Nations – FAO*) qui a proposé une « ontologie géopolitique » suite à un travail de collecte d'information sur les territoires⁹³. Cette ontologie, publiée en 2008, permet de structurer explicitement ces données et de faciliter leur partage (Caracciolo et al. 2006). Elle constitue ainsi une référence mondialement reconnue (Kim et al. 2013).

De manière concrète, l'ontologie distingue deux classes d'objets : les territoires et les groupes de territoires. Les territoires correspondent aux pays et nations membres de la FAO, qu'ils soient autonomes, contestés ou non. Les groupes correspondent à des ensembles géographiques, économiques ou organisationnels.

⁹³ Cela a abouti en 2002 à la création d'un portail « Profils de pays de la FAO » qui rassemble différentes statistiques nationales dans les domaines de l'agriculture et du développement : <http://www.fao.org/countryprofiles/fr/>

L'ontologie précise également les propriétés des territoires. Tout d'abord, il y a les codes de références qui permettent d'identifier des territoires de manière unique dans un référentiel. Il existe une multitude de référentiels : UN M.49, ISO-3166 Alpha-2 et Alpha-3, code PNUD, code GAUL, FAOSTAT, AGROVOC et FAOTERM, ainsi que l'identifiant DBPedia (pour l'intégration au web de données). Il existe d'autres identifiants, comme le code Insee en France⁹⁴, le NUTS utilisé dans l'Union européenne ou le FIPS aux États-Unis. Différents services, qu'ils soient publics, communs ou privés, attribuent également leurs propres identifiants à des territoires, comme OpenStreetMap, GeoNames, Wikidata, des bibliothèques nationales, Facebook, Michelin Voyages, Encyclopædia Britannica Online...

Si les codes sont des attributs pratiques pour permettre l'identification non ambiguë d'un territoire, les discussions humaines désignent essentiellement les territoires par leur nom. Cependant, l'usage d'un nom pour identifier un territoire est moins robuste. D'une part, la diversité linguistique pose problème puisque les territoires n'ont pas les mêmes noms dans les différentes langues, sans compter les problèmes d'encodage générés par les caractères accentués ou non-latin. Ensuite, il est fréquent d'avoir plusieurs territoires homonymes, notamment pour les villes (par exemple Sidney au Canada, Brest en Biélorussie), ou les EPCI (il y a en France deux « Communauté de Communes des Quatre Rivières », et avant 2018, il y avait trois « Communauté de Communes des Trois Rivières »). Enfin, un territoire peut posséder un nom court et un nom officiel dans une même langue (« Russie » ou « Fédération de Russie »).

La FAO intègre également comme attributs des territoires leurs coordonnées géographiques (latitude maximale, latitude minimale, longitude maximale, longitude minimale), des données statistiques de base (superficie du pays, superficie des terres, terrains agricoles, population et PIB), ainsi que les informations sur les noms et codes de leurs devises et les adjectifs de la nationalité. Les territoires sont aussi repérés temporellement avec leur période d'existence. Enfin, l'ontologie de la FAO présente aussi trois types de relations entre territoires : l'appartenance d'un territoire à un groupe de territoires, les territoires frontaliers et la succession historique d'un territoire à un autre, comme la Russie qui succède à l'URSS.

⁹⁴ Le code Insee pose toutefois un problème puisqu'il associe un code à deux chiffres aux départements, mais également aux régions. Par exemple, le code 44 correspond à la fois à la région Grand Est et au département Loire-Atlantique. Pour éviter les ambiguïtés, nous faisons habituellement précéder les codes régions par R (RX pour les régions d'avant 2016)

Dans la stratégie de la définition d'une ontologie, c'est-à-dire dans l'objectif de servir de référence partagée, il nous paraît nécessaire de réutiliser au maximum des ontologies existantes et reconnues. Toutefois, certains ajustements nous paraissent nécessaires. D'une part, nous ne distinguerons pas les territoires des groupes de territoires. De notre point de vue, un groupe de territoire est lui-même un territoire. Cette approche nous permet de décrire le monde par une hiérarchie de territoires en évitant les problèmes sémantiques liés au regroupement : la France regroupe 18 régions, s'agit-il d'un territoire ou d'un groupe de territoires ? Ainsi, dans notre modèle, la notion de groupe n'existe pas, mais les territoires peuvent s'intégrer dans d'autres territoires plus vastes. Afin d'organiser cette hiérarchie, il nous semble nécessaire de disposer d'un attribut précisant l'échelle du territoire (ville, département, région, pays...). Par ailleurs, nous proposons également d'ajouter deux informations géographiques : les coordonnées du centroïde et les frontières délimitant ce territoire.

Pour résumer, nous avons défini le territoire simplement comme une aire géographique. Un territoire possède des attributs d'identification (codes et noms) dont au moins l'un des deux est nécessairement indiqué (les codes étant à privilégier pour éviter les ambiguïtés). Les territoires possèdent également des attributs géographiques (frontières, centroïde, latitudes et longitudes extrêmes) et un large panel d'attributs de caractérisation (échelle, surface, population, PIB, période d'existence...). Les territoires entretiennent également des relations avec d'autres territoires (frontières communes, relation d'inclusion et succession historique).

b) Acteur

La notion d'acteur est polysémique. Pour nos réflexions, il s'agit de pouvoir désigner les personnes du territoire qui manipulent des ressources, on pourrait à ce titre parler d'agents (Shahrokni et al. 2015). Il peut s'agir de personnes morales ou de personnes physiques. Pour éviter d'avoir à complexifier notre ontologie, et de manière similaire aux territoires, nous utilisons également le terme d'acteur pour pouvoir en désigner un groupe, par exemple un secteur d'activité.

La caractérisation d'un acteur présente plusieurs similarités avec celle d'un territoire. Ainsi, les acteurs peuvent être identifiés par des codes (comme le numéro SIRET ou le code de la Nomenclature d'activités française - NAF), ou par un nom (éventuellement dans différentes langues). Ils peuvent être géolocalisés, avec une adresse ou des coordonnées. Ils ont aussi une période d'existence et peuvent être caractérisés par de nombreux attributs : le nombre d'employés pour une entreprise, son chiffre d'affaires, son secteur d'activité... Les acteurs

peuvent être liés entre eux (un groupe peut posséder plusieurs succursales) ou se succéder. Enfin, ils sont généralement ancrés sur un ou plusieurs territoires, par exemple une collectivité qui exerce ses compétences dans ses limites administratives, ou même simplement une entreprise implantée sur un terrain.

Nous identifions donc plusieurs types de propriétés pour les acteurs : des attributs d'identification (codes et noms), de géolocalisation (adresse, coordonnées), un large panel d'attributs de caractérisation (période d'existence, nombre d'employés, chiffre d'affaires...), et des relations avec d'autres acteurs (hiérarchie, succession) et des territoires.

c) Produit

De manière formelle, nous définissons un produit comme étant l'objet d'un échange entre des acteurs d'un territoire. Dans cette approche, ce terme peut ainsi tout aussi bien désigner des objets, la matière qui les compose, de l'énergie ou des services. Même si l'étude du métabolisme est avant tout une approche matérielle et énergétique, les services contribuent au métabolisme du territoire et doivent donc pouvoir être également considérés. Finalement, cette définition inclut un large éventail de concepts : du minerai brut, des objets finis, des déchets, de l'électricité, des services rendus, le transport de marchandises, des transactions financières, des échanges de données... Nous nous focalisons toutefois dans nos travaux sur les produits physiques qui constituent directement le métabolisme d'un territoire.

Afin de pouvoir désigner le produit, un nom, qui peut être libre, est nécessaire. Afin de pouvoir classifier l'information statistique, le nom du produit doit s'inscrire une nomenclature qui dresse une liste limitée et hiérarchisée de possibilités (dans différentes langues), souvent en association avec un code (voir par exemple la nomenclature NST2007 en Figure 10).

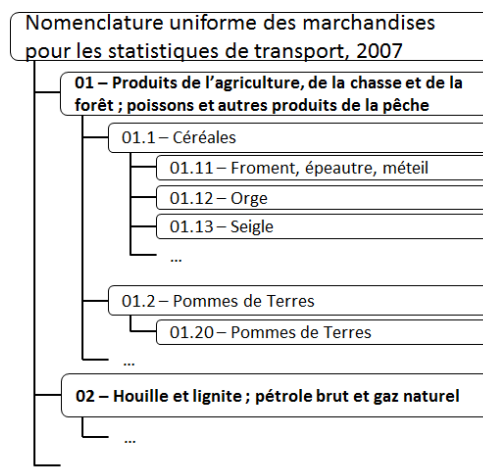


Figure 10 : Extrait de la structure de la nomenclature NST2007

Les produits peuvent également être caractérisés par des valeurs physiques ou économiques comme leur durée de vie ou des coefficients qui vont permettre de changer l'unité de quantification⁹⁵. Ils peuvent être également associés à d'autres produits dans d'autres nomenclatures pour permettre la conversion de l'une à l'autre, avec éventuellement des coefficients de conversion (Chapitre 3/I.3).

Ainsi pour résumer les produits sont des concepts à minima nommés et faisant fréquemment référence à une nomenclature qui hiérarchise les produits et leur associe éventuellement un code. Les produits peuvent être associés à des caractéristiques complémentaires, notamment des coefficients qui permettront le passage entre différentes d'unités, et des références vers d'autres produits pour la conversion dans d'autres nomenclatures.

d) Stock

Un stock est défini comme une quantité (dans une ou plusieurs unités) d'un produit appartenant à un acteur et/ou localisé sur un territoire à un instant donné. Dans des cas particuliers, nous pouvons définir le stock sur une période, plutôt qu'à un instant donné, par exemple pour caractériser le temps de transit d'un objet dans une plateforme logistique.

Il est possible d'ajouter d'autres attributs à un stock (une référence, sa géolocalisation...). Toutefois, les informations sur le produit, la temporalité, la(les) quantité(s) et unité(s), et l'acteur ou le territoire associé sont généralement suffisantes pour décrire de manière satisfaisante un stock.

e) Flux

La notion de flux possède plusieurs définitions selon les disciplines. Trois idées clefs peuvent y être associées :

- Un échange entre deux acteurs (matière, argent...), éventuellement entre l'environnement et le système socio-économique,
- Un déplacement avec un sens, depuis une origine vers une destination, ainsi qu'éventuellement le chemin correspondant.
- La traversée d'une surface ou d'une frontière (une « porte ») par de la matière ou de l'énergie, selon une définition plus physique.

⁹⁵ Par exemple 1 ordinateur = 8 kg = 1000 € = 300 kg eq. CO2 // 1 kg de fer = 7860 cm³...

Ces idées se recoupent, mais ne sont pas équivalentes, et ne peuvent pas être décrites exactement de la même façon. Dans tous les cas, définir un flux requiert d'indiquer le produit dont il est constitué, sa quantité (dans une ou plusieurs unités) et le cadrage temporel du flux (une période ou une date ponctuelle). Il est également nécessaire de contextualiser le flux en l'associant au territoire sur lequel il se manifeste, et/ou à l'acteur qui le manipule. En fait, ce dernier point amène à devoir distinguer deux types de flux :

- Ceux dont on connaît l'origine (en termes d'acteur et/ou de territoire) **et** la destination, et éventuellement le chemin. Nous les désignerons comme des « flux de chemin » (*PathFlow*). Les statistiques de transport de marchandises entrent dans ce cas de figure.
- Ceux qui se manifestent à un endroit donné, sans que cela implique de connaître à la fois l'origine et la destination de manière claire. Nous les qualifierons de « flux de porte » (*GateFlow*). Plusieurs notions courantes du vocabulaire économique et environnemental reposent implicitement sur cette notion : lorsqu'on parle d'extraction, de recyclage, de rejets, d'émissions, de production ou de consommation, c'est souvent en association avec un seul acteur ou un seul territoire.

En plus de ces informations minimales, d'autres attributs peuvent compléter la description d'un flux notamment le moyen de transport (route, rail, eau, air, pipeline, fil électrique...).

f) Transformation

La transformation de la matière et de l'énergie est le fondement des processus économiques et techniques. Dans une approche modélisatrice et simplificatrice, une transformation peut être assimilée à un système « boîte noire », qui convertit des produits en entrées en d'autres produits en sortie, sur une certaine période (Host 1989). Cette conversion est soumise à des lois de conservation à minima physiques (celles de la masse et de l'énergie), et parfois socio-économiques (comme la conservation de la quantité de monnaie⁹⁶). Les transformations peuvent être liées entre elles, les sorties des unes devenant les entrées des autres, permettant de conceptualiser des interactions complexes avec rétroactions pouvant

⁹⁶ La conservation de quantité la monnaie ne s'applique qu'à une échelle micro-économique. À l'échelle macro, les privilèges de création monétaire octroyés à certains acteurs économiques rendent mouvant ce principe de conservation.

former des macro-systèmes. Les territoires et les acteurs peuvent ainsi être assimilés à de tels macro-systèmes de transformation.

Les transformations constituent le support d'une modélisation dynamique des flux et stocks, que ce soit pour des processus élémentaires ou pour des entités macro (territoires ou acteurs), là où les concepts précédents ne permettent qu'une description comptable statique. Les transformations permettent de connecter explicitement ces flux et ces stocks de produits entre eux.

Ainsi, de manière formalisée, une transformation va être décrite à minima par une référence (nom, code, territoire ou acteur), une série de flux (précisant ainsi les produits et les quantités) en entrée, une série de flux en sortie et une période. Il peut être nécessaire de considérer également des variations du stock interne, surtout à des échelles macro. Dans des approches plus complexes, une transformation peut être plus précisément décrite à travers un tableau (ou matrice) entrées/sorties, ce qui permet de lier les flux en entrées avec ceux en sortie. Au regard du fonctionnement métabolique d'un territoire, les entrées et sorties peuvent être classées en trois catégories (Giljum & Hubacek 2009) : les produits fonctionnels échangés (la matière première venue d'ailleurs en entrée, les produits finis vendus/exportés en sortie), les déchets récupérés ou fournis par d'autres acteurs, et les flux depuis ou vers l'environnement (la matière première extraite en entrée, les rejets et émissions en sortie).

En poursuivant la réflexion, il nous semble pertinent en plus d'un tableau entrées/sorties, d'associer aux transformations un tableau sorties/sorties. Dans la logique productive, les déchets et rejets vers l'environnement sont générés pour servir la production de produits fonctionnels. À travers ce type de tableau, il est possible d'associer ces produits fonctionnels (en sortie) aux déchets et rejets générés pour leur production (en sortie également). La formalisation de cette approche demande toutefois un développement spécifique que nous n'avons pas l'ambition de mener ici.

g) Généralisation

À travers l'étude de chacun des six concepts, nous avons pu identifier certains points communs. D'une part, plusieurs propriétés se retrouvent dans différents concepts : les attributs d'identification par des noms ou par des codes, les attributs temporels, et la possibilité de hiérarchiser les éléments entre eux.

Ensuite, pour chacun des concepts, nous avons pu identifier des propriétés que nous appellerons fonctionnelles, c'est-à-dire qui doivent être renseignées ou qui sont porteuses d'un sens plus complexe comme les relations hiérarchiques. Nous avons également identifié des propriétés non fonctionnelles qui peuvent être considérées comme de simples caractéristiques élémentaires : la population ou le PIB d'un territoire, le nombre d'employés d'une entreprise ou d'un acteur... En fait, il ne semble pas possible de déterminer a priori toutes les propriétés qui peuvent être mobilisées dans une étude. Nous assumons ainsi que cette liste de propriétés identifiées n'est pas exhaustive. La littérature pose le débat entre une description précise des flux, mais qui demande potentiellement du temps pour la formaliser, et une description minimale, plus simple à obtenir, mais plus limitée dans l'exploitation qui peut en être faite (Adoue 2004 ; Georgeault 2015). Notre ontologie offre ainsi la souplesse de pouvoir s'adapter aux deux approches : des données minimales sont requises pour donner du sens aux concepts, mais ceux-ci peuvent être librement précisés et enrichis de propriétés non fonctionnelles supplémentaires.

Nous notons que de nombreuses propriétés doivent pouvoir être repérées dans le temps. Lorsqu'on parle de la population d'un territoire ou du nombre d'employés pour un acteur, cela n'a réellement de sens que si l'on sait à quel moment cela correspond. À défaut d'être explicite, l'interprétation est faite qu'il s'agit du moment présent.

Enfin, les concepts de Flux et de Stock sont quantifiés : ils ne prennent de sens que si leur quantité est précisée, qu'elle soit économique (en euros) ou physique (en kg).

Afin d'exploiter ces similarités entre les concepts sous-jacents du métabolisme, nous en introduisons deux nouveaux, plus abstraits, mais qui vont permettre de généraliser certains propos : Les « objets du métabolisme d'un territoire » (*MTOjects*) qui regroupent l'ensemble des six concepts du métabolisme territorial (Territoire, Acteur, Produit, Flux, Stock, Transformation), et les « objets quantifiés » (*MTQuantified*) qui désignent les flux et les stocks. Cette généralisation nous sera particulièrement utile par la suite pour construire le modèle de données et développer des fonctionnalités algorithmiques.

Ainsi, à travers cette analyse, nous avons ainsi défini les bases d'une ontologie du métabolisme territorial. Nous formulons ainsi une hypothèse structurante : les données décrivant le métabolisme d'un territoire décrivent en fait des instances des six concepts sous-jacents ou des relations entre eux. L'ontologie constitue une première étape dans la

normalisation de l'information, nous allons par la suite la dériver en un modèle de données qui pourra être implémenté au sein d'un SI. Il n'est pas prévu dans nos travaux de définir cette ontologie à travers un langage de représentation formel (OWL par exemple) ; nous préférons d'abord évaluer sa pertinence à travers la structuration d'une base de données classique, avant de consacrer notre attention sur un formalisme plus avancé et plus complexe.

2) Analyse fonctionnelle

L'ontologie présentée précédemment nous permet de décrire la structure de l'information à manipuler. Toutefois, elle ne nous renseigne pas sur la façon dont cette information doit être manipulée. Aussi, sur la base des différents types de traitements récurrents identifiés (Chapitre 3), nous réalisons une analyse fonctionnelle pour aller vers une plus grande explicitation des concepts et fonctionnalités attendues du SINAMET. Il s'agit ainsi de dresser les grandes lignes d'un cahier des charges et de pouvoir y faire référence par la suite sans toutefois pouvoir les détailler pleinement pendant nos travaux.

a) Disposer d'une base pour centraliser les données

Une des principales fonctionnalités attendues d'un SINAMET est de pouvoir rassembler plusieurs jeux de données au même endroit, leur multiplicité (Chapitre 3/II.1) étant l'un des freins les plus importants à l'analyse des données. Ainsi :

- (F1) Le SINAMET inclut une base de données centralisée permettant de structurer l'information avec un modèle de données constant.
- (F2) L'importation des données depuis différents types de sources (tableurs, autres bases de données, entrées manuelles...) vers cette base centralisée est facilitée et inclut une vérification de l'intégrité et de la cohérence des données importées.

b) Gérer des instances et de leurs propriétés

À travers la définition de l'ontologie (section I.1), nous formulons l'hypothèse que les données relatives au métabolisme d'un territoire peuvent être assimilées à des propriétés relatives à six types d'instances différents : des territoires, des acteurs, des produits, des flux, des stocks et des transformations. En fait, ce sont plutôt sept types, en distinguant les flux de chemin et les flux de porte. Ainsi :

- (F3) Le SINAMET offre les fonctionnalités pour créer les sept types d'instances du métabolisme et d'y associer une liste non limitée de propriétés.

- (F4) Chaque propriété est constituée d'un nom et d'une valeur et peut être repérée dans le temps (par une date ou une période) et sourcée.
- (F5) La valeur d'une propriété peut être un attribut (texte, nombre, coordonnées géographiques) ou être une relation avec une autre instance de la BDD.
- (F6) Certaines propriétés renvoient à un sens plus complexe qui est implémenté à travers le SINAMET : unicité, hiérarchisation, voisinage, alias, coefficient de conversion... Ces propriétés sont identifiées Annexe 3.

c) Sélection des données

Les fonctionnalités présentées précédemment concernent la façon de structurer les données à travers des instances et des propriétés dans une base centralisée. Pour être efficace, le SINAMET doit également inclure des fonctionnalités pour facilement sélectionner les données dans cette base. Ainsi :

- (F7) Les instances non quantifiées (territoire, acteur, produit) peuvent être sélectionnées par leur nom ou leur code, ceux-ci étant inscrits dans des référentiels multiples (langue et nomenclature).
- (F8) Les instances peuvent être sélectionnées en lot dans la base de données selon leurs attributs et la logique de l'ontologie (type, temporalité, territoire ou produit associés...).
- (F9) Il est possible de naviguer et d'accéder aux données via les relations que les instances entretiennent, par exemple l'organisation hiérarchique.

d) Analyse des données

À partir du jeu de données préparé qui peut être extrait de la BDD du SINAMET, il s'agit désormais de mener les opérations permettant d'aboutir aux données finales qui pourront être interprétées, sans qu'il ne nous soit toutefois possible de définir a priori toutes les typologies d'analyse disponibles. Ainsi :

- (F10) Les données peuvent être manipulées à travers différentes structures selon le type d'analyse à mener (objets, tableaux ou structures récursives – Chapitre 1/III.2).
- (F11) Des fonctionnalités permettent de réaliser les opérations de conversion des unités ou des nomenclatures.

- (F12) Des fonctionnalités permettent de synthétiser les données comme transformer une liste d'objets quantifiés en une valeur unique ou un profil temporel⁹⁷.
- (F13) L'exportation des résultats de l'analyse sous différentes formes (tableur, shapefile...) et leur représentation sont facilitées.
- (F14) Il est facilement possible de construire de nouvelles analyses personnalisées pour augmenter les possibilités de compréhension des données.

e) Utilisateurs et interface graphique

Les fonctionnalités attendues et présentées précédemment sont explicitées au regard des besoins de cohérence de l'information dans sa manipulation, indépendamment de l'utilisateur. Toutefois, ces fonctionnalités nécessitent des compétences particulières en termes de traitement de données pour être appréhendées et utilisées. Si on peut faire l'hypothèse que manipuler un tableur simple est à la portée d'un public large, les personnes en mesure de manipuler l'information à travers une structure plus complexe sont bien moins nombreuses. Généralement, il s'agit des gens avec des notions de programmation. Or, il nous semble nécessaire de pouvoir rendre intelligibles les résultats des analyses de flux auprès d'un panel plus large d'acteurs (Chapitre 2/V.3) et notamment par la production interactive de graphique (Chapitre 3/III.2). Ainsi :

- (F15) Une interface graphique ergonomique est disponible dans le SINAMET. Elle permet d'avoir une vue globale des données stockées, et d'accéder à chaque instance et à ses propriétés à travers des fonctions de recherche.
- (F16) Les algorithmes de traitement de données développées par les personnes qui le peuvent (les programmeurs) doivent pouvoir être réutilisés et paramétrés facilement, y compris par des non-programmeurs à travers une interface graphique.

En conclusion de cette analyse fonctionnelle, nous avons identifié différentes fonctionnalités qui nous paraissent être fréquemment mobilisées ou utiles à mobiliser pour étudier le métabolisme des territoires. Réunir ces fonctionnalités au sein d'un système d'information unique nous semble ainsi nécessaire pour réussir l'automatisation du traitement des données.

⁹⁷ Il s'agit d'une liste de valeurs associées à des dates

II. Conception du SINAMET

L'ontologie et l'analyse fonctionnelle permettent de donner un cadre à l'information : vocabulaire, relations, contraintes d'intégrité, attentes des utilisateurs... Le SINAMET que nous voulons concevoir doit ainsi s'inscrire dans ce cadre.

Dans les parties suivantes, nous formulons plusieurs propositions pour implémenter concrètement le SINAMET : implémentation d'une base de données centralisée, utilisation des techniques de programmation orientée objet, mise en correspondance des modèles de données pour faciliter leur importation et leur structuration, et déploiement d'une architecture modulaire.

Ces propositions sont des choix techniques, réalisés en partie de manière itérative (Chapitre 1/III.1) et consolidés par le développement d'un prototype logiciel et son utilisation sur des données réelles (voir Partie III).

1) Définition du modèle de données d'une base centralisée

Pour constituer la base centralisée de données (F1), nous devons définir un modèle de données adapté, c'est-à-dire une structure de données combinée à une ontologie. Dans notre état de l'art (Chapitre 1/III.4), nous faisons émerger trois types d'approches qui correspondent à cette intention d'une BDD centralisée :

- Les Entrepôts De Données (EDD), qui utilisent classiquement un modèle entité-association et une ontologie définie à partir des données à centraliser,
- Wikidata, qui s'appuie également sur un modèle entité-association mais avec une ontologie générique définie a priori,
- Les techniques du web de données, qui offrent beaucoup de souplesse pour définir, partager et faire évoluer des ontologies, mais au prix d'une structure de données plus complexe à appréhender.

Nos travaux étant encore exploratoires, nous préférons nous appuyer sur le modèle entité-association (Chapitre 1/III.2) : tout en étant assez simple et commun, il est suffisamment riche pour structurer une information d'une grande complexité. D'autres travaux reposent sur des structures de données différentes, comme un hypercube (Pauliuk et al. 2019) ou le web de données (Davis 2012), mais la première nous semble conceptuellement trop limitée, et la seconde encore trop complexe.

Une limite des entrepôts de données est la nécessité de définir une ontologie commune à partir des données déjà existantes et ayant leur propre modèle. Si nous avons montré à travers notre ontologie qu'il existe un nombre limité de concepts (six) pour décrire le métabolisme d'un territoire, nous n'avons pas pu établir de liste finie de propriétés pour caractériser ces concepts. Finalement, il nous semble nécessaire d'hybrider les EDD avec l'approche Wikidata. Nous définissons a priori les classes du modèle et les propriétés fonctionnelles identifiées (F3), tout en donnant la possibilité d'ajouter à ce modèle d'autres propriétés non fonctionnelles, c'est-à-dire dont il n'est pas possible d'en implémenter un sens complexe (F6). En plus de permettre une souplesse dans la liste des propriétés⁹⁸, l'approche Wikidata permet également de qualifier leurs valeurs. Pour nous, il s'agit essentiellement de renseigner la temporalité et la source de cette valeur (F4). Contrairement à Wikidata, il ne nous semble pas nécessaire d'avoir une liste plus longue de qualificatifs.

Nous proposons ainsi le modèle logique de données Figure 11, qui complète l'Annexe 3. Les six concepts fondamentaux sont dérivés en sept classes d'objets : la description d'un flux de chemin ou d'un flux de porte n'étant pas équivalente, cela nous oblige à distinguer les deux classes d'objets. Les classes sont associées entre elles selon la logique de l'ontologie. Deux classes mères (MTOBJECT et MTQUANTIFIED) sont également définies dans le modèle et permettent aux classes d'objets concernées d'hériter des propriétés communes, et donc de simplifier le développement et la maintenance du système d'information. Nous rendons disponibles les trois types de propriétés identifiées (F5) : attribut primitif (texte, nombre), géographique, ou relation avec un autre objet du métabolisme.

⁹⁸ La terminologie de notre modèle et celle de Wikidata diffèrent légèrement : Pour nous (pour Wikidata), le modèle consiste en des instances (*item*) définies par des propriétés (*statement*) ayant un nom (*property*), une valeur (*value*) et des qualificatifs de la valeur (*qualifier*). Notre choix des termes repose sur la volonté d'utiliser l'approche orientée objet présentée plus loin.

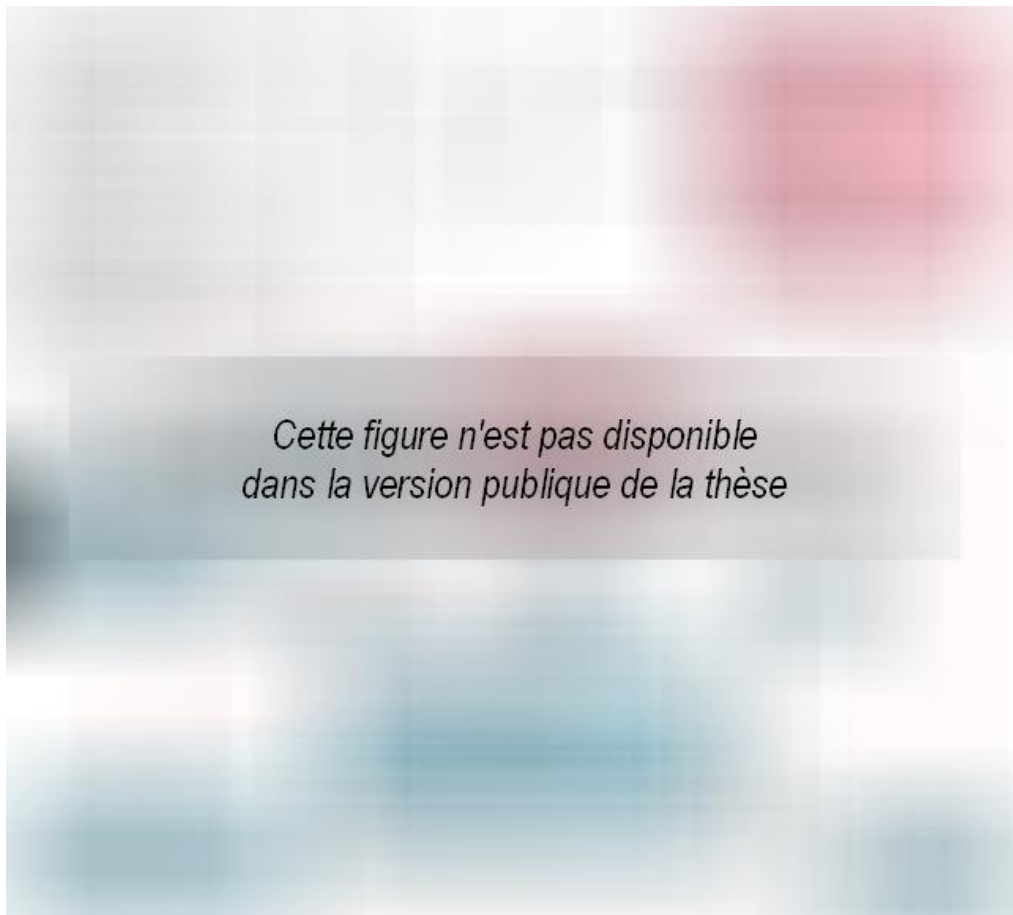


Figure 11 : Modèle logique de données implémenté dans le SINAMET

2) Usage des techniques de programmation orientée objet et abstractions complémentaires

Le paradigme orienté objet offre des avantages dans la manipulation de données et permet de compléter le modèle entité-association par des fonctionnalités de traitement à travers des méthodes (Chapitre 1/III.3). Ainsi, conformément à notre ontologie, chaque instance du métabolisme (chaque flux, chaque produit, chaque territoire...) est représentée par un objet informatique pour lequel il est simple d'accéder aux propriétés, qui peuvent être elles-mêmes d'autres objets.

Pour pouvoir synchroniser les objets avec la base de données et en abstraire la gestion, nous recourons à une solution ORM. L'usage d'un ORM ne permet toutefois pas d'abstraire complètement les liens avec une BDD. La synchronisation des objets et les requêtes à exécuter doivent encore être explicitées. Cependant, notre modèle de données, qui conceptualise les propriétés comme des objets à part entière, est de fait plus complexe à interroger via des requêtes.

Pour pallier ces difficultés, nous implémentons plusieurs fonctionnalités complémentaires au sein du SINAMET. Le but est d'abstraire une partie de cette complexité pour sélectionner et manipuler plus simplement les objets de données et leurs relations (F6) ; (F7) ; (F8) ; (F9). Nous implémentons ainsi une interface avec la base de données, des méthodes spécifiques pour les classes d'objets et des fonctions de conversion et de synthèse. Nous illustrons ces points par quelques exemples.

L'interface avec la base de données (*Database Interface* - DBI) vise à permettre d'ajouter ou extraire des données de la base sans avoir à construire explicitement la requête. Nous présentons la méthode d'importation de données un peu plus loin. En ce qui concerne leur extraction, il est fréquent de vouloir récupérer des instances individuellement par des noms ou des codes pour les territoires, les acteurs ou les produits (F7), ou sous forme de liste selon les territoires associés, la temporalité ou les produits concernés pour les flux ou les stocks (F8). Nous implémentons des fonctions comme par exemple *DBI.get_territory(name)*⁹⁹ ou *DBI.get_flows(territory)* pour récupérer respectivement un territoire selon son nom ou des flux par leur territoire. Au total, ce sont une vingtaine de fonctions disponibles pour récupérer les différents types d'objets, individuellement ou collectivement, selon leurs propriétés.

Les méthodes complémentaires associées aux objets vont permettre quant à elles de naviguer plus facilement dans les propriétés et les relations entre objets (F9). Par exemple la fonction *MTObject.get_property(name, date)* va permettre d'obtenir la valeur de la propriété d'un objet à une date donnée, *MTObject.is_in(mtoject)* va vérifier si un objet (un territoire par exemple) est inclus dans un autre, la fonction *Territory.get_children(scale)* va renvoyer tous les territoires d'une certaine échelle contenus dans un autre, comme toutes les villes d'une région. Certaines méthodes permettent aussi de changer la structure des données, comme *Product.get_tree()* qui revoit un produit et ses sous-produits structurés sous la forme d'un arbre, ce qui est pratique lorsqu'on veut manipuler la hiérarchisation entre les objets (F10).

Des fonctionnalités de conversion de nomenclature ou d'unités (F11) sont également embarquées dans le SINAMET. Sous réserve de l'utilisation des bonnes propriétés, ces opérations de conversion sont réalisées automatiquement à l'aide d'un appel à une seule fonction : *Product.to_nomenclature(nomenclature)* et *Flow.get_quantity(unity)*. En ce qui concerne les fonctionnalités de synthèse (F12), nous implémentons une classe Profil qui permet

⁹⁹ La fonction *DBI.get_territories(*)* existe également pour sélectionner plusieurs territoires.

de réduire une liste de flux en sa valeur sur une période donnée – *Profil.get_value(list_flows, unity, year)* – ou d’en récupérer une représentation temporelle (liste de valeurs associées à une liste de périodes) que nous appelons un *profil* et qui peut être combiné avec d’autres (additionné, multiplié, divisé...).

Nous avons présenté quelques fonctions utiles pour sélectionner et manipuler les objets de données selon leurs propriétés. Ainsi, pour réaliser une nouvelle étude, il n’est pas utile de développer à nouveau tous les algorithmes idoines, ni à construire les requêtes. Cela résulte en un gain de temps et une plus grande facilité conceptuelle pour construire les analyses de données.

3) Tableau d’association pour l’importation des données

En plus de l’inspiration pour la conception du modèle de données du SINAMET, la conception d’entrepôts de données nous permet d’interroger deux autres aspects particuliers de notre démarche : la granularité de l’information et la conception des extracto-chargeurs (Chapitre 1/III.4).

Le fonctionnement d’un EDD amène à devoir définir une granularité de l’information, par exemple spatiale ou temporelle. Or, nous montrons qu’il n’existe pas de granularité générale idéale pour les études de métabolisme (Chapitre 3/I). Notre approche est donc d’éviter d’imposer une granularité particulière pour le stockage des données dans la base afin de conserver le maximum d’information contenue dans les données sources. Grâce à notre modèle de données, nous pouvons ainsi décomposer et hiérarchiser les territoires, les acteurs et les produits de manière aussi fine que souhaité, et la temporalité des flux et des stocks peut être définie au jour près¹⁰⁰. Ce sont les fonctionnalités complémentaires incluses dans le SINAMET qui vont permettre de synthétiser les données si besoin, mais à la volée après stockage.

Concernant les extracto-chargeurs (ETL), leur objectif s’inscrit dans les mêmes termes que notre démarche : Il s’agit de pouvoir importer des données de différentes sources externes ayant leur propre structuration et ontologie, au sein d’une base centralisée ayant un modèle de données fixé. L’étude de ces composants nous permet de formuler quatre idées clefs :

¹⁰⁰ Ce qui est suffisant dans nos travaux, mais qui devra être affiné pour étudier des dynamiques journalières.

- La forme la plus simple d'ETL consiste en une mise en association des intitulés des données sources avec ceux du modèle de la base centralisée¹⁰¹.
- Les données sources qu'il faut intégrer ne sont pas toujours adéquatement formatées ou structurées. Il peut être nécessaire d'y inclure des algorithmes de préparation des données¹⁰².
- Les ETL sont spécifiques à chaque source de données. L'effort d'implémentation de ces ETL est largement dépendant de la complexité de la phase de préparation.
- Dans la pratique (de notre état de l'art et de notre expérience), la conceptualisation d'ETL est une tâche un peu technique, essentiellement confiée à des programmeurs. Si cette affirmation peut être nuancée pour les formes les plus simples d'ETL, nous ne chercherons toutefois pas les rendre accessibles (dans leur conception) au grand public.

Tout comme les ETL des EDD, ceux du SINAMET implémentent différentes étapes : l'ouverture des données sources, leur préparation, l'association avec le modèle de données commun, et leur chargement dans la base centralisée.

L'ouverture des données ne pose généralement pas de difficultés, la plupart du temps il s'agit simplement d'ouvrir un tableur. La préparation peut poser beaucoup plus de problèmes, nous l'avons déjà mentionné (Chapitre 1/II.3). Les phases d'association et de chargement demandent également un peu d'attention. Le lecteur pour retrouver l'implémentation d'un module ETL comme exemple commenté en Annexe 4.

Notamment, nous avons opté pour conceptualiser le SINAMET sur la base d'une approche orientée objet. Aussi, l'association avec le modèle de la base consiste en la création d'instances de différents types (Territoire, Produit, Flux...) et en l'ajout de leurs propriétés. Pour associer les propriétés des instances avec les valeurs des données sources, nous utilisons un tableau d'association { mots-clefs → valeurs }. Cela constitue une approche simple et efficace

¹⁰¹ Par exemple à partir d'une liste d'enregistrements de territoires qui indiquent en colonne n°1 le nom, en n°2 le code, en n°3 la population, on peut définir le tableau d'association :

Territoire["nom"] ← colonne 1 // Territoire["code"] ← colonne 2 // Territoire["population"] ← colonne 3

Il suffit alors à l'ETL de parcourir les données sources puis de créer pour chaque ligne une instance de territoire avec les propriétés correspondantes aux colonnes.

¹⁰² Par exemple, les données de la bases Agreste (production agricole) présentent les territoires sous la forme « code – nom », or il est plus adéquat d'avoir les champs « code » et « nom » séparés. Une étape de préparation est ainsi ajoutée pour dissocier les deux.

(F2). Certains mots-clefs font référence à des propriétés fonctionnelles (voir Annexe 3), mais il est possible d’user de n’importe quel autre mot-clef pour créer une nouvelle propriété non fonctionnelle.

Les propriétés sont fréquemment sourcées et datées. Nous donnons donc la possibilité de renseigner une source, en lien avec ses métadonnées (Pauliuk et al. 2019), et une temporalité. Cette temporalité peut être définie de quatre manières différentes : par une date ponctuelle (*date_point*) ou par une période via une date de départ et de fin (*date_start* + *date_end*), via une année (*year*) ou via une année et un mois (*year* + *month*). Dans les deux derniers cas, l’information est automatiquement interprétée par le SINAMET pour être ramenée à la forme standard d’une période dans notre modèle de données (*date_start* + *date_end*).

Certaines propriétés du modèle sont des relations entre les instances. Or, les données sources ne contiennent pas des instances en tant que telles, mais des références aux instances par leur code ou par leur nom notamment. Certains mots-clefs du tableau d’association sont donc adaptés pour permettre d’exprimer au SINAMET d’interpréter cette subtilité, par exemple « TerritoryCode » ou « TerritoryName ». Cela permet d’associer automatiquement un territoire de la base de données à partir de son nom ou de son code. Autre subtilité, les codes peuvent être selon différents référentiels (Insee, Nuts...), les noms en différentes langues et les quantités en différentes unités. Nous introduisons donc une notation à travers le symbole arbitrairement réservé « @ ». Ainsi, il devient possible de préciser certains mots-clefs, par exemple « Code@insee », « Code@nuts3 », « Name@fr » ou « Quantity@kg »¹⁰³.

Concernant le chargement des données, la principale difficulté réside dans la jointure des données. Par exemple, nous pourrions disposer de trois tableaux, l’un présentant le nom et code Insee de territoires, le deuxième le nom et le code NUTS-3, et le dernier le code Insee et la population. Évidemment, ces données sont à lier, il ne s’agit pas de créer des nouvelles instances de territoires pour chacun des jeux de données. Le SINAMET dispose ainsi des fonctionnalités qui vont permettre de comparer les données existantes avec les nouvelles à charger et d’en déduire s’il s’agit de créer de nouvelles instances ou simplement de les mettre à jour. L’usage d’une clef primaire simplifie le processus : si un objet avec cette clef primaire existe dans la

¹⁰³ D’un point de vue de la structure de données, cela revient finalement à définir un tableau d’association pour certaines valeurs d’un tableau d’association.

base de données, alors il faut le mettre à jour. À défaut, le SINAMET créera une nouvelle instance.

4) Architecture modulaire

La conception d'ETL souligne la particularité de chaque source qui nécessite son propre composant logiciel pour être intégrée à d'autres. De l'autre côté, en sortie de la base de données, chaque typologie d'analyse des données est également spécifique. Ce constat nous amène à vouloir implémenter une architecture logicielle modulaire : puisqu'il n'est pas possible de prévoir a priori toutes les sources de données, ni toutes les typologies d'analyses qui pourraient être mobilisées, il faut donner la possibilité d'ajouter de nouvelles sources ou analyses a posteriori.

Cette conception modulaire est renforcée par la séquentialité du traitement des données à travers le SINAMET : d'abord l'importation des données à travers les ETL, puis leur extraction pour les analyser et les exporter. Ce processus est illustré Figure 12 de manière générale et sur un exemple précis. Concrètement, il s'agit d'abord d'importer les données en deux temps : d'abord les données de contexte (territoires, acteurs, produits) puis les données quantitatives sur les flux ou les stocks. Une fois importées et structurées dans la base, ces données peuvent être facilement extraites de manière paramétrable pour être analysées, visualisées et exportées. Les scripts des modules ETL pour les départements et le transport de marchandises, ainsi que celui du module « Graphe Régional », qui appuient l'exemple Figure 12 sont présentés et commentés en Annexe 4 et en Annexe 5. Nous pouvons souligner la concision des scripts permise par le SINAMET qui offre des fonctions de haut niveau aux développeurs des modules pour abstraire une grande partie du traitement des données.

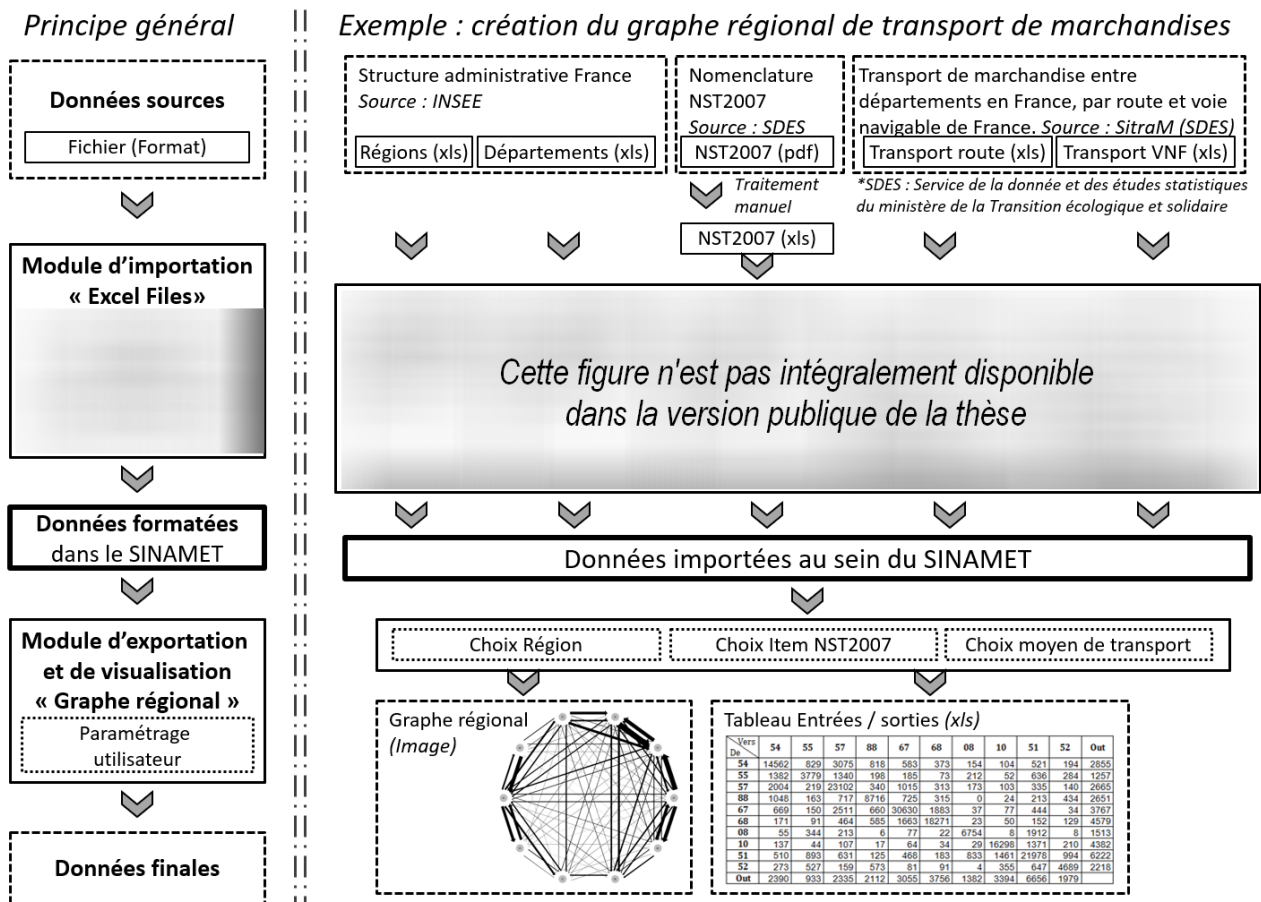


Figure 12 : Les étapes du traitement de données à travers le SINAMET, cas général et exemple

En posant la base de données centralisée comme point de passage systématique des données, nous dissocions les phases d'importation-préparation de celles d'extraction-analyse-exportation. Éventuellement, une phase de traitement complémentaire peut être mobilisée lorsqu'il s'agit d'inférer les données de la base centralisée pour en produire d'autres à partir d'un modèle. Les différentes phases sont implémentées distinctement de façon modulaire. Cette architecture logicielle basée sur un noyau logiciel générique et des modules spécifiques est illustrée Figure 13. Elle permet de construire différentes analyses à partir d'une même source, ou d'utiliser les algorithmes d'une même méthode d'analyse de données pour différentes sources pour d'autres contextes. En d'autres termes, le SINAMET permet de capitaliser le travail réalisé : chaque module peut être librement développé et partagé avec d'autres utilisateurs, leur évitant de devoir développer à nouveau les opérations du traitement de données (F16).

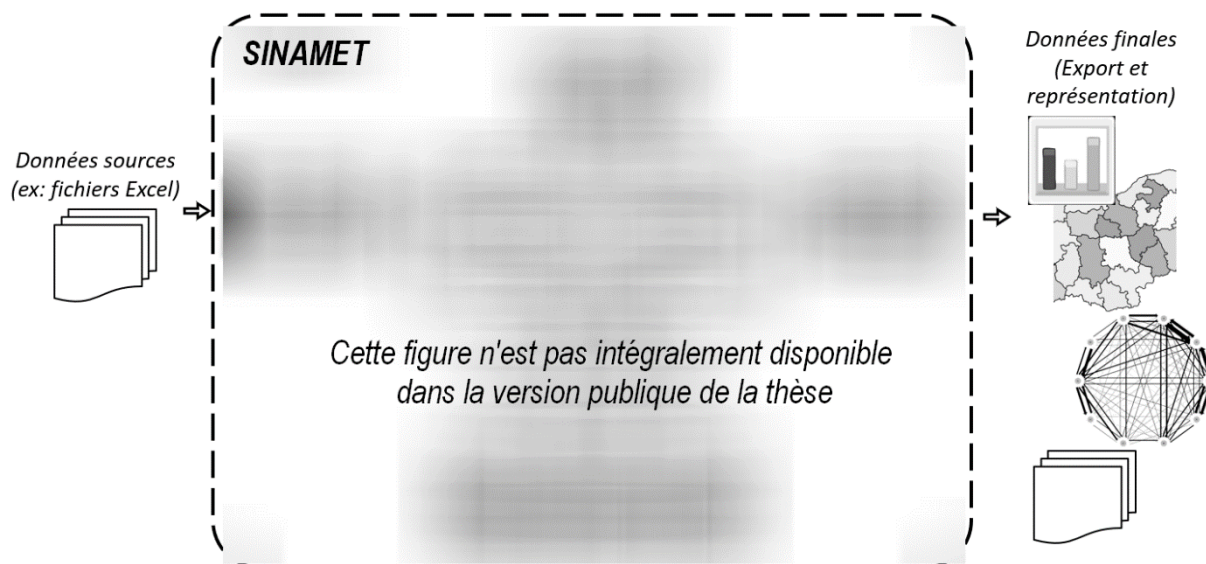


Figure 13 : Architecture logicielle du SINAMET

En plus de la capitalisation des travaux, les modules permettent d'abstraire complètement ou presque le traitement de données. Ils peuvent être mobilisés sans qu'il soit nécessaire d'en comprendre les rouages techniques. Éventuellement, il peut y avoir à préciser quelques paramètres pour utiliser les modules et renseigner sur le contexte (choix d'un territoire, des produits analysés, de la période...), mais cela peut facilement se faire à travers une interface graphique. Le SINAMET donne donc la possibilité, à des utilisateurs qui seraient peu familiers du traitement des données, d'utiliser quand même les travaux existants et de les adapter de manière interactive (à travers le paramétrage graphique) à leurs contextes et intérêts (F16). Cela va exactement dans le sens de notre appel à renforcer l'accessibilité de l'information de façon interactive (Chapitre 3/III.2).

5) Information géographique et SIG

L'association des flux à une représentation spatiale peut aider à l'analyse du métabolisme d'un territoire (Barles 2014 ; Ioppollo et al. 2014). Cependant, le SINAMET n'a pas vocation à reproduire les fonctionnalités d'un Système d'Information Géographique (SIG). En fait, les deux outils sont complémentaires, le premier étant plutôt destiné aux analyses quantitatives et qualitatives sur les flux, l'autre aux analyses spatiales. Afin de pouvoir exploiter cette complémentarité entre SIG et SINAMET, il nous paraît nécessaire de disposer des fonctionnalités minimales permettant d'importer de l'information géographique (F5), de la stocker et de l'exporter sous un format adéquat (notamment Shapefile) à travers le SINAMET

(F13). Ces fonctionnalités sont accessibles à travers différentes librairies (notamment GeoSQLAlchemy et GeoPandas, voir section suivante).

Même si cela n'a pas encore pu faire l'objet de nos travaux, il serait pertinent par la suite d'envisager un outil intégré, qui reprenne les fonctionnalités du SINAMET au sein d'un SIG¹⁰⁴. En effet, un certain nombre de fonctionnalités géographiques et de concepts implémentés dans les SIG peuvent aussi être mobilisés pour l'analyse du métabolisme : gestion de bases de données, de différentes structures ou de différentes couches, mesure de distances, visualisation cartographique, géocodage d'adresses, recherche d'imbrications spatiales...

III. Prototype du SINAMET

L'analyse fonctionnelle et les techniques mentionnées précédemment nous permettent de proposer un cadre conceptuel pour une manipulation facilitée des données liées aux flux physiques sur un territoire. Néanmoins, en l'état, ce cadre reste théorique. Afin de vérifier sa validité sur des cas d'études réels, nous avons développé un prototype logiciel du SINAMET pendant nos travaux. Ce prototype a été développé en partie avec le soutien de stagiaires, Olivier Saint-Eve et Maxime Eckstein, que nous remercions pour leur contribution.

Ce prototype reste incomplet dans ses fonctionnalités et dans son ergonomie. En effet, il n'était pas possible, dans le temps de nos travaux, de développer un logiciel pleinement opérationnel. Cela aurait demandé un travail de développement informatique conséquent, au détriment de la réflexion académique qui est le principal constituant du travail de thèse. Toutefois, sans prototype, il n'aurait pas été possible de valider notre approche et d'obtenir aussi facilement les résultats qui sont présentés dans la Partie III.

Aussi, dans cette partie, nous prenons le temps d'expliquer les constituants principaux de notre prototype logiciel et ses limites par rapport à nos ambitions. Cela permet également d'aider le lecteur dans sa compréhension du cadre théorique présenté précédemment.

1) Langage de programmation et SGBD

D'un point de vue théorique, la réalisation d'un système d'information peut se faire en utilisant différents langages de programmation : la logique des algorithmes n'en est pas

¹⁰⁴ Le contraire paraît plus complexe à mettre en œuvre étant donné le degré de maturité des SIG, au regard de notre SINAMET encore au stade de prototype.

dépendante et l'approche que nous présentons pourrait être implémentée à travers différents langages.

Toutefois en pratique, les langages se distinguent par différentes caractéristiques : caractère ouvert ou propriétaire, vitesse d'exécution, clarté et simplicité du code, niveau d'abstraction, disponibilité des bibliothèques, fonctionnalités, popularité... Afin de ne pas consacrer trop de temps dans nos travaux au développement du prototype, nous avons retenu un langage présentant un code simple avec un haut niveau d'abstraction et disposant d'un large panel de bibliothèques : Python.

Python est un langage de programmation sous licence libre et l'un des plus populaires du moment, si ce n'est le plus populaire (Cass 2018). Il est reconnu pour être un langage qui apporte une grande productivité pour le programmeur grâce à des fonctionnalités de haut niveau et un large panel de bibliothèques disponibles (Thomsen & Bach Pedersen 2009), tout en étant plutôt élégant dans son écriture. Une des faiblesses du langage Python est sa rapidité d'exécution : En tant que langage interprété, il mobilise un temps d'interprétation au fil de l'exécution qui le ralentit, cela pouvant devenir significatif sur des grandes masses de données. Il est toutefois possible de précompiler une partie du code pour en éviter l'interprétation systématique et donc accélérer l'exécution des algorithmes.

De la même façon qu'il faut choisir un langage de programmation, il convient également de choisir un Système de Gestion de Base de Données (SGBD). Parmi les plus communs, nous pouvons citer : SQLite, MySQL, PostgreSQL. Nous avons retenu ce dernier car il permet de stocker de l'information géographique et il est assez simple d'installation. Grâce à l'usage d'une solution ORM, la gestion de la base de données est complètement abstraite : il est finalement possible de changer assez facilement de SGBD si cela s'avère nécessaire (Chapitre 1/III.3).

2) L'usage de bibliothèques

Comme nous le mentionnions, la disponibilité de bibliothèques (ou bibliothèques) est un critère de choix des langages de programmation. Ces bibliothèques vont permettre de mobiliser des fonctionnalités de haut niveau, sans avoir à les implémenter. C'est évidemment un gain de temps énorme, indispensable pour envisager la réalisation d'un prototype dans les délais de nos travaux.

En choisissant Python comme langage de programmation, cela nous donne accès à un large panel de bibliothèques. Parmi elles, voici celles que nous mobilisons car elles nous permettent d'apporter des solutions à plusieurs attentes de l'analyse fonctionnelle :

- SQLAlchemy, & GeoAlchemy : Solution ORM permettant de gérer une base de données, & contenant des données géographiques, à travers une programmation orientée objet (F1) ; (F3).
- Matplotlib : Construction et affichage de graphiques (F13) ; (F14).
- Pandas, & GeoPandas : Permet une manipulation efficace des données, & des données géographiques, sous forme de *DataFrame*¹⁰⁵, ainsi que l'import et export de tableaux (F2) ; (F13) ; (F14)
- Anytree : Permet la structuration de données sous forme d'arbre (F10).
- PyQt : Interface utilisateur graphique (F15).
- Shapely : pour la gestion de données géographiques tels les points, lignes et polygones (F5)

3) Structure du code source

Fidèlement à l'architecture logicielle, le code source est organisé en deux parties : le noyau et les modules. Le noyau comprend la définition des objets, la gestion de la base de données à travers la solution ORM et la DBI, l'interface graphique, et différentes fonctions supports pour naviguer dans les données ou les synthétiser, et pour créer des modules.

Les modules sont quant à eux articulés en groupe autour d'un dossier contenant des données sources originales, et où chaque module contribue à importer ces données ou à les analyser, éventuellement en combinaison avec d'autres sources. Ils sont interfacés avec le noyau à travers une structure définie et constituée de trois opérations successives (voir Annexe 4/I) : paramétrage (*setting*), analyse de données (*execute*), export et information utilisateur (*output*). Ces opérations vont faire le liant entre les données, les fonctionnalités du SINAMET et les fonctionnalités d'autres bibliothèques pour répondre à l'objectif du module (importer, traiter ou analyser/exporter des données).

¹⁰⁵ Une *DataFrame* s'apparente à un tableau d'enregistrements enrichi en fonctionnalités.

4) État général de l'implémentation

Dans le temps imparti, nous n'avons pas pu développer toutes les fonctionnalités que nous jugeons nécessaires pour le SINAMET. Notamment, ne sont pas aboutis : la conversion entre nomenclatures, la gestion des intitulés de produits libres, l'implémentation des stocks et des transformations, la possibilité d'user de coefficients entre unités qui soient variables dans le temps, ou encore la notion de territoires frontaliers ou successifs. Fruit d'un travail itératif et exploratoire, le code source du prototype contient également des morceaux obsolètes ou qui ne sont pas pleinement opérationnels.

Il n'en reste pas moins que l'essentiel de notre conception est actuellement implémenté dans ce prototype : il est possible de traiter l'information à travers différents types d'instances aux propriétés datées et sourcées. L'importation des données à travers un tableau d'association est plutôt simple et efficace. Les principales fonctions de recherche et de sélection de données sont fonctionnelles et également simples à manipuler, tout comme la hiérarchisation des instances qui permet par exemple de mener facilement des analyses régionales ou nationales sur des données départementales (voir Chapitre 7).

Un autre élément indispensable pour le bon usage d'un logiciel est la documentation. Elle est encore incomplète, à l'image de notre prototype, et doit donc être travaillée pour rendre le SINAMET exploitable par d'autres utilisateurs.

5) L'implémentation des notions de transformations et de stock

Notre ontologie met en avant la notion de transformation comme fondation du métabolisme territorial. Toutefois, son implémentation à travers le SINAMET n'est pas encore claire. Nous présentons dans le Chapitre 8 un essai sur l'étude des flux alimentaires à l'échelle de l'Eurométropole en partant de la consommation humaine qui peut être assimilée à une transformation. En fait, ce sont les relations entre les transformations, leurs propriétés et les autres notions du métabolisme qui sont à clarifier. L'articulation avec les tableaux entrées/sorties nous semble également essentielle mais n'a pas pu être approfondie.

De la même manière, nous n'avons pas réalisé d'études impliquant la gestion de stocks. Nous pensons que l'approche sur des stocks est très similaire à la gestion des autres objets quantifiés que sont les flux, et en particulier les flux de porte, sans douter qu'il y ait certainement certaines subtilités que nous n'avons pas encore identifiées.

Malgré cela, les quatre autres notions (flux, territoire, acteur, produit) du métabolisme ont été suffisantes pour aborder certaines applications sous l'angle de la comptabilité matière (voir Partie III). Ainsi, si la clarification des notions de transformation et de stock peut élargir le potentiel applicatif du SINAMET, elles ne sont pas pour autant nécessairement mobilisées dans toutes les études.

6) Temps d'exécution

Une limite importante de notre prototype est sa rapidité d'exécution. En termes d'importation de données, nous arrivons à traiter la création d'environ 100 000 à 200 000 instances par heure, là où certains jeux comme la base SitraM contiennent plus de 500 000 instances. Par ailleurs, les requêtes ne sont pas pleinement optimisées et peuvent prendre plus d'une dizaine de secondes à être exécutées, requêtes qu'il peut parfois être nécessaire de répéter quelques fois à quelques centaines de fois (par exemple pour chaque département) selon les typologies d'analyses.

Cette lenteur nous a amenés à développer quelques stratégies de mise en cache qui permettent de limiter les requêtes sur la base de données, mais ces stratégies pourraient encore être optimisées. L'usage d'une solution ORM explique également une partie de ces performances limitées, mais s'en affranchir aurait demandé un travail complémentaire de développement logiciel trop important.

Toutefois, soulignons que si la durée d'exécution est importante, il s'agit d'une exécution automatisée qui ne requiert donc pas de supervision humaine une fois lancée. La gêne est donc bien moindre que si cela nécessitait une présence active, et est donc acceptable pour un prototype. Une fois que les données ont été importées (opération longue, mais à ne réaliser qu'une seule fois), la plupart des analyses que nous avons menées ne prennent que quelques dizaines de secondes ou quelques minutes pour être exécutées, rares sont celles qui demandent plus de temps.

7) Interface graphique utilisateur

Notre prototype intègre une interface graphique utilisateur (*Graphic User Interface* – GUI) qui facilite largement son utilisation (F15) même si son ergonomie et ses fonctionnalités peuvent encore largement être améliorées (Figure 14). Cette GUI inclut une barre de navigation pour accéder aux différents modules (F16), et un affichage multi-fenêtres qui permet à chaque

module de disposer de sa propre fenêtre. Plusieurs fonctions génériques sont accessibles depuis le menu, comme la recherche d'acteurs, de produits ou de territoires par leur code ou leur nom, l'affichage d'informations générales sur les données contenues dans la base, la création de modules, la suppression des données d'une source...

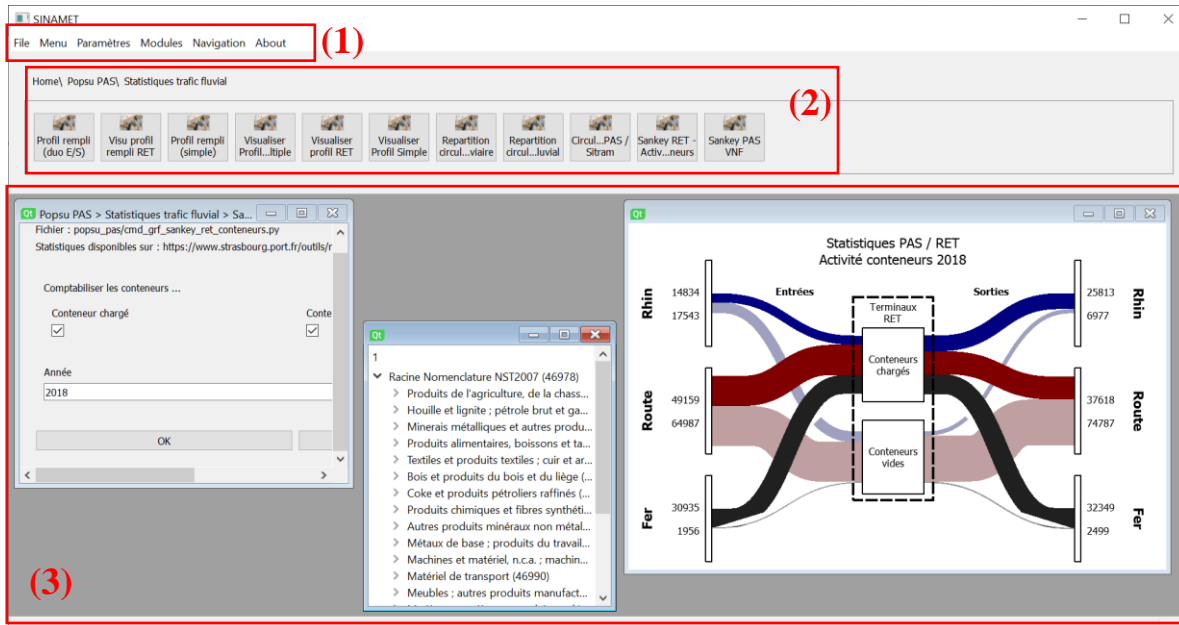


Figure 14 : Interface générale du prototype de SINAMET.

1 : Menu général, 2 : Navigation et accès aux modules, 3 : Affichage multi-fenêtres

La GUI permet notamment de visualiser les hiérarchies entre des instances, et d'accéder aux propriétés des instances de manière individuelle (Figure 15). Cela est utile pour s'assurer que les données contenues dans la base correspondent à ce qui est attendu.

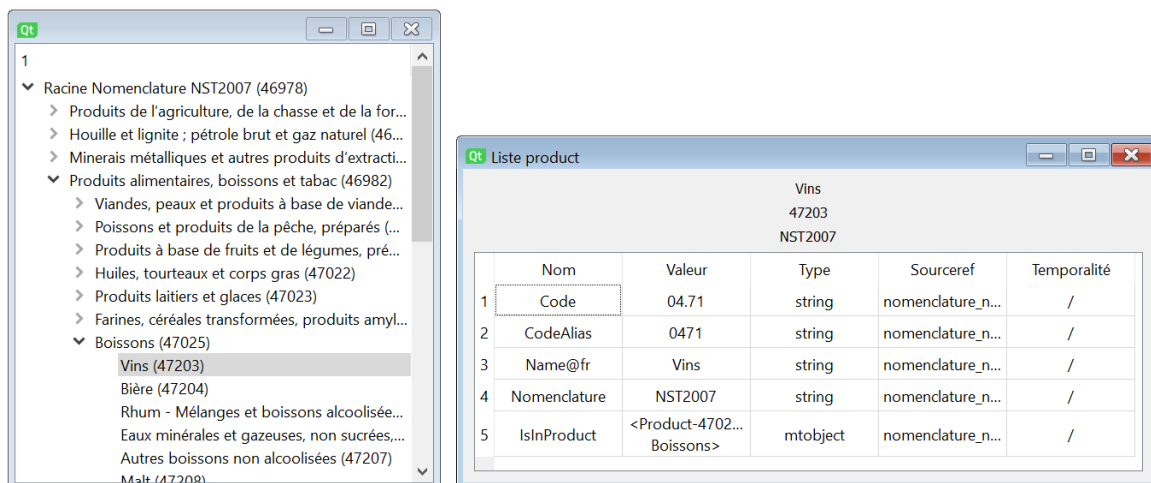


Figure 15 : Affichage de la hiérarchie des produits de la nomenclature NST2007 (à gauche) et détail d'un produit spécifique (à droite)

Le prototype intègre également les outils qui vont permettre de construire facilement des fenêtres graphiques pour paramétrer les analyses. L'appel du module « Graphe régional » (introduit Figure 12 p. 143) va générer une fenêtre de paramétrage, puis va créer le graphique correspondant (Figure 16).

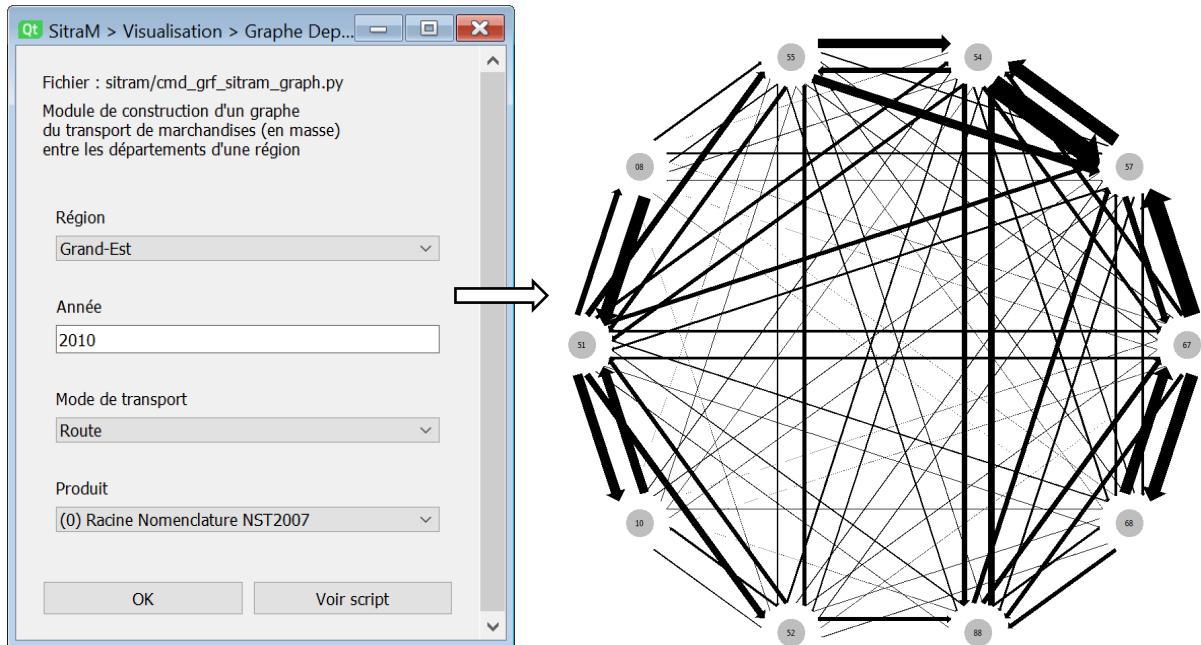


Figure 16 : Exemple du module de création d'un graphe régional du transport de marchandises. Le module est paramétré, puis la construction graphique est opérée automatiquement.

IV. Conclusion du quatrième chapitre

Dans ce chapitre, nous présentons notre démarche qui nous a permis, à partir du besoin d'automatiser le traitement des données dans les études de métabolisme, de concevoir un système d'information adéquat et d'en implémenter un prototype. Pour cela, nous avons d'abord analysé le sens de l'information en définissant une ontologie et puis détaillé les principales fonctionnalités attendues. Nous avons ensuite proposé différentes approches concrètes qui correspondent au cadre fixé. Enfin, un travail assez conséquent de développement a permis d'implémenter et de tester le prototype de SINAMET.

Même si le développement doit être poursuivi pour obtenir un outil réellement opérationnel, ce prototype nous a déjà permis de traiter des données relatives à différents cas d'études et a ainsi contribué à produire de l'information. Nous présentons ces cas d'études dans le chapitre suivant. Cela nous permettra de mieux conclure en fin de thèse sur les apports et les limites du SINAMET.

*Partie III Exemples
d'applications*

Le travail présenté précédemment nous permet de disposer de différents outils pour l'analyse et la compréhension du métabolisme d'un territoire :

- Une liste d'applications des analyses de flux de matières et d'énergie,
- Un ensemble de huit dimensions qui invitent à questionner plus systématiquement ce métabolisme,
- Les techniques de visualisation de données comme support privilégié pour rendre accessible une information complexe, malgré certaines limites dimensionnelles,
- Un ensemble de points d'attentions dans le processus de traitement des données, notamment en ce qui concerne leur utilisabilité,
- Un prototype d'un Système d'Information pour l'Analyse du Métabolisme des Territoires (SINAMET) qui permet de manipuler de grands jeux de données relatifs aux flux physiques.

Dans ce chapitre, nous mobilisons ces différents outils, et en particulier le SINAMET, sur quatre cas d'étude. Dans les temps de nos travaux, nous n'avons pas pu mener une étude complète du métabolisme d'un territoire. Nous avons préféré concentrer notre attention sur différentes études au périmètre plus restreint, mais qui permettent de montrer la validité et le potentiel de notre approche et sa capacité à être généralisée, et aussi de mieux cerner ses limites. Les cas d'études retenus sont ainsi :

- La consommation d'énergie du patrimoine bâti de l'Eurométropole de Strasbourg
- Le transport de marchandises via les installations portuaires de Strasbourg
- La sensibilité à l'échelle de certains indicateurs de la méthodologie EWMFA
- Le circuit de la matière alimentaire à Strasbourg

Chapitre 5 Consommation d'énergie du patrimoine bâti de l'Eurométropole de Strasbourg

Nous avons mené il y a quelques années une étude de la consommation d'énergie du patrimoine bâti de l'Eurométropole de Strasbourg (EMS), notamment à des fins de réalisation d'un bilan carbone (Ribon 2013). Cette étude avait déjà montré les difficultés inhérentes au traitement des données, notamment les difficultés de rétro-analyse des résultats, le manque de reproductibilité de l'exercice et les divergences d'interprétation selon l'unité de restitution. L'approche retenue pour apporter une réponse à ces problématiques fut de développer un programme dédié (des macros) visant à faciliter le traitement des données et l'interprétation des résultats.

L'approche développée dans cette thèse est finalement similaire, nous montons toutefois largement en généralité, puisqu'il s'agit de pouvoir traiter l'ensemble des données relatives aux flux de matières et d'énergie à l'échelle d'un territoire. Les données de facturation du patrimoine de l'Eurométropole de Strasbourg sont ainsi un cas d'application possible de nos travaux, que nous avons souhaité reprendre pour l'approfondir. Nous nous inscrivons dans une démarche d'exploration des données (Chapitre I/IV.1) et nous cherchons à l'extraire de l'information exploitable et utile à partir des données à travers une diversité de représentations. Nous n'avons ainsi pas de problématique spécifique a priori, mais plutôt diverses interrogations qui vont nous aider à explorer et assimiler ces données.

Nous commençons ainsi par définir le cadrage de l'étude, c'est-à-dire le périmètre et les données sources, puis nous expliquerons comment le SINAMET est mobilisé sur ce cas d'étude. Nous présentons ensuite les résultats et nous concluons par une discussion sur notre approche.

I. Cadrage de l'étude

1) Périmètre

L'étude concerne le patrimoine bâti¹⁰⁶ géré par la Ville ou l'Eurométropole de Strasbourg (VdS, EMS), c'est-à-dire ici, l'ensemble des bâtiments pour lesquels elles assurent le paiement des factures d'énergie. S'il s'agit de deux institutions juridiquement distinctes, les services administratifs de la ville et de la métropole sont fusionnés. Par abus de langage, nous parlerons de « la » collectivité en référence aux deux entités réunies.

Comme nous l'indiquions précédemment (Chapitre 3/I), la définition du périmètre est dépendante des données sources que nous avons à disposition. Nous présentons les jeux de données mobilisés dans la section suivante, cela nous permet de formaliser le périmètre de cette étude dans le Tableau 7.

Territoire // Niveau de détail	Eurométropole de Strasbourg // Bâtiment <i>Uniquement 1055 bâtiments directement gérés par la collectivité</i>
Période // Pas de temps	2009 à 2017 // journalier ou mensuel selon les données <i>Le pas de temps journalier est défini par les dates des factures</i>
Matières ou énergies étudiées	Flux énergétiques : Électricité, gaz, chauffage urbain, fioul
Unités de quantification	<u>Flux</u> : kWh EF ; kWh EP ; t eq. CO2 ; euro <u>Besoin en chauffage</u> : DJU18 <u>Bâtiment</u> : m ² SHON pour les bâtiments
Processus spécifiques étudiés	La consommation d'énergie du patrimoine bâti de la collectivité

Tableau 7 : Périmètre de notre analyse sur les données de facturation d'énergie

2) Données sources

Les données sources principales de cette étude proviennent des services de la collectivité. Nous avons dans un premier temps récupéré uniquement la liste des factures d'énergies (Tableau 8), puis dans un deuxième temps, notamment en raison des données incomplètes (voir

¹⁰⁶ Nous excluons donc les informations de consommation liées au patrimoine non bâti (feux de signalisation, éclairage public...)

Figure 18), les services nous ont transmis le bilan énergie (Tableau 9) et la liste du patrimoine bâti (Tableau 10). Pour compléter ces données principales, nous mobilisons les données climatiques qui reflètent le besoin en chauffage à travers des degrés-jours unifiés (DJU18¹⁰⁷) (Tableau 11). Nous mobilisons également des données non structurées, notamment des coefficients pour chaque type d'énergie : émissions de gaz à effet de serre¹⁰⁸ et ratio Énergie Primaire (EP) / Énergie Final (EF)¹⁰⁹.

Nom de la source :	Factures énergie EMS	Producteur :	EMS
Détails :	Factures d'énergie pour chaque point de comptage du patrimoine (bâti et non bâti), distinguant EMS et Ville de Strasbourg		
Unité de quantification :	kWh final (PCS si applicable) ; euros TTC		
Période :	2009 - 2017	Pas de temps :	Date facture
Territoire couvert :	EMS	Niveau de détail :	Par point de comptage
Classification des produits :	Parmi : Chauffage urbain / Electricité (tarif bleu, vert, jaune) / Gaz naturel / Fioul		
Format des données :	1 fichier .xlsx (liste d'enregistrements) (1 fichier)		
Synthèse des données :	145797 lignes ; 754 GWh ; 92 millions d'€ ; plus de 2200 adresses		
Obtention :	Demandé à l'EMS, auprès du Département Approvisionnement Énergie et Magasin (Mission Achats Opérationnels)		

Tableau 8 : Données de facturation du patrimoine de l'EMS

¹⁰⁷ Les DJU (référence chauffage 18°C pour notre étude) permettent de quantifier le besoin quotidien en chauffage. Ils représentent la somme des différences quotidiennes entre la référence et la moyenne des températures extrêmes. Ainsi, une journée de température moyenne 5°C ajoute 13 DJU18. Seules les valeurs de températures < 18°C (car chauffage, il existe un concept similaire pour la climatisation) sont comptabilisées.

¹⁰⁸ Il s'agit des coefficients de 2015 disponibles sur <http://www.bilans-ges.ademe.fr/>. Le coefficient d'émission du chauffage urbain (218 g eq. CO2/kWh) est une valeur moyenne estimée à partir des données règlementaires : Arrêté du 22/03/2017 modifiant l'arrêté du 15/09/2006 relatif au diagnostic de performance énergétique [...].

¹⁰⁹ Il vaut 1, sauf pour l'électricité où il vaut 2,58 : Arrêté du 26/10/2010 relatif aux caractéristiques thermiques et aux exigences de performance énergétique des bâtiments nouveaux et des parties nouvelles de bâtiments.

Nom de la source :	Bilan énergie & eau EMS	Producteur :	EMS
Détails :	Consommation de chaque bâtiment distinguant EMS et VdS		
Unité de quantification :	kWh final (PCS si applicable) ; euros TTC ; m3 (pour eau froide)		
Période :	2015 - 2017	Pas de temps :	Mensuel
Territoire couvert :	EMS	Niveau de détail :	Par bâtiment
Classification des produits :	Parmi : Chauffage urbain / Électricité (tarif bleu, vert, jaune) / Gaz naturel / Fioul / Eau froide		
Format des données :	1 fichier .xlsx (mix entre liste d'enregistrements et tableau 2D) (1 fichier)		
Synthèse des données :	15250 lignes, 97 millions d'€		
Obtention :	Demandé à l'EMS, auprès du Département Approvisionnement Énergie et Magasin (Mission Achats Opérationnels)		

Tableau 9 : Données du Bilan énergie et eau patrimoine de l'EMS

Nom de la source :	GIMA	Producteur :	EMS
Détails :	Informations sur chaque bâtiment du patrimoine (catégorie, surface, année d'acquisition, direction...)		
Unité de quantification :	m ² SHON (Surface Hors Œuvre Nette)		
Période :	/	Pas de temps :	/
Territoire couvert :	EMS	Niveau de détail :	Bâtiment
Classification des produits :	/		
Format des données :	1 fichier .xlsx (liste d'enregistrements) (1 fichier)		
Synthèse des données :	1794 bâtiments, 1,6 million m ² de surface totale		
Obtention :	Demandé à l'EMS, auprès du Service Ingénierie de la Construction		

Tableau 10 : Données sur le patrimoine bâti de l'Eurométropole de Strasbourg (EMS)

Nom de la source : DJU Entzheim	Producteur : Cegibat
Détails : Rigueur climatique / Besoins en chauffage	
Unité de quantification : Degrés-Jour Unifiés chauffage référence 18°C (DJU18)	
Période : 2009 - 2017	Pas de temps : Mensuel
Territoire couvert : Station météo Entzheim (67)	Niveau de détail : /
Classification des produits : Non applicable	
Format des données : 1 fichier .xls (tableau 2D) (1 fichier)	
Synthèse des données : En moyenne 2715 DJU18 / an ; min/max = [2309 ; 3138]	
Obtention : https://cegibat.grdf.fr/simulateur/calcul-dju	

Tableau 11 : Données climatiques (DJU18)

II. Utilisation du SINAMET

1) Préparation des données

La préparation des données sources (Chapitre 1/II.3) nous amène à remarquer différents problèmes à résoudre avant de pouvoir pleinement utiliser les données. Il y a des problèmes de date, d'adresse, de jointure, ou encore des données de consommation négatives ou nulles.

a) Gestion des dates

La première typologie de problème sur les données sources concerne les périodes de facturation. Par exemple, pour 176 factures, cette période est négative, c'est-à-dire avec une date de fin antérieure à la date de début. D'autres factures présentent des dates inconsistantes comme le « 13/10/2140 » dont il s'agit plus probablement du « 13/10/2014 ». Nous n'avons pas toutefois pas les moyens de corriger manuellement ces données, nous décidons donc de les exclure de notre analyse, malgré le biais que cela entraîne.

Par ailleurs, notre convention pour définir les périodes est d'inclure les dates de début et de fin. Nous pourrions décider que la date de fin est exclue. Justement, dans les données sources, cette convention n'est ni explicite, ni constante, et varie d'une facture à l'autre parfois pour un

même contrat. Ainsi, nous avons dû développer un algorithme spécifique pour analyser les dates de chaque contrat d'énergie et éventuellement corriger leur enchaînement. Dans le cas du fioul, il n'y a même pas de période facturée : ce sont simplement les jours de livraisons qui sont indiqués. Aussi, les périodes associées sont par approximation définies comme étant entre deux livraisons. Cet algorithme était nécessaire pour éviter des perturbations numériques et graphiques liées à un mauvais conventionnement de la définition des périodes (voir Figure 17).

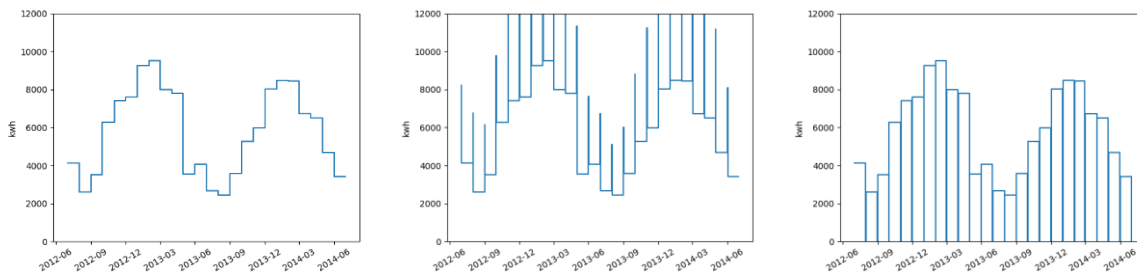


Figure 17 : Conséquences graphiques d'une mauvaise convention de définition des périodes sur la consommation d'un bâtiment quelconque, a) résultat attendu, b) périodes qui se chevauchent sur un jour, c) périodes écartées d'un jour

Les problèmes présentés précédemment sont facilement détectables par un algorithme dédié à la préparation des données. Leur correction n'est toutefois pas toujours possible automatiquement. Par ailleurs, d'autres dates pourraient être erronées sans que nous n'ayons les moyens de le vérifier. Plusieurs techniques pourraient être mobilisées pour renforcer la qualité des données. Toutefois, le temps nécessaire à l'implémentation de ces techniques n'est pas justifié dans nos travaux, qui s'accommoderont donc de ces incertitudes.

b) Géocodage de l'adresse

Les données sources font référence à l'adresse des bâtiments, ce qui peut être utile pour des productions cartographiques, ou des analyses par quartier. Toutefois, plusieurs problèmes ont été rencontrés en ce qui concerne le géocodage des adresses :

- les adresses du fichier initial étant abrégées (par exemple « bld » au lieu de « boulevard »), il a fallu les retravailler pour les avoir de façon complète et exploitable,
- le moteur de géocodage utilisé (Nominatim¹¹⁰) manquait de souplesse, sensible parfois aux virgules mêmes,

¹¹⁰ Outil de recherche d'adresses exploitant les données d'OpenStreetMap <https://nominatim.openstreetmap.org/>

- certains noms de rue étaient erronés, que ce soit dans le fichier initial, ou même dans la base de données de recherche Nominatim,
- certaines adresses ont disparu, suite notamment à des opérations de réaménagement urbain.

Ces différents problèmes n'ont pas empêché de géocoder l'essentiel des adresses, à quelques exceptions près, mais au prix d'un important travail manuel. D'autres outils plus performants que Nominatim ou comme la Base Adresse Nationale¹¹¹, pourraient aider à réduire significativement l'effort associé au géocodage des adresses.

c) Données de consommation négatives ou nulles

Plusieurs factures présentent des consommations (en kWh ou en €) négatives ou nulles. Plusieurs explications sont possibles :

- il s'agit d'une régulation sur des consommations précédentes estimées trop à la hausse,
- il s'agit d'installations sur l'espace public (feux de signalisation, éclairage...) dont on suppose le fonctionnement constant, la facturation est ainsi forfaitaire et la consommation en kWh n'est pas précisée dans les données,
- les données de facturation du gaz (Tableau 8) ne sont plus disponibles en kWh à partir de septembre 2015, le logiciel de facturation aurait changé.

Les installations non bâties (feux de signalisation, éclairage...) ne sont pas incluses dans notre étude. Pour le reste, nous intégrons quand même les données négatives ou nulles qui peuvent être nécessaires pour compléter ou corriger certaines consommations.

d) Jointure des données

Chaque bâtiment est normalement identifié dans une base de données appelée GIMA (Tableau 10), toutefois, les données de facturation dont nous disposons (Tableau 8) n'incluent pas de clef de jointure permettant de faire référence à l'instance GIMA. Après plusieurs demandes auprès du service de la collectivité, nous avons finalement obtenu les données du bilan énergie qui incluent la clef de jointure (Tableau 9). Ces données ne sont toutefois pas équivalentes aux données de facturation. Nous avons dû ainsi nous résoudre à devoir jongler entre deux configurations exclusives :

¹¹¹ <https://adresse.data.gouv.fr/> disponible sous Licence Ouverte depuis le 1er janvier 2020.

- une définition des bâtiments pour chaque adresse de facturation unique du Tableau 8, en y associant la facture correspondante du même tableau,
- une définition des bâtiments à partir des données GIMA du Tableau 10, en y associant les données du bilan énergie du Tableau 9.

L'usage des données de la base GIMA nous permet d'accéder à des informations sur la surface et la typologie des bâtiments. Néanmoins, la qualité des données du bilan énergie est moindre et la période couverte plus restreinte, comme nous le verrons dans les résultats.

2) Chargement du contexte et des données de consommation

Une fois les problèmes mentionnés précédemment résolus (ou contournés), il s'agit de charger les données préparées au sein de la base centralisée du SINAMET. Nous chargeons d'abord les données de contexte (les produits, les acteurs et les territoires), puis les données quantitatives sur les flux (Chapitre 4/II.4).

La définition des produits est ici assez simple, puisqu'il n'y a que quatre types de flux (chauffage urbain, électricité, gaz naturel ou fioul), l'électricité étant toutefois dissociée en trois catégories selon son tarif : bleu, jaune et vert (par ordre croissant de puissance). L'eau froide, dont les consommations sont indiquées dans le bilan énergie et eau (Tableau 9) est également intégrée, même si les données ne sont pas exploitées ici. Ces produits sont hiérarchisés dans une nomenclature spécifique à l'exercice que nous appellerons arbitrairement « FluidesBatiment ».

Les acteurs sont au nombre de deux : la Ville et l'Eurométropole de Strasbourg (VdS et EMS). Dans un modèle plus avancé, nous pourrions distinguer les différentes directions qui se répartissent l'usage du patrimoine.

Enfin, nous définissons chaque bâtiment comme étant un territoire (selon l'une des deux configurations présentées précédemment), que nous pouvons associer à chaque commune de l'Eurométropole. Notre modèle territorial est donc constitué de trois niveaux : EMS, Communes, Bâtiments.

Une fois les données de contexte importées dans la base de données, nous pouvons ajouter les données de consommations. Ainsi, chaque facture ou ligne du bilan énergie (selon la configuration) est associée à un produit énergétique, à la Ville ou à l'EMS, et à un bâtiment. Les quantités, en kWh EF et en euros, sont également précisées.

III. Résultats obtenus

Tout ce travail préalable de collecte, de préparation, et de stockage des données étant achevé, nous allons pouvoir analyser les données en nous appuyant sur leur représentation. Notre objectif est d'explorer ces données, en jouant avec les différentes dimensions du métabolisme et avec différentes techniques de visualisation (Chapitre 3/III). Ces analyses sont toutes réalisées à l'aide du SINAMET¹¹².

1) Vision globale de la consommation d'énergie

Nous présentons Figure 18 le profil de consommation global du patrimoine, c'est-à-dire une représentation de la consommation en fonction du temps. Nous représentons les deux jeux de données présentés Tableau 8 et Tableau 9.

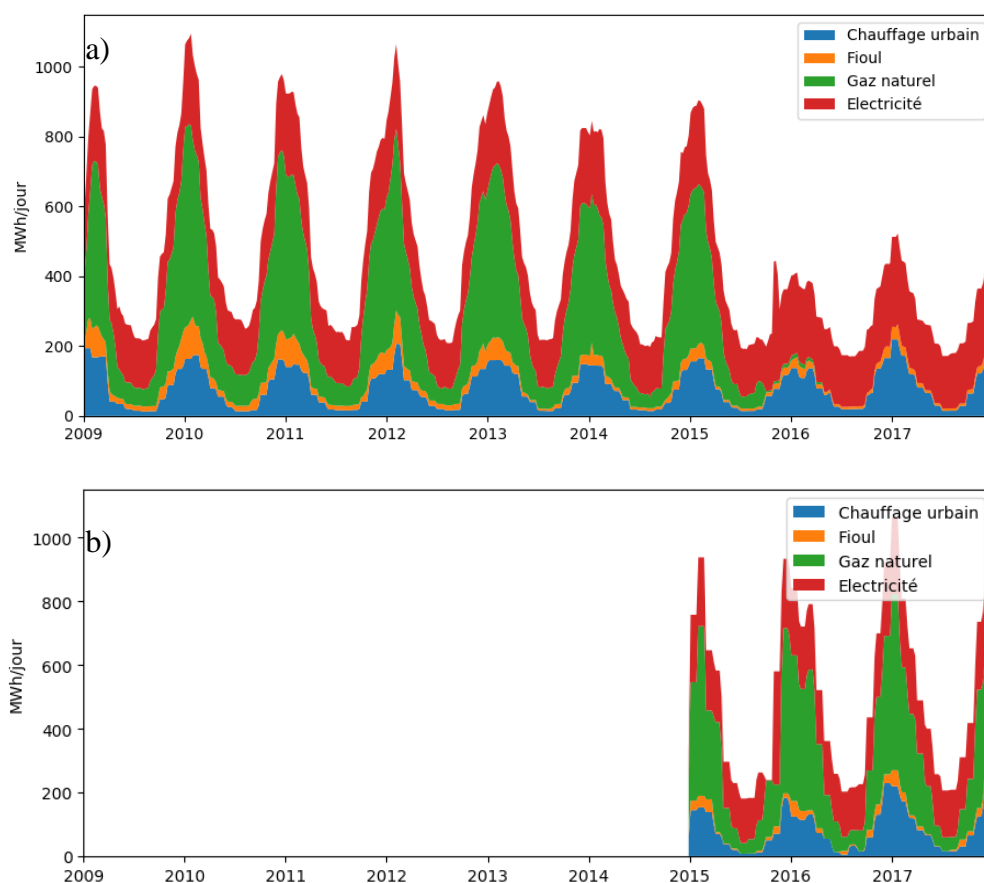


Figure 18 : Profils de consommation (énergie finale) du patrimoine bâti, (a) d'après les données de facturation, Tableau 8 (b) d'après le bilan énergie,

¹¹² Nous avons en plus utilisé un Système d'Information Géographique pour représenter sur une carte les données de la Figure 24 p. 158.

Ce profil de consommation nous permet de remarquer que :

- L'essentiel des données de consommation de gaz (en kWh) est manquant dans les données de facturation à partir de septembre 2015. Le bilan énergie ne contient quant à lui pas de données antérieures à 2015. Sauf mention explicite, ce seront les données de facturation qui seront utilisées.
- Le fioul est marginal dans la consommation globale.
- La consommation d'énergie est très variable au cours de l'année : cela est lié à la variation des températures et met en évidence l'importance du chauffage dans la consommation globale. Nous notons une base fixe d'environ 200 MWh/jour et des pics pouvant aller jusqu'à 1000 MWh/jour.

Nous précisons les profils spécifiques du gaz et de l'électricité Figure 19, celui du chauffage urbain étant bien visible Figure 18. Ainsi, nous remarquons que le chauffage urbain et le gaz connaissent une grande variabilité saisonnière : la consommation en hiver est plus de 10 fois celle en été. A contrario, les variations de la consommation d'électricité sont plus réduites : elle est multipliée par environ 1,5 entre l'été et l'hiver.

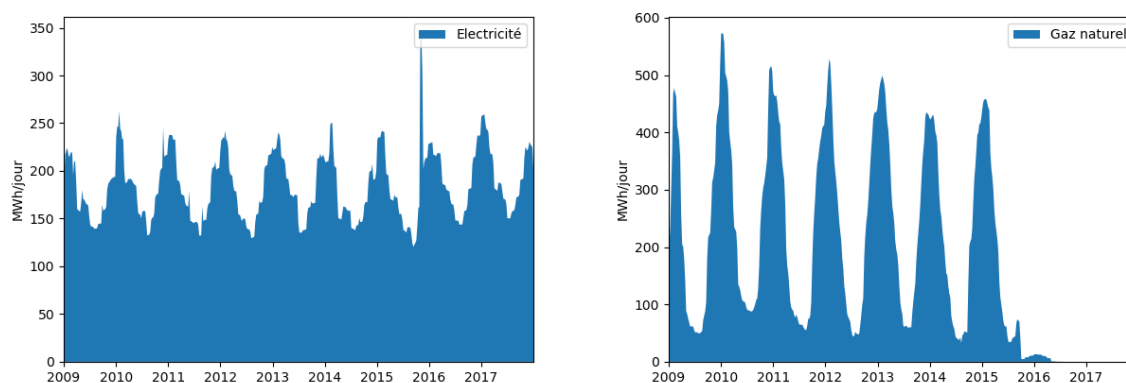


Figure 19 : Profil de consommation (énergie finale) du patrimoine bâti pour le gaz et de l'électricité.

Nous représentons également le bilan sur 2014 de la part des différents types de fluides énergétiques dans la consommation globale, pour différentes unités (Figure 20). Cette représentation permet d'apporter d'autres éléments de compréhension :

- Au regard de la consommation globale, la part variable de consommation¹¹³, en grande partie liée au chauffage, représente 58 % de la consommation totale.

¹¹³ Calcul : Part Variable (100 GWh en 2014) = Total (173 GWh en 2014) – Part fixe (200 MWh/jour)

- La priorisation de la réduction des consommations en fonction du type de fluide n'est pas évidente, notamment vis-à-vis de l'électricité : s'il s'agit d'un type d'énergie faiblement émetteur de carbone (en raison du parc nucléaire), c'est aussi une énergie qui coûte plus cher et dont le rendement global (EF / EP) est plus faible.

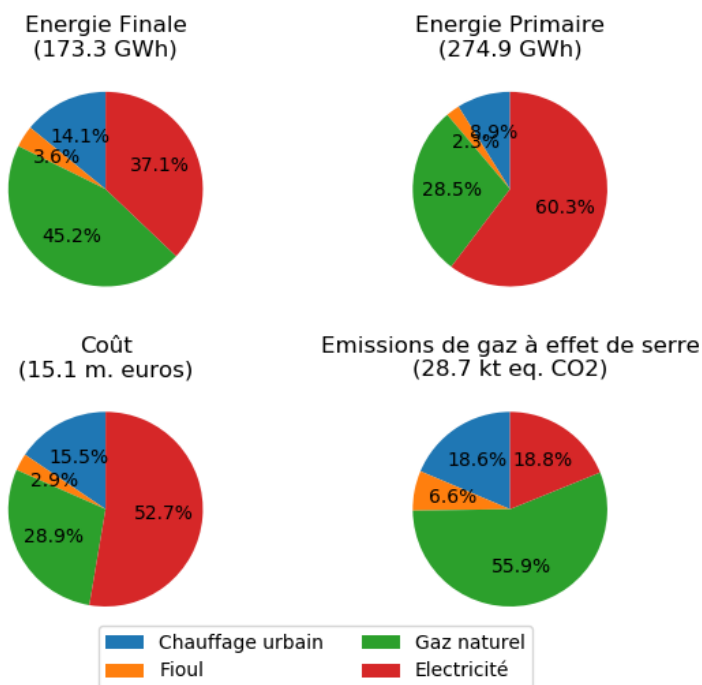


Figure 20 : Répartition des consommations du patrimoine bâti selon l'unité (en 2014).

Pour compléter la question du coût de l'énergie, nous représentons son évolution globale pour les différentes énergies (Figure 21). Nous remarquons la forte croissance (récemment stabilisé) du prix de l'électricité, là où celui des autres énergies est resté plus stable sur les dix dernières années.

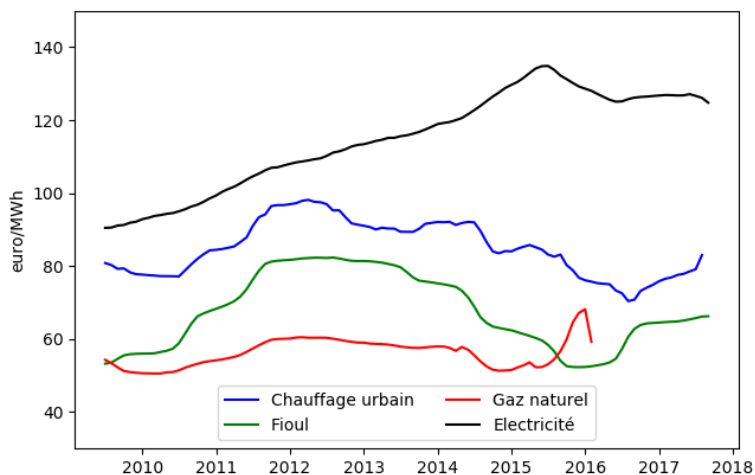


Figure 21 : Évolution du coût moyen des différents types d'énergie (lissé sur 12 mois).

Si nous regardons plus spécifiquement l'électricité (Figure 22), nous remarquons également la saisonnalité de son prix qui est plus important en hiver¹¹⁴. C'est contracyclique à d'autres types de flux comme le chauffage urbain (Figure 23) qui est caractérisé par un coût d'abonnement important et donc plus significatif par kilowattheure lorsque l'on consomme moins (en été).

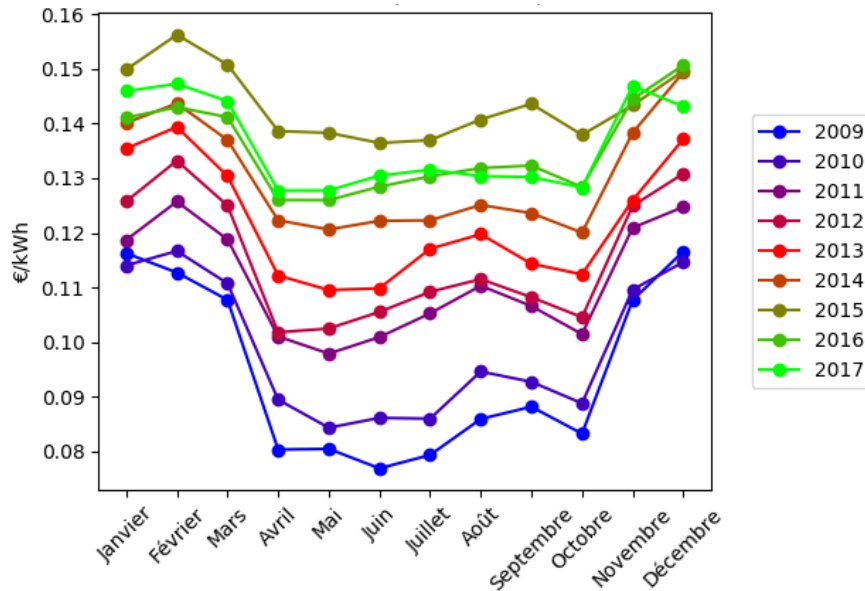


Figure 22 : Coût mensuel moyen de l'électricité.

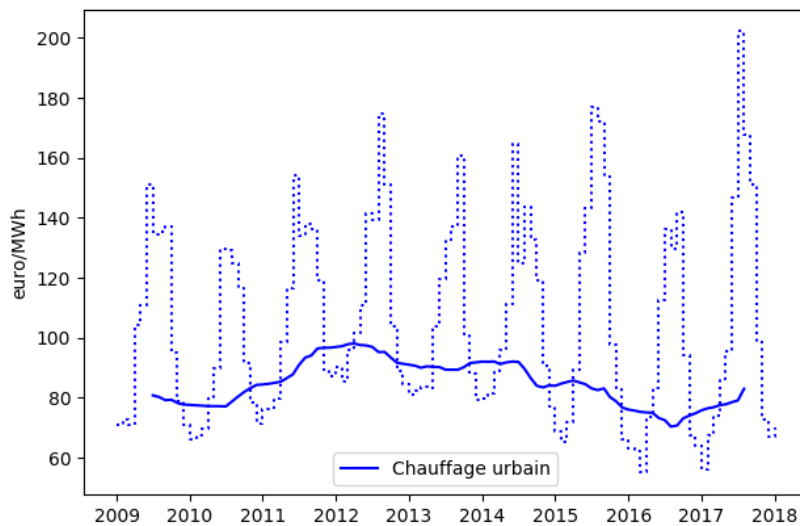


Figure 23 : Coût moyen du chauffage urbain.

En pointillé : mois par mois – En continu : lissé sur 12 mois.

¹¹⁴ Cela ne concerne toutefois que les bâtiments qui nécessitent une forte puissance électrique. Le phénomène est inversé pour les petits bâtiments (bureau et logements) pour qui le kWh en hiver coûte moins cher qu'en été.

Nous poursuivons la vue d'ensemble à travers une cartographie de la consommation d'énergie (Figure 24). Cela permet de faire ressortir dix-sept bâtiments dont la consommation dépasse 1,5 GWh/an : piscines, station de pompage et de relevage d'eau, centre administratif et autres grands équipements : halls, musées, cités...

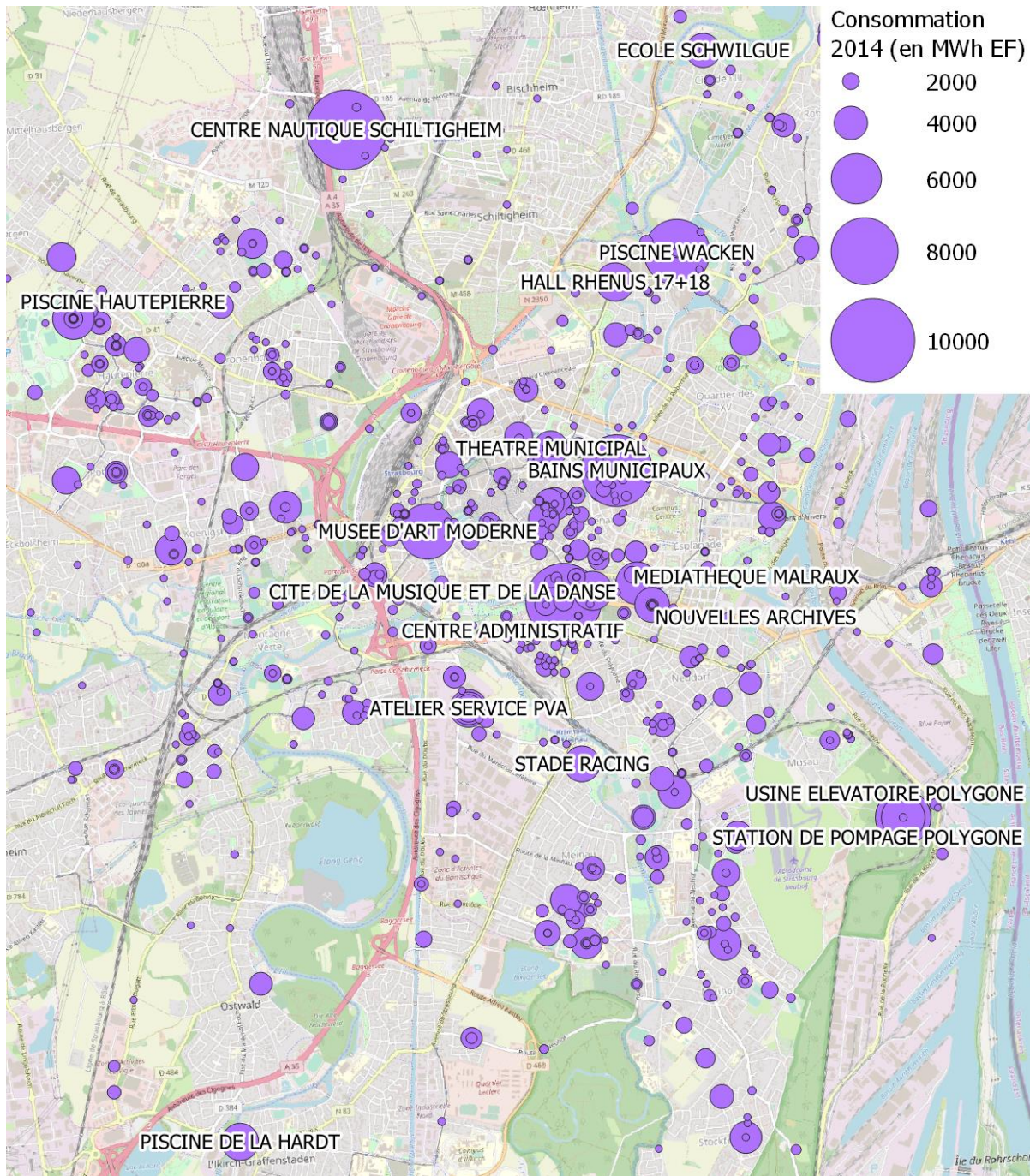


Figure 24 : Consommation d'énergie (kWh EF) du patrimoine (en 2014). Les 17 bâtiments nommés sont ceux ayant une consommation d'énergie supérieure à 1500 MWh EF.

En regardant la consommation cumulée des bâtiments (Figure 25), nous notons que ces dix-sept bâtiments représentent plus d'un tiers de la consommation totale du patrimoine bâti, et 10 % des 1037 bâtiments pèsent pour 75 % la consommation¹¹⁵.

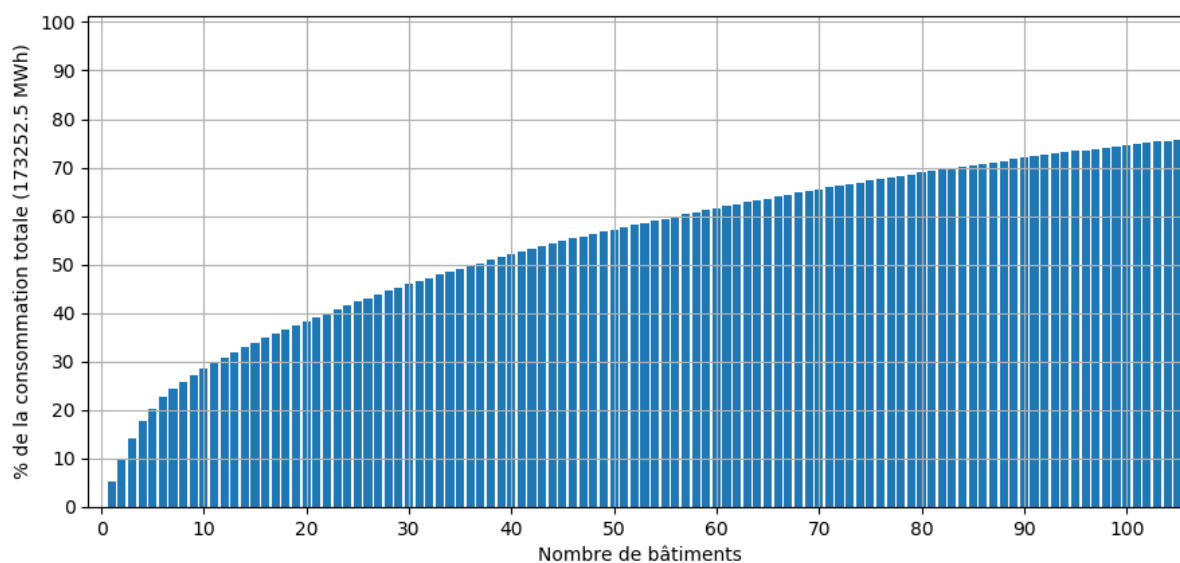


Figure 25 : Consommation cumulée des bâtiments les plus importants (en 2014).

Lecture : 50 % de l'énergie consommée par l'ensemble du patrimoine (1037 bâtiments au total) est à destination des 36 bâtiments les plus consommateurs.

2) Recherche de spécificités

Les résultats obtenus précédemment nous permettent d'avoir une vision plutôt générale de la consommation d'énergie par le patrimoine, même si nous avons déjà identifié des bâtiments spécifiques : les plus consommateurs d'énergie. Dans cette section, notre objectif est plutôt d'aller à la recherche d'anomalies dans la consommation d'autres bâtiments particuliers.

Nous observons Figure 16 un prix général de l'électricité saisonnier. Ce prix n'est toutefois pas homogène pour tous les bâtiments. En récupérant le prix pour chaque mois pour chaque bâtiment, nous mettons en avant certains excès. L'exemple le plus marquant est celui des « Vestiaire et tribune du stade de l'III » dont le prix de l'électricité peut dépasser les 250 €/MWh (Figure 26), bien au-delà du prix moyen constaté de 150 €/MWh en hiver (Figure 22). Nous ignorons toutefois la cause. Nous soupçonnons un dépassement régulier de puissance souscrite, mais il ne nous est pas possible de le vérifier avec les données disponibles.

¹¹⁵Cela n'est pas sans rappeler le principe de Pareto qui annonce qu'environ 80 % des effets sont le produit de 20 % des causes.

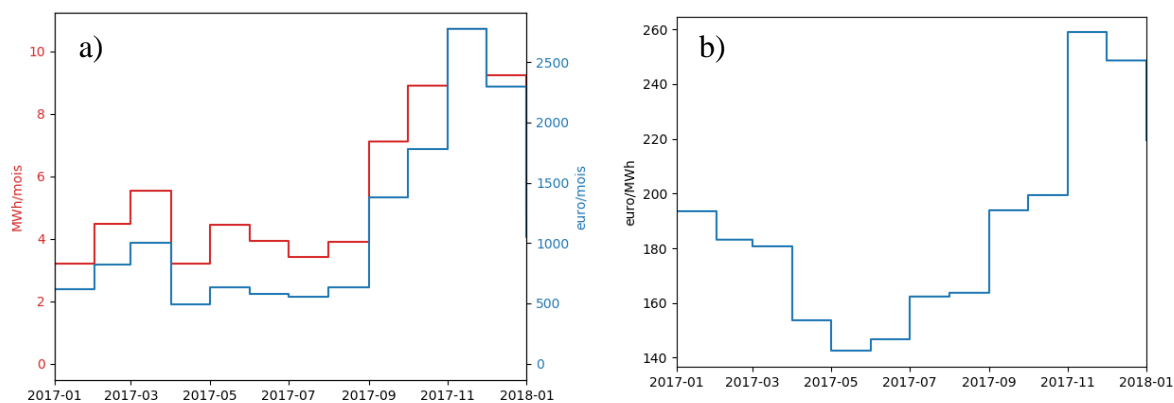


Figure 26 : Consommation électrique des « Vestiaire et tribune du stade de l'Ill » en MWh et euros (a) et coût moyen de l'électricité en euro/MWh (b)

Pour poursuivre la recherche de bâtiments atypiques, nous examinons le ratio de consommation d'énergie rapporté à la surface du bâtiment et selon leur typologie. Les données suivantes sont ainsi obtenues sur la base du bilan énergie et eau, ainsi que des données GIMA. Nous représentons cette approche Figure 27 en précisant le nom des bâtiments les plus remarquables. Cette figure doit toutefois être lue avec une grande précaution : ces bâtiments remarquables n'ont pas forcément une consommation anormale. Notamment, ils peuvent être liés à des installations extérieures (accueils, centres sportifs), or, nous les rapportons à des mètres carrés intérieurs (par exemple, le pavillon d'accueil qui apparaît en haut de la figure ne fait que 26 m²). D'autre part, la surconsommation du stockage des musées est liée à une exigence de régulation stricte de température et d'hydrométrie, et celle de la station de pompage est liée au fonctionnement permanent de pompes pour le réseau d'eau. En fait, les anomalies ont plutôt tendance à s'afficher dans la tranche 0-2000 kWh/an. En zoomant sur une partie de cette tranche, nous remarquons la diversité des consommations des écoles maternelles, des piscines et des vestiaires sportifs. Les piscines (et probablement les vestiaires) sont des équipements aux dynamiques particulières avec des consommations importantes et qui méritent donc une analyse spécifique (Ribon 2013) mais que nous ne mènerons pas ici. En revanche, la grande diversité des intensités de consommations des écoles maternelles, dont certaines dépassent 1000 kWh/m²/an, nous pose beaucoup plus question¹¹⁶. Toutefois, à nouveau, à partir des seules données, nous ne pouvons apporter d'explications. Pour cela, il faudrait s'assurer de la véracité des données reflétant la consommation et la surface et mener une enquête de terrain.

¹¹⁶ Pour apporter un point de comparaison, les bâtiments résidentiels sont considérés comme énergivores (classe G et dernière du Diagnostic de Performance Énergétique) à partir de 450 kWh/m²/an.

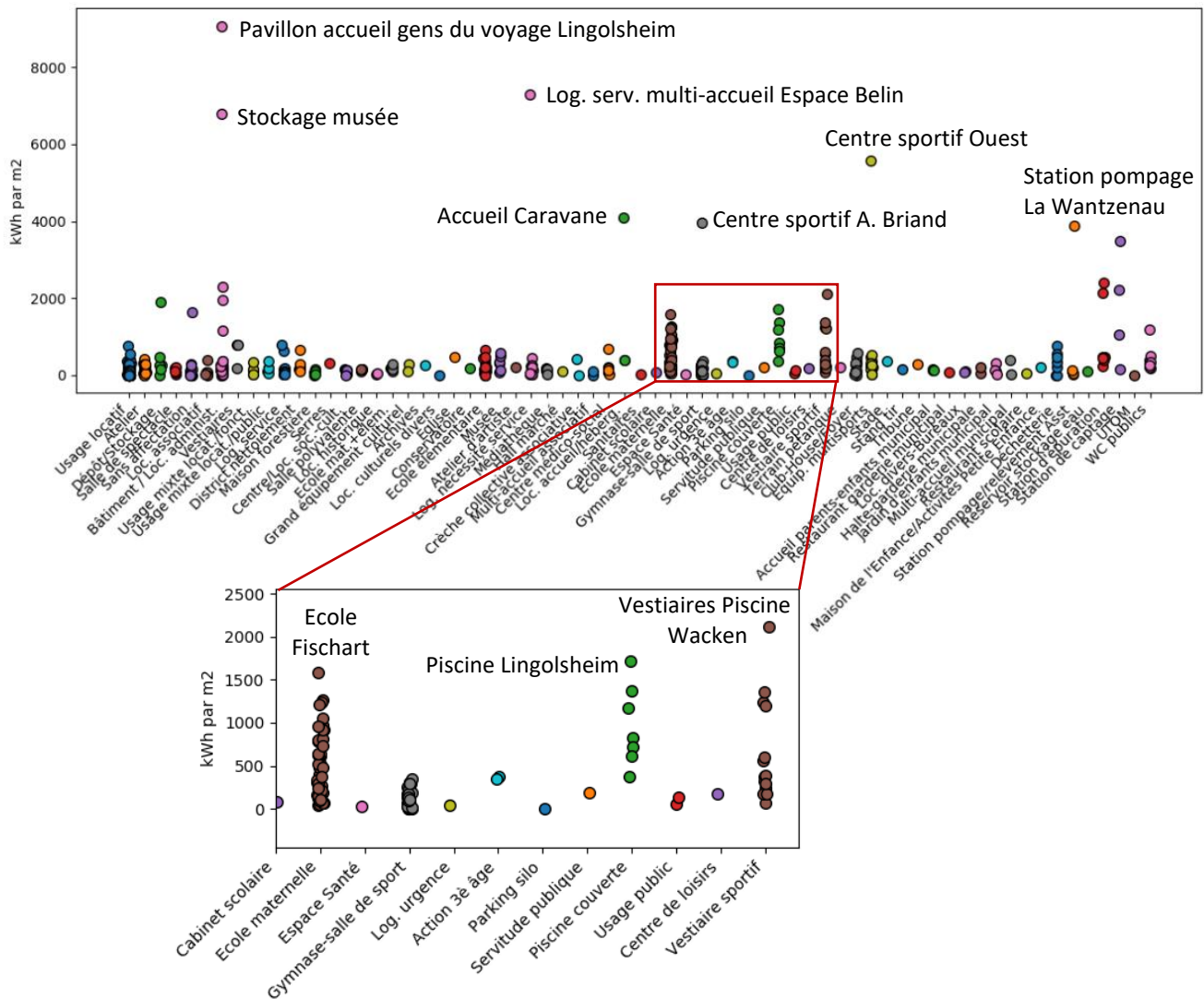


Figure 27 : Consommation des bâtiments gérés par la collectivité en 2016 (en kWh EF/m²) selon leur typologie (un point correspond à un bâtiment)

Nous terminons notre recherche d'anomalie par une analyse de la sensibilité climatique des bâtiments. La consommation d'énergie des bâtiments est en grande partie liée au chauffage, cette partie est donc sensible aux variations climatiques. Cela se voit particulièrement sur la Figure 28 qui représente la consommation d'un bâtiment (le centre administratif, qui est le plus gros consommateur du patrimoine) en fonction des besoins de chauffage (quantifiés en DJU18), mois par mois. La corrélation linéaire n'est pas parfaite (en 10 ans, des progrès ont été réalisés pour réduire la consommation d'énergie) mais visible. Pour ce bâtiment, nous identifions par régression linéaire une part variable (une pente d'environ 1,75 MWh/DJU18), essentiellement due au chauffage, et une part fixe (400 MWh/mois) qui concerne plutôt l'électricité spécifique : éclairage, parc informatique, autres appareils électriques... (Ribon 2013).

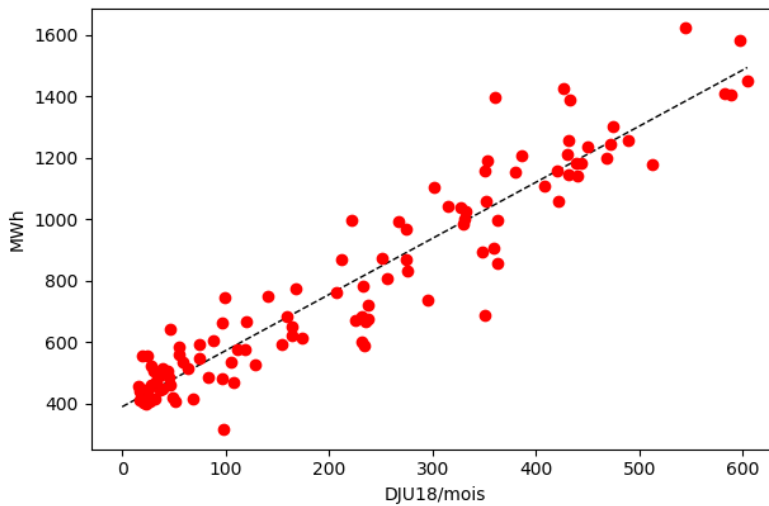


Figure 28 : Sensibilité climatique du centre administratif de l'Eurométropole de Strasbourg, pour chaque mois entre 2009 et 2017.

Pour généraliser l'analyse, nous calculons cette régression linéaire pour chaque bâtiment en 2016, rapportée à la surface bâtie. Cela nous permet de construire la Figure 29 montrant la part variable et fixe des différents bâtiments du patrimoine. Plusieurs bâtiments émergent ainsi. Le centre administratif présente la plus grande part variable du patrimoine et une part fixe importante. Les piscines, musée, halls, station de captage sont également des bâtiments aux dynamiques particulières et pour lesquels l'usage de ratio par mètre carré n'est pas pertinente, ou du moins à discuter. Ce qui nous interpelle est le nombre de bâtiments avec une part fixe négative, et même quelques-uns avec une part variable négative. Ce n'est pas a priori cohérent. La mensualité des consommations présentées dans le bilan énergie semble être en fait celle de la date des factures, et non des périodes facturées. Nous n'avons toutefois pas les moyens d'examiner en détail ce problème, mais cela nous fait douter de la qualité des données sources.

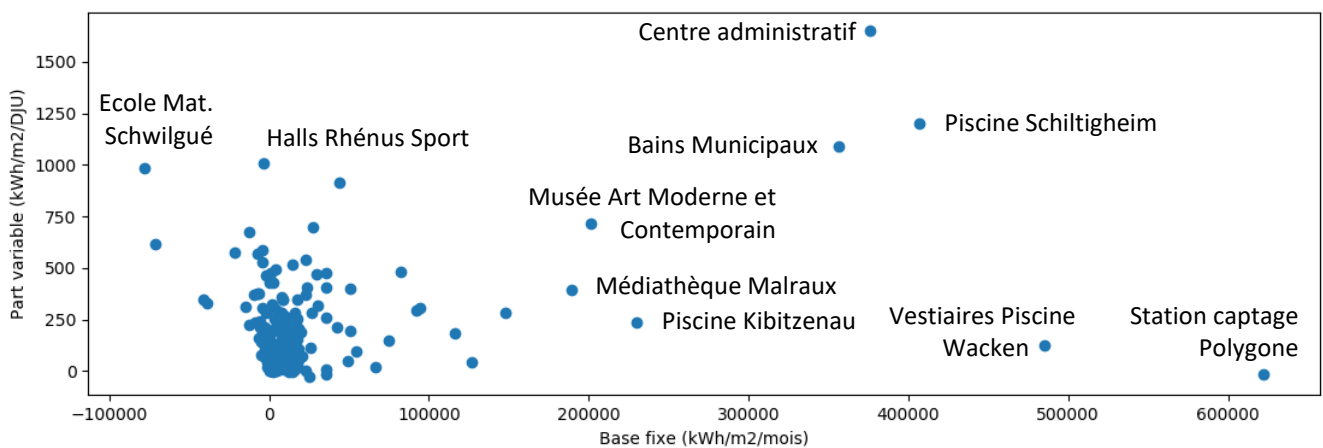


Figure 29 : Sensibilité climatique des différents bâtiments (en 2016)

IV. Discussions sur l'approche mobilisée

L'étude que nous avons menée ici concernait un type de données particulier : la consommation d'énergie du patrimoine bâti d'une collectivité. Il ne s'agissait toutefois pas tant de mener une analyse poussée, mais plutôt d'expérimenter le potentiel applicatif du SINAMET et de nos réflexions sur le traitement des données du métabolisme d'un territoire. Ainsi, au regard des éléments développés dans les premiers chapitres, cette étude nous permet d'appuyer nos propos en montrant :

- la diversité des axes d'analyses possibles,
- l'appui des représentations graphiques et interactives dans une démarche descriptive,
- le potentiel applicatif de ces résultats,
- les difficultés dans le management des données,
- le rôle du SINAMET dans cette étude.

Nous avons pu examiner les données selon différentes dimensions (Chapitre 3/III.1) : quantitatives, selon différentes unités, selon différents types de flux, en distinguant les différents éléments du territoire, ou encore à différents moments. Nous n'avons pas ici distingué les différents acteurs que sont la Ville et l'Eurométropole, car ça n'apporte pas d'information intéressante pour le lecteur peu familier du contexte, mais nous l'avons fait pour l'institution dans un autre rapport. Le moyen de transport n'est pas ici une dimension intéressante à étudier distinctement, il est propre à chaque fluide énergétique (électricité par câble, fioul par camion, gaz et chauffage urbain par conduites dédiées). Il nous reste l'intention d'usage du fluide, qui n'est pas explicite, mais nous arrivons à déduire des informations sur la part (importante) du chauffage dans ces usages.

Nous avons fait le choix d'appuyer essentiellement notre analyse sur des représentations graphiques (Chapitre 3/III.2). Nous voulions montrer au lecteur qu'une grande diversité de représentations peut aider à appréhender une information complexe avec certaines subtilités. Ce qui est moins explicite, mais pourtant également important, c'est la possibilité donnée par le SINAMET de reproduire facilement les figures présentées dans d'autres contextes. Il est ainsi possible de recréer facilement les graphiques présentés précédemment, mais en changeant les données sources ou les paramètres (bâtiment étudié, fluide, période...). Les graphiques créés sont aussi interactifs : il est possible de zoomer dessus (par exemple Figure 27), et

éventuellement d'afficher des labels¹¹⁷. Les possibilités graphiques sont ainsi très grandes ; les résultats présentés n'en sont qu'une petite partie.

Certaines des applications que nous mentionnions (Chapitre 2/IV) sont directement mobilisables ici. Nous observons ainsi une marge de manœuvre en termes d'amélioration des procédés de certains bâtiments comme l'isolation thermique ou le prix de l'énergie. Nous pouvons également évaluer des impacts environnementaux comme la consommation de ressources fossiles ou les émissions de gaz à effet de serre. Bien sûr, les différentes figures réalisées peuvent appuyer une communication de la part de l'institution et aider à construire une vision partagée de la consommation de son patrimoine bâti. La réglementation impose également un suivi et une information sur les bâtiments les plus consommateurs, ce suivi pourrait être réalisé sur la base des analyses présentées.

Si les résultats ont un potentiel intéressant, le travail qui a été mené pose aussi la question du management de la donnée (Chapitre 1/II.3). Ainsi, nous avons décrit nos difficultés pour obtenir les données : elles étaient soit incomplètes, soit de qualité discutable, soit avec des subtilités invisibles sans les connaissances du terrain, comme celui d'équipements spécifiques comme les piscines. Les échanges par mail avec le service producteur des données n'ont pas été très fluides non plus. Le travail de préparation des données a été également important, soulignant une utilisabilité des données sources qui pourrait être améliorée. Cela ne nous a pas empêchés de réussir à extraire de l'information de ces données, même s'il convient toujours d'être prudent sur les résultats d'une analyse menée uniquement sur des données. Ainsi, seule une enquête spécifique et sur le terrain peut nous aider à comprendre les causes réelles de ces dynamiques et déterminer si elles sont normales ou non (Chapitre 1/IV).

Enfin, le SINAMET (Chapitre 4) a joué un rôle important dans cette étude, mais s'il ne transparaît pas directement dans les résultats. La structure prédéfinie des données et les fonctionnalités annexes du logiciel nous ont largement facilité la tâche quant à la construction des différents résultats présentés en offrant une grande intuitivité pour sélectionner et manipuler des données. L'automatisation nous permet également de réutiliser les algorithmes, par exemple les composantes a) et b) de la Figure 18 ont été produites avec *exactement* le même algorithme, ce sont juste les données sources chargées dans le SINAMET qui ont changé. L'ensemble de

¹¹⁷ Ce qui rend en pratique la figure illisible, mais avec les possibilités de zoom, nous avons pu reporter les intitulés des principaux bâtiments.

l'étude est donc très facilement reproductible pour toute organisation devant gérer les consommations d'un ensemble de bâtiments, c'est-à-dire au moins toutes les collectivités.

V. Conclusion du cinquième chapitre

Cette étude permet ainsi d'esquisser le potentiel important de notre approche sur différents volets : compréhension des flux mobilisés, identifications d'anomalies, possibilité de généralisation et de reproduction... Néanmoins, son impact transformatif n'est pas encore démontré : bien que les résultats aient été partagés à des agents de la collectivité, il n'y a pas eu encore d'analyse approfondie sur les différents cas identifiés, alors que potentiellement certaines économies pourraient être facilement accessibles. Finalement, nous revenons bien à nos propos sur le cycle de l'information (Chapitre 1/II.2) : si nous avons effectivement facilité l'analyse des données et la production de l'information, il nous faut encore réussir à transformer cette information en une connaissance qui puisse entraîner la mise en place des actions adéquates.

Chapitre 6 Transport de marchandises par les installations portuaires dans le bassin de vie de Strasbourg

Ce chapitre vient présenter des travaux réalisés dans le cadre du programme de recherche en urbanisme POPSU Métropoles (Plateforme d'Observation des Projets et Stratégies Urbaines¹¹⁸). Ce programme associe des équipes de recherches avec des agents de différentes métropoles françaises et d'autres institutions pour mener un travail d'observation et co-construction d'une compréhension partagée du territoire. À Strasbourg, ce programme s'articule autour de la place des espaces portuaires dans la dynamique de métropolisation.

Notre contribution a notamment été d'étudier les flux de marchandises circulant par voie fluviale et de réaliser différentes cartographies, sur la base de travaux d'étudiants, afin de caractériser différentes dimensions portuaires et en lien avec le territoire (réseaux hydriques, voies logistiques, attraits touristiques, structure économique...). Nous reprenons, en Annexe 6, quelques cartographies qui viennent appuyer nos propos centrés ici sur la thématique des flux. Les autres cartographies réalisées seront éditées sous forme d'un atlas à l'issue du programme.

Nous commencerons cette étude par quelques informations de contextualisation sur les espaces portuaires et le Port Autonome de Strasbourg. Puis nous présenterons la manière dont nous avons mobilisé le SINAMET sur cette thématique, ainsi que les résultats obtenus. Enfin, nous discuterons de ces résultats, des interprétations possibles et des limites de l'exercice.

I. Éléments de contexte

Le port de Strasbourg est une zone d'activité située à l'ouest de Strasbourg, le long du Rhin et donc de la frontière allemande, dont le développement à l'endroit actuel remonte à 1900 (Beyer & Debrie 2011). À cette époque, la ville est rattachée à l'Allemagne (depuis le traité de Francfort de 1919) et connaît une forte urbanisation. Afin d'appuyer cette urbanisation, le port, initialement situé au sud de la ville sur un des bras de l'Ill, vient s'accoler au Rhin, tirant partie

¹¹⁸ <http://www.popsu.archi.fr/popsu-metropoles/accueil>

et contribuant aux travaux de canalisation du Rhin (voir carte Annexe 6/I). Ces travaux permettent de réduire les aléas de son lit, offrant ainsi à Strasbourg la possibilité de se développer à l'est sur des zones auparavant fréquemment inondées. Ils viennent également améliorer sa navigabilité. À l'époque, les bateaux restent de loin le moyen le plus efficace pour le transport de marchandises, malgré le train également en plein développement. Les canaux du Rhône au Rhin et de la Marne au Rhin relient Strasbourg au reste du territoire français depuis 1832 et 1853, et les nouvelles infrastructures construites autour de 1900 (les bassins du commerce et de l'industrie et la capitainerie) viennent ainsi offrir de nouvelles opportunités économiques pour Strasbourg et renforcer son développement. Par la suite redevenu français, le port va poursuivre son développement, doublant même de superficie après la Seconde Guerre mondiale. Aujourd'hui, il est le deuxième port fluvial de France, après Ports de Paris, avec près de 8 millions de tonnes de marchandises qui y transitent annuellement.

La zone portuaire est gérée par le Port Autonome de Strasbourg (PAS), entité administrative sous tutelle conjointe de l'État et de la Ville de Strasbourg. Le PAS gère l'allocation de ses espaces fonciers aux activités industrielles ainsi que l'entretien et l'exploitation d'une partie des infrastructures portuaires, sur le site de Strasbourg, mais aussi à Lauterbourg et à Marckolsheim.

Si la zone du port est encore aujourd'hui à dominante industrielle, une dynamique transformatrice est en cours. En effet, depuis une trentaine d'années, il y a une volonté politique forte de faire de Strasbourg une métropole européenne et transfrontalière. Le rapprochement avec la ville de Kehl, situé de l'autre côté de la frontière se renforce au fil du temps, que ce soit en terme politique, mais surtout urbanistique (Beyer & Lacoste 2017). En 2012, le Conseil municipal valide le Schéma directeur des Deux-Rives qui définit un développement résidentiel le long de l'axe Strasbourg-Kehl, accompagné par une ligne de tramway transfrontalière qui a été inauguré en 2017. De fait, la métropole strasbourgeoise va bientôt connaître une urbanisation continue jusqu'au-delà de la frontière.

II. Cadrage de l'étude

1) Périmètre de l'analyse de flux

Notre étude porte sur les flux transitant par les installations portuaires du Port Autonome de Strasbourg et dans le département du Bas-Rhin (67). Nous focalisons ainsi notre attention sur le transport de marchandises par voie navigable, mais nous aborderons également brièvement le transport routier et ferroviaire.

Le cadrage temporel (2009-2018 ou 2017-2019 selon les données) et spatial (département et infrastructures du PAS) nous est imposé par les données sources que nous avons pu récupérer. Les unités de quantification (tonnes et conteneurs) sont également directement issues de ces données sources. Nous résumons ainsi le périmètre de l'étude dans le Tableau 12.

Territoire // Niveau de détail	1) France métropolitaine, en détail le Bas-Rhin // Département 2) Installations du Port Autonome de Strasbourg // Sans détail
Période // Pas de temps	1) 2009 à 2018 // Annuel 2) 01/2017 à 08/2019 // Mensuel
Matières ou énergies étudiées	Toutes marchandises, éventuellement par conteneurs
Unités de quantification	Tonne ; Conteneurs EVP (Equivalent Vingt Pieds)
Processus spécifiques étudiés	En particulier les marchandises transportées par voie navigable

Tableau 12 : Périmètre de notre analyse sur les flux de marchandises dans les installations portuaires

2) Données sources

Les jeux de données sources mobilisés sont présentés Tableau 13, Tableau 14 et Tableau 15.

Nom de la source : SitraM	Producteur : MTES
Détails : Flux de marchandises, par route et voie navigable	
Unité de quantification : Tonnes ; Tonnes × kilomètres ; Nombre d'observations	
Période : 2009-2018	Pas de temps : Annuel
Territoire couvert : France	Niveau de détail : Département
Classification des produits : Nomenclature NST2007	
Format des données : (12 fichiers)	Transport intérieur : 1 tableur .xlsx (liste d'enregistrements) / an Imports internationaux : 1 tableur .xlsx (liste d'enregistrements) Exports internationaux : 1 tableur .xlsx (liste d'enregistrements)
Obtention : Disponible sur demande et devis auprès du MTES	

Tableau 13 : Données du Système d'Information sur le Transport de Marchandises (SitraM) –
MTES = Ministère de la Transition Ecologique et Solidaire

Nom de la source : Trafics du PAS	Producteur : PAS
Détails : Trafics fluviaux (marchandises et passagers)	
Unité de quantification : Tonnes (marchandises) ou unités (bateaux et passagers)	
Période : 01/2017 – 08/2019	Pas de temps : Mensuel
Territoire couvert : Infrastructures PAS	Niveau de détail : /
Classification des produits : Nomenclature spécifique (dérivée de NST-R, puis NST2007)	
Format des données : (32 fichiers)	1 fichier .pdf (<i>sic !</i>) / mois structuré en liste d'enregistrements
Obtention :	https://www.strasbourg.port.fr/outils/mediatheque/graphiques-et-statistiques-mensuelles/

Tableau 14 : Données sur le trafic fluvial du Port Autonome de Strasbourg (PAS)

Nom de la source :	Activité conteneurs PAS	Producteur :	PAS
Détails :	Entrées et sorties de conteneurs, selon mode de transport (Rhin, Route, Fer)		
Unité de quantification :	Conteneur « Equivalent Vingt Pieds » (EVP)		
Période :	01/2017 – 08/2019	Pas de temps :	Mensuel
Territoire couvert :	Terminaux RET	Niveau de détail :	/
Classification des produits :	Conteneurs chargés et vides		
Format des données :	1 fichier .pdf (<i>sic !</i>) / mois structuré en tableau 2D (32 fichiers)		
Obtention :	https://www.strasbourg.port.fr/outils/mediatheque/graphiques-et-statistiques-mensuelles/		

*Tableau 15 : Données sur l'activité conteneur du Port Autonome de Strasbourg (PAS) –
RET : Rhine Europe Terminal*

3) Utilisation du SINAMET

a) Préparation des données

La difficulté principale dans la préparation des données fut d'exploiter les fichiers statistiques PDF disponibles sur le site du port. Cela a demandé l'implémentation de techniques spécifiques pour l'analyse automatique de ce type de fichier. Cela était plus intéressant que de reprendre les données à la main mais pas forcément plus rapide.

Également dans les données SitraM, nous avons quelques lignes pour lesquelles l'origine ou la destination des marchandises ne sont pas renseignées. Nous décidons d'exclure ces quelques flux de notre analyse.

b) Modèle territorial, des acteurs et nomenclatures

Afin de manipuler les données SitraM, nous devons au préalable charger le modèle territorial et la nomenclature des produits. La notion d'acteur n'intervient pas ici. Ainsi le modèle territorial consiste en l'ensemble des départements français, associé à leur région. Les différents pays du monde sont aussi mobilisés pour les imports et exports. La nomenclature des produits est la NST2007, avec sa hiérarchisation.

Pour les données du PAS, nous n'avons pas la notion de territoire. Tous les flux présents sur les statistiques sont donc associés à l'acteur ad-hoc « PAS ». La nomenclature des produits, bien que liée à la NST2007, est présentée d'une telle façon qu'il nous paraît plus cohérent de créer une nomenclature spécifique à ces données, notamment pour y inclure les conteneurs.

Une fois les données de contexte chargées, il s'agit ensuite de charger les sources présentées précédemment qui décrivent des flux de marchandises. Les données SitraM décrivent des flux de chemin (PathFlow) car nous en connaissons l'origine et la destination, les statistiques PAS décrivent quant à elles des flux de porte (GateFlow) et nous savons seulement s'il s'agit de flux en entrée (/déchargement) ou en sortie (/chargement).

III. Résultats

1) Dans les départements français

Les données de la base SitraM nous permettent d'interroger la place du transport de marchandises¹¹⁹ par voie fluviale en France pour chaque département. Nous représentons cela sous forme d'une carte Figure 30, nous remarquons que seule une petite moitié de la France mobilise les voies d'eau pour le transport de marchandises malgré des apparentes possibilités dans le Sud-Ouest avec le canal de Garonne / canal du Midi, et dans l'Ouest en Bretagne et sur une partie de la Loire.

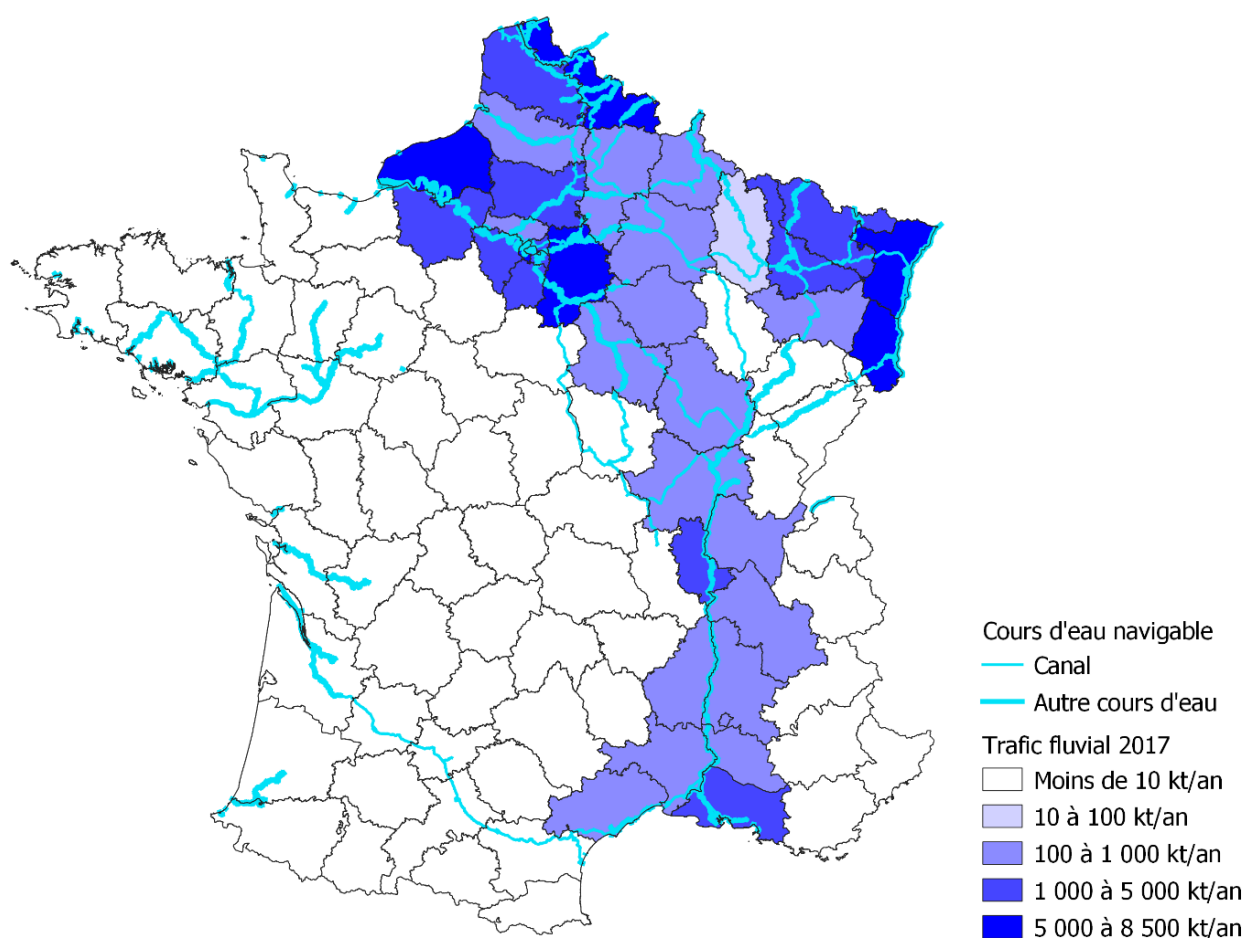


Figure 30 : Trafic fluvial de marchandises (chargement + déchargement) des départements français.

Sources : SitraM 2017 (trafic de marchandises), Système d'Information sur l'Eau 2018 (cours d'eau)

¹¹⁹ Nous considérons le trafic de marchandises comme la somme (en masse) des chargements et des déchargements. Le tonnage réellement transporté est donc généralement la moitié de ce trafic, aux importations et exportations près.

Parmi les départements qui utilisent les voies navigables, nous faisons ressortir les plus spécifiques dans le Tableau 16. La cartographie par département n'est pas vraiment adaptée pour faire ressortir l'important trafic en Île-de-France, en particulier dans les Hauts-de-Seine, en Seine-et-Marne et à Paris. Nous observons dans ce tableau différentes caractéristiques territoriales : la grande part modale du trafic de marchandises par voie navigable, par rapport à la route¹²⁰ pour certains départements, jusqu'à 20 % à Paris et en Hauts-de-Seine, et l'usage important des voies navigables pour du trafic interne au département, donc de proximité, en Seine-Maritime et dans le Rhône.

Département	Part modale du trafic fluvial	Part interne du trafic fluvial	Trafic fluvial (en t)	Trafic fluvial (en t/hab)
Bas-Rhin	11%	7%	8 332 049	3,70 *
Hauts-de-Seine	22% *	2%	5 094 083	1,59
Seine-Maritime	9%	43%	8 333 979 *	3,33
Nord	5%	24%	7 385 563	1,42
Seine-et-Marne	8%	39%	6 906 313	2,45
Haut-Rhin	12%	0%	5 442 597	3,57
Rhône	5%	46%	4 137 470	1,11
Eure	11%	22%	3 589 347	2,97
Moselle	7%	5%	3 482 988	1,67
Paris	20%	0%	1 968 946	0,45

Tableau 16 : Départements les plus significatifs au regard du transport de marchandises par voie navigable. Source : données SitraM 2017. Population Insee 2017/2018

Part modale = tonnage voie navigable / (tonnage voie navigable + tonnage route)

*–données ferroviaires non disponibles–. // Part interne = part du tonnage dont l'origine ET la destination sont sur le département même. // * : valeur maximale en France*

¹²⁰ Nous n'avons malheureusement pas accès aux données ferroviaires

Enfin, nous représentons les tonnages des différentes marchandises transportées par voie navigable en France (Figure 31). Les matériaux de construction sont largement en tête avec près de 35 millions de tonnes chargées et déchargées. Puis les céréales (~ 11 millions de tonnes), les produits pétroliers (~ 6 millions de tonnes) et des marchandises de nature indéterminée, il s'agit souvent des conteneurs (~ 6 millions de tonnes). Nous l'avons déjà remarqué : les matériaux de construction représentent une grande partie du métabolisme en termes de masse (Chapitre 2/IV.7), et ils ont tendance à éclipser l'importance d'autres flux dont la densité économique (rapport prix/poids) ou l'empreinte environnementale sont bien plus importantes.

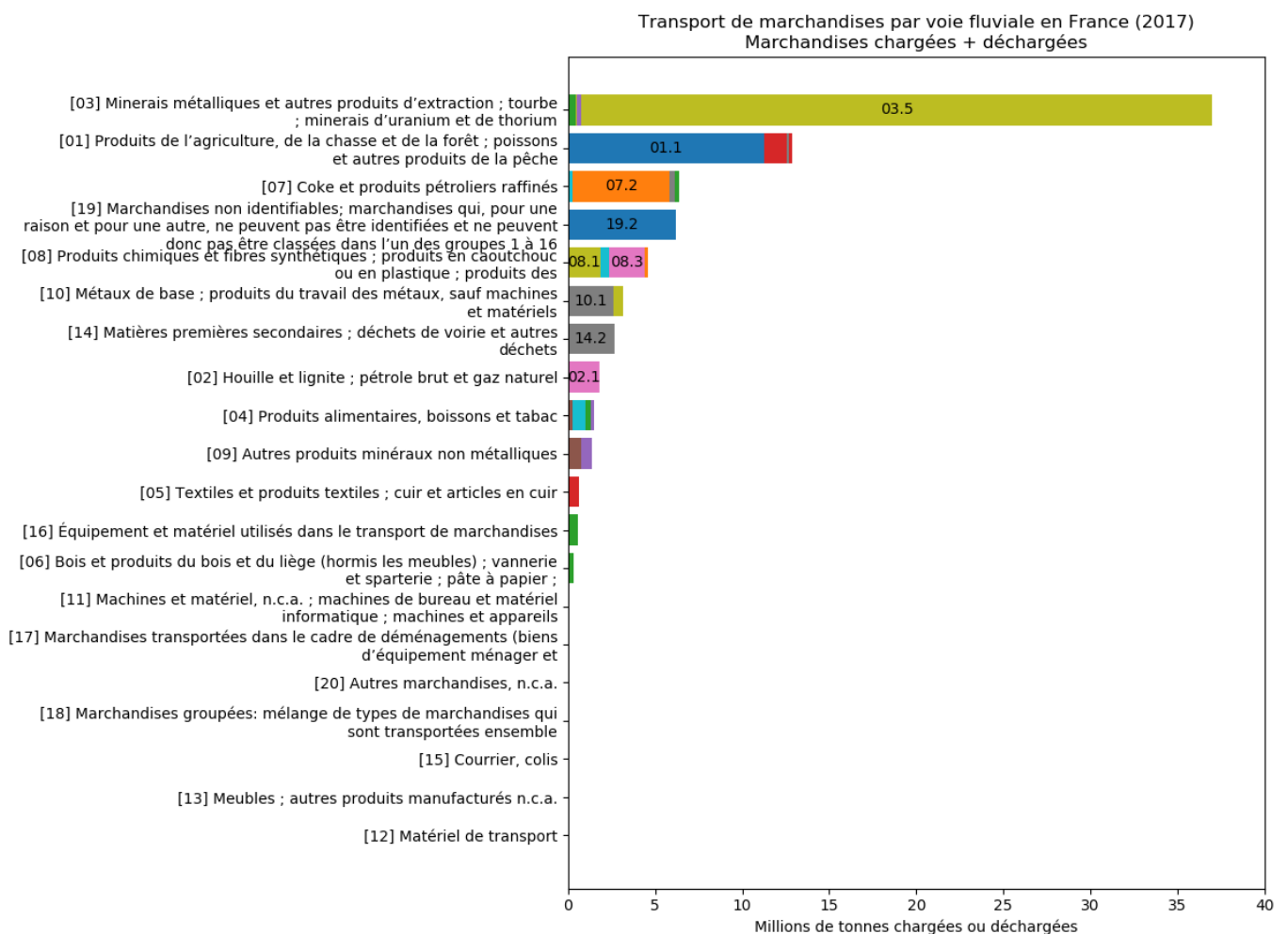


Figure 31 : Tonnage des différentes marchandises transportées par voie navigable en France. Source : SitraM 2017. Items référencés : 03.5 : Pierre, sables, graviers, argiles, tourbe et autres produits d'extraction n.c.a. // 01.1 : Céréales // 07.2 : Produits pétroliers raffinés liquides // 19.2 : Autres marchandises de nature indéterminée // 08.1 : Produits chimiques minéraux de base // 08.3 : Produits azotés et engrais (hors engrais naturels) // 10.1 : Produits sidérurgiques et produits de la transformation de l'acier (hors tubes et tuyaux) // 14.2 : Autres déchets et matières premières secondaires // 02.1 : Houille et lignite

2) À l'échelle du Bas-Rhin

Nous notons dans Tableau 16 que notre département d'étude se démarque en termes de trafic fluvial : il est presque le département où le trafic fluvial est le plus important, il l'est rapporté à son nombre d'habitants. Le trafic interne est toutefois bien plus marginal que pour d'autres départements. Nous représentons Figure 32 les quantités importées et exportées, ainsi que le transport intérieur dans le Bas-Rhin. À titre de comparaison, nous représentons également les trafics au sein du PAS, nous y reviendrons dans la section suivante.

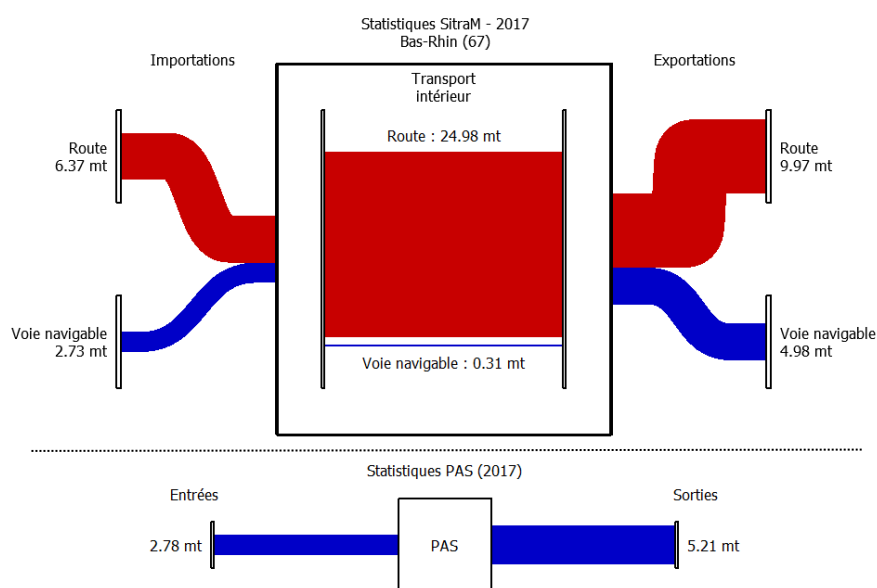


Figure 32 : Importations, exportations et trafic intérieur de marchandises dans le Bas-Rhin (par route et voie navigable), trafic de marchandises du Port Autonome de Strasbourg (voies navigables) en 2017.

Sources : SitraM, PAS

En observant l'évolution des tonnages recensés, nous notons une certaine stabilité des importations et des exportations sur les dix dernières années. Nous pouvons ainsi supposer un régime permanent dans le fonctionnement du territoire. Le zoom sur le transport par voie navigable nous permet de relever une tendance à la hausse des importations, une tendance à la baisse des exportations, et surtout une année 2018 hors normes¹²¹, nous y reviendrons dans la section suivante.

¹²¹ C'est pour cette raison que nous choisissons 2017 en année de référence sur certains graphiques, plutôt que 2018 dont nous avons également les données.

Transport de marchandises - Bas-Rhin (67)

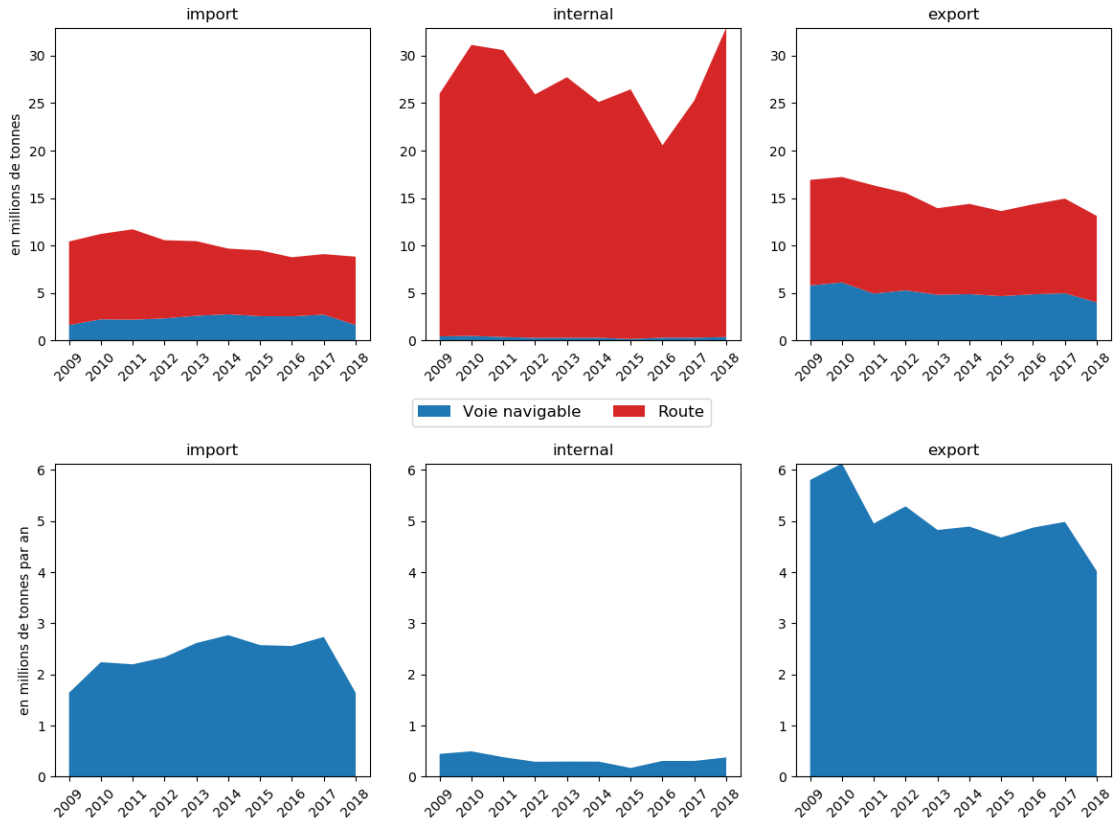


Figure 33 : Évolution des quantités de marchandises transportées dans le Bas-Rhin par route et voie navigable, entre 2009 et 2018. Source : SitraM

Nous examinons plus en détail la nature des flux et leur origine pour l'année 2017. En importation (Figure 34), les principaux flux sont des produits pétroliers en provenance d'Allemagne, des Pays-Bas et de Belgique (~ 1 000 kt/an), des céréales en provenance du Haut-Rhin et d'Allemagne (~ 400 kt/an), des marchandises de diverses natures de Belgique et des Pays-Bas (~ 300 kt/an), des sables, graviers, scories d'Allemagne (~ 200 kt/an).

Importations par voie navigable,

Bas-Rhin 2017

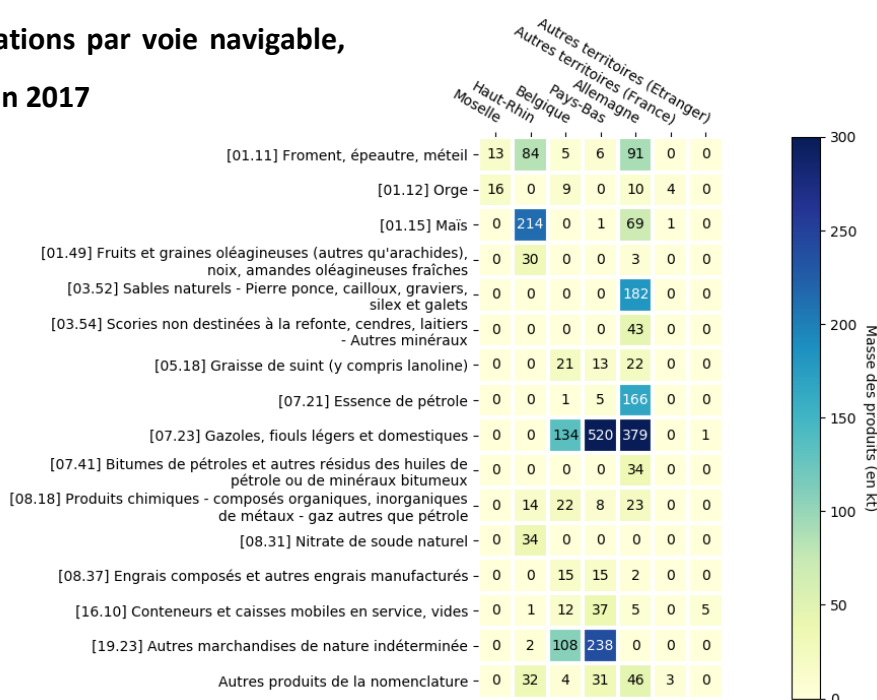


Figure 34 : Nature et origine des marchandises importées dans le Bas-Rhin par voie navigable en 2017 (total : 2732 kt)

En exportation (Figure 35), nous avons des matériaux de construction, essentiellement du sable et des graviers à destination de l'Allemagne, des Pays-Bas, de la Belgique (~ 3 000 kt/an), des marchandises de diverses natures vers la Belgique et les Pays-Bas (~ 500 kt/an), du maïs vers les Pays-Bas et l'Allemagne (~ 300 kt/an). L'origine (et la destination) des marchandises est toutefois à interpréter avec nuance : il s'agit du dernier point de transfert modal et non de l'origine réelle de la marchandise (Chapitre 3/II.3).

Exportations par voie navigable,

Bas-Rhin 2017

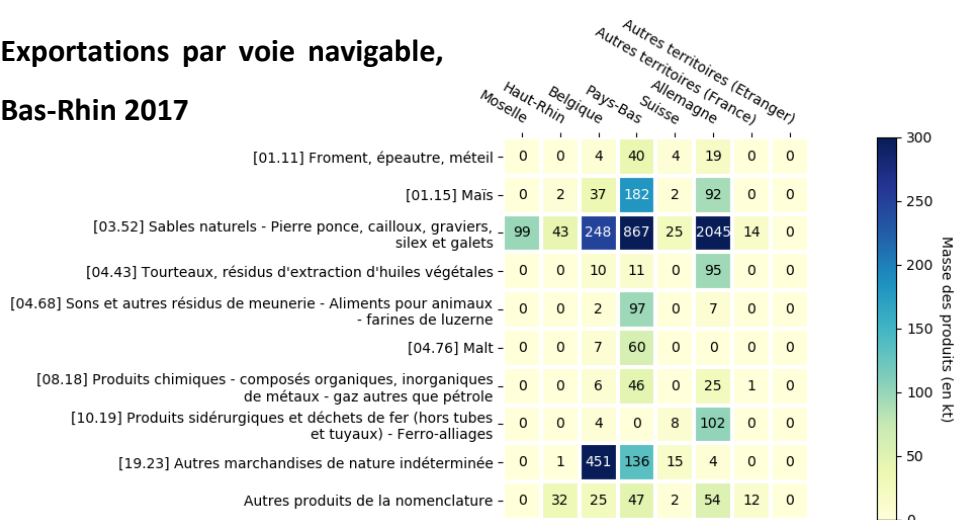


Figure 35 : Nature et destination des marchandises exportées depuis le Bas-Rhin par voie navigable en 2017 (total : 4981 kt)

Sans aller autant dans le détail que pour le trafic fluvial, nous représentons Figure 36 l'origine et la destination des marchandises importées et exportées par route. Cela nous permet notamment de remarquer qu'environ la moitié des échanges de marchandises du département l'est avec un territoire frontalier : Meurthe-et-Moselle, Moselle, Vosges, Haut-Rhin, et Allemagne¹²².

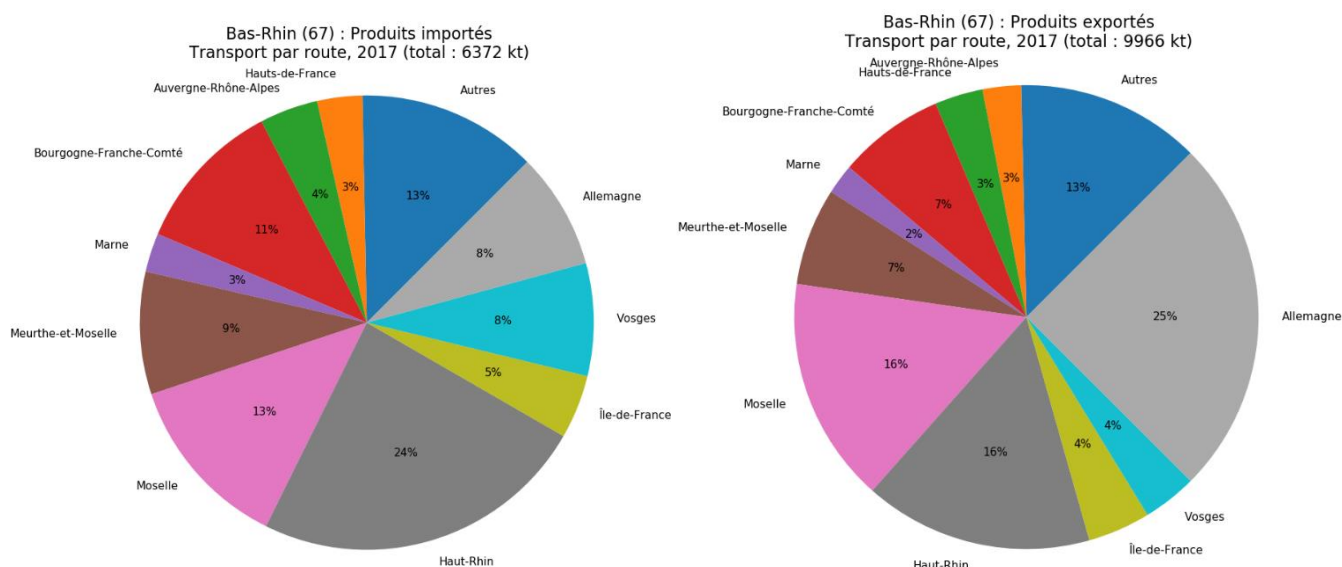


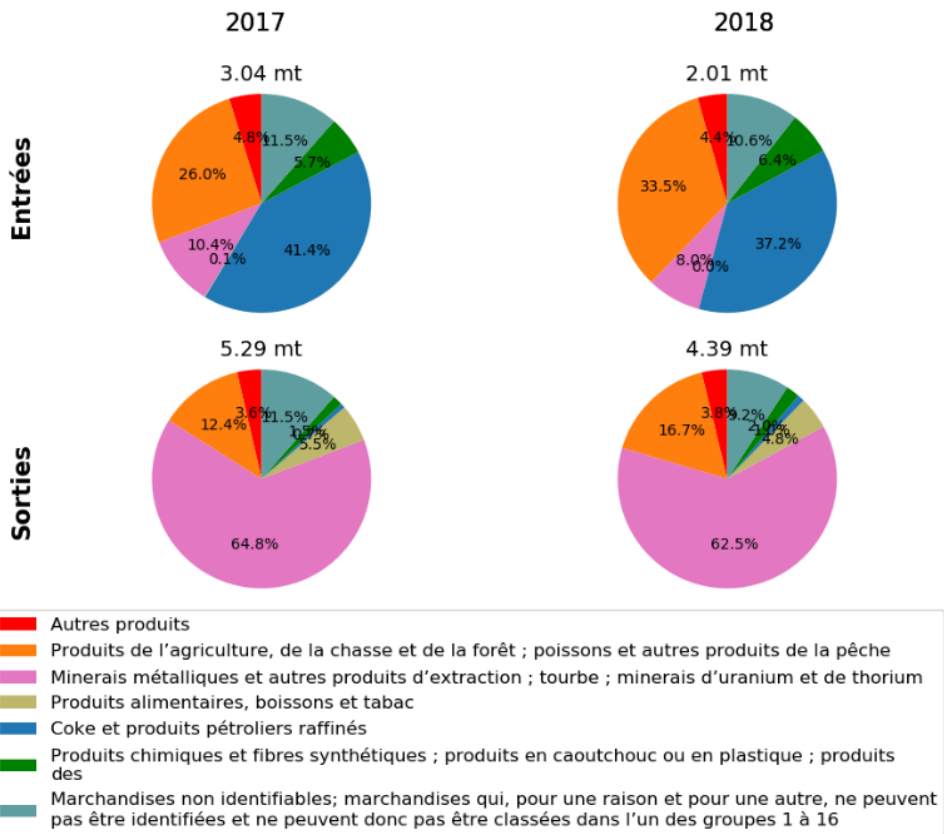
Figure 36 : Origine (destination) des marchandises importées (exportées) par route dans (depuis) le Bas-Rhin en 2017. Source : SitraM

3) Dans le Port Autonome de Strasbourg

Nous avons représenté Figure 32 p. 186 les trafics fluviaux du PAS à côté de ceux du département. D'après ces données, environ 95 % du trafic fluvial du département transite en fait par les installations du PAS. Les flux décrits par les deux jeux de données (SitraM et PAS) devraient donc coïncider fortement : c'est bien ce que nous observons Figure 37, malgré des intitulés de nomenclatures différents et des regroupements de catégories dans les statistiques du PAS.

¹²² Nous ne disposons toutefois pas de plus de précision spatiale sur les échanges avec les pays étrangers.

a) **Trafic fluvial (Statistiques SitraM pour le Bas-Rhin)**



Trafic fluvial (Statistiques du PAS)

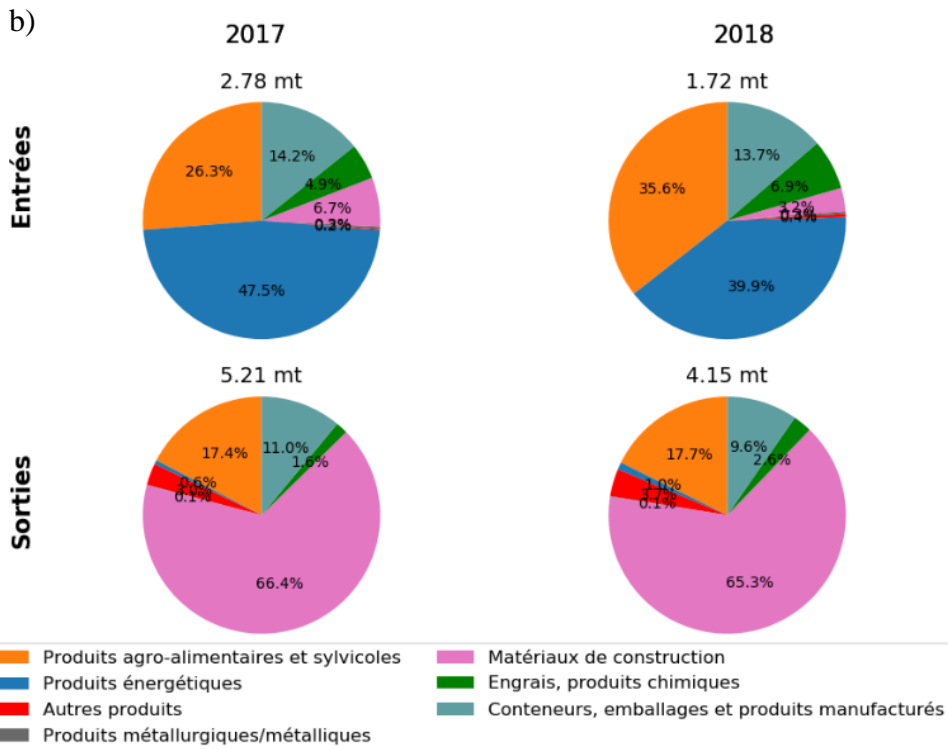


Figure 37 : Comparaison entre les statistiques de transport de marchandises par voie fluviale, a) Dans le Bas-Rhin (Source : SitraM) ; b) Dans les installations du PAS (Source : PAS)

Ainsi, les statistiques du PAS peuvent compléter les données SitraM en nous renseignant sur les dynamiques saisonnières, comme l'illustre la Figure 38. Nous observons une très forte baisse de l'activité entre août et décembre 2018, notamment des produits pétroliers en import, et des matériaux de construction en export. En fait, cette baisse d'activité reflète un épisode de basses eaux du Rhin qui a empêché la circulation de certains bateaux (CCNR 2019).

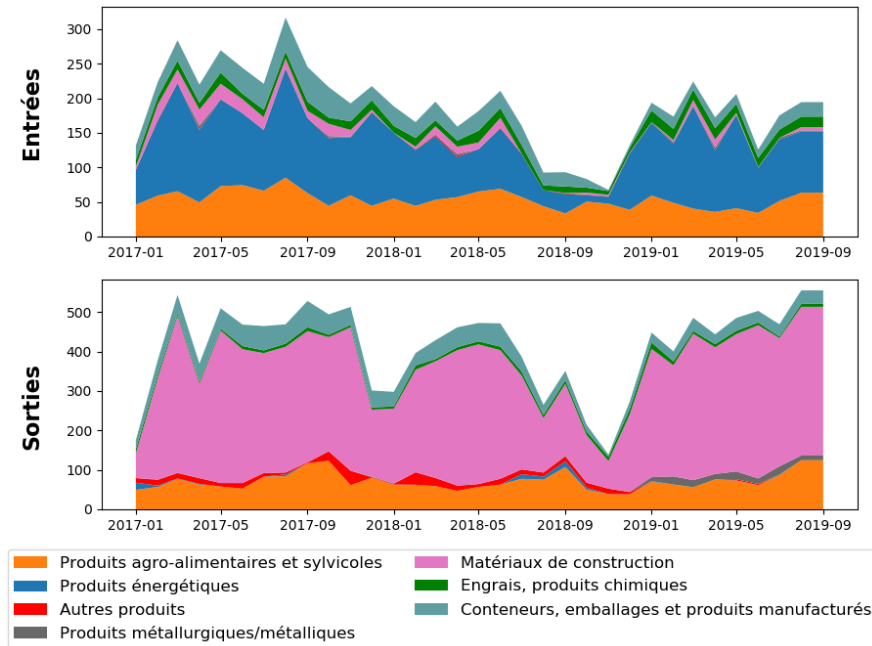


Figure 38 : Trafic fluvial de marchandises du Port Autonome de Strasbourg (en kilotonne par mois). Source : PAS

Cet épisode est encore plus flagrant à travers le trafic de conteneurs (Figure 39), qui fut presque nul par voie fluviale en octobre 2018. D'après les responsables du PAS avec qui nous avons pu échanger, cela s'est fait au bénéfice du transport par voie ferrée qui a été sollicité au maximum de ses capacités à cette période.

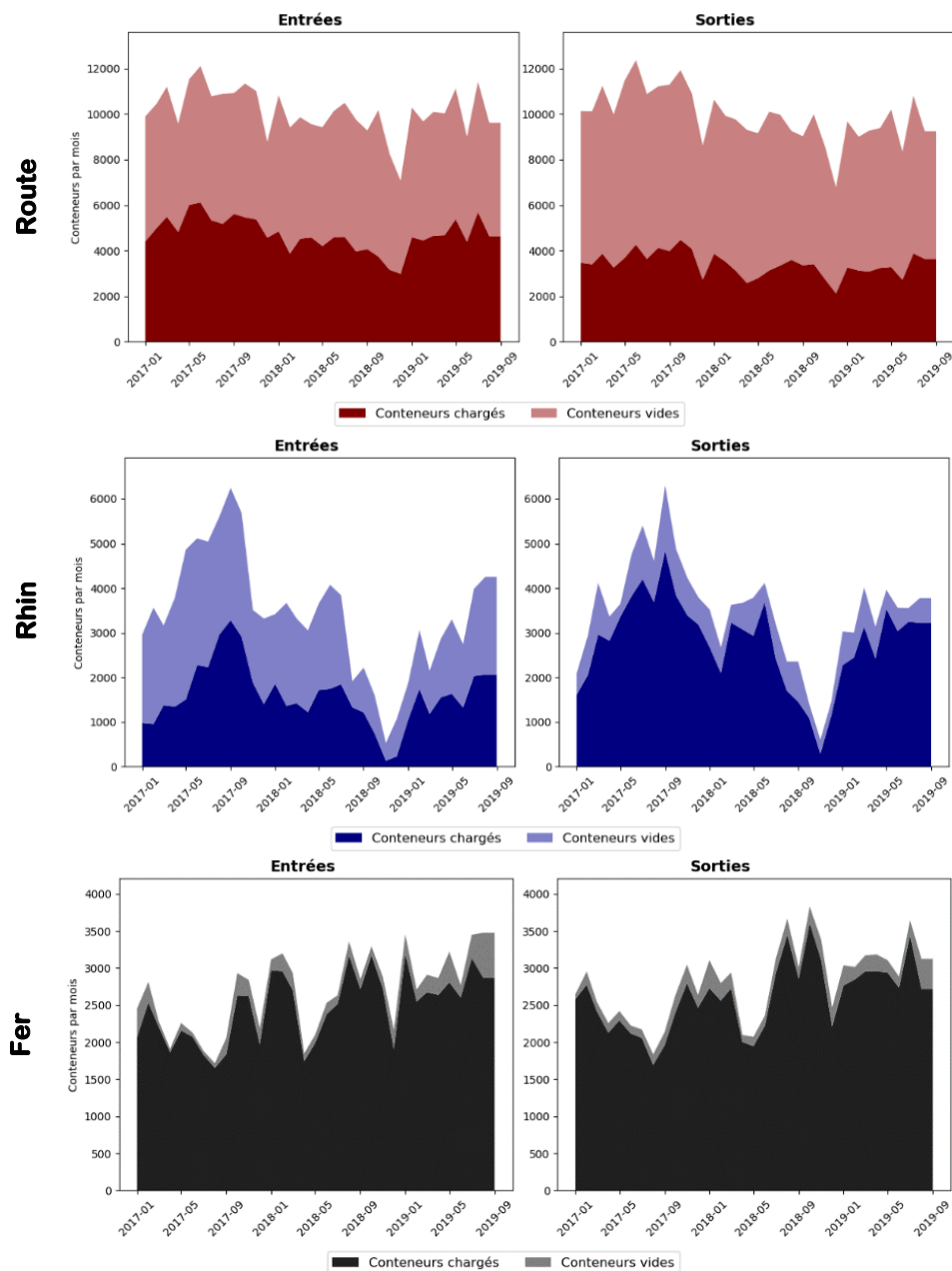


Figure 39 : Trafic de conteneurs dans les installations du PAS
selon les différents modes de transport. Source : PAS

Pour synthétiser l'activité des deux plateformes trimodales de conteneurs dans le port, nous représentons le bilan 2017 du trafic de conteneurs en Figure 40. Environ 210 000 conteneurs ont transité par ces infrastructures cette année-là, mais ce qui est frappant est la proportion de conteneurs vides qui y transite également : presque la moitié de tous les conteneurs. Cela laisse ainsi penser à une sous-optimalité logistique, car transporter des conteneurs vides présente a priori peu d'intérêt. Optimiser cela risque toutefois d'être complexe puisque cette dynamique fait intervenir plusieurs acteurs : les entreprises logistiques, le PAS et les armateurs propriétaires des conteneurs.

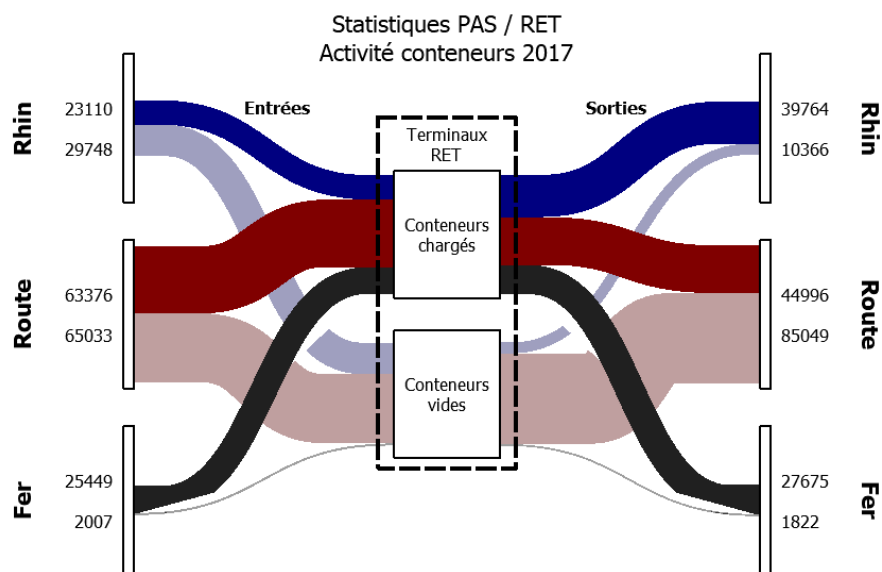


Figure 40 : Trafic de conteneurs dans le Port Autonome de Strasbourg en 2017.

RET = Rhin Europe Terminals. Source : PAS

Enfin, nous avons trouvé une seule une information¹²³ sur le trafic de marchandises par voie ferrée à travers le PAS, pour le premier semestre 2016. Sans pouvoir distinguer les entrées des sorties, cela nous donne un aperçu de ce qui est transporté. Les principales marchandises en termes de tonnage sont pour moitié des conteneurs (divers), pour un quart des céréales et pour un huitième des produits métallurgiques. En 2018, ce sont environ 1,3 million de tonnes de marchandises qui transitent par les installations ferroviaires du port.

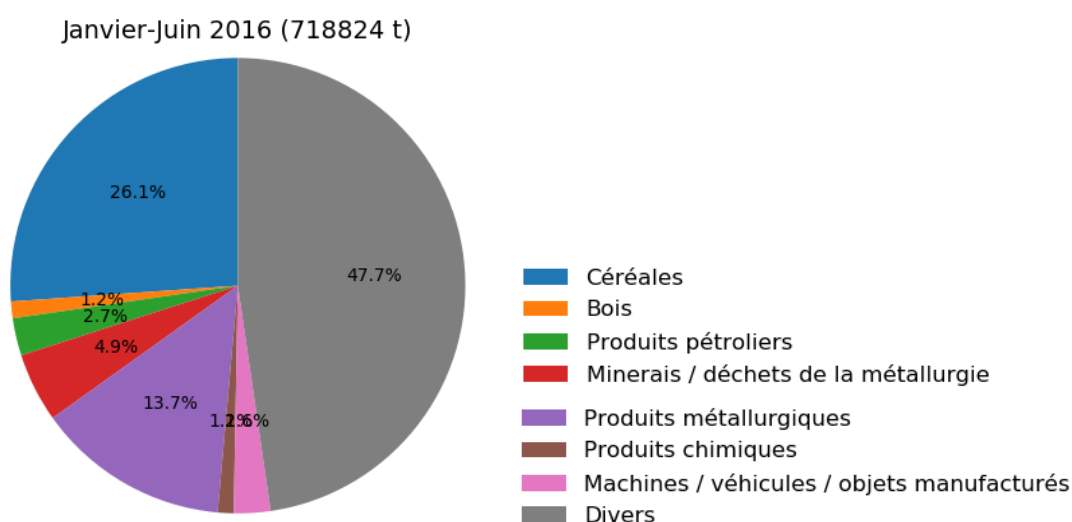


Figure 41 : Fret ferroviaire à travers le PAS entre janvier et juin 2016. Source : PAS

¹²³ Nous pensons d'ailleurs que cette information a été intégrée par erreur dans les données diffusées, puisqu'elle n'est présente que dans le seul fichier des données de juin 2017.

IV. Discussions

Au regard de ces résultats produits, et des applications identifiées (Chapitre 2/IV), nous pouvons révéler certains aspects du fonctionnement des espaces portuaires.

1) Compréhension scientifique et vision partagée

Ce travail contribue aux objectifs du projet POPSU en permettant de renforcer et partager la connaissance du territoire entre différents acteurs clefs (EMS, PAS, VNF, DREAL...). Les données viennent ainsi nous révéler les principales marchandises qui continuent de transiter par la voie d'eau sur le territoire. Pour faire simple : conteneurs¹²⁴, cailloux, céréales, produits pétroliers et chimiques.

Elles nous révèlent également les échanges par voie d'eau sont surtout avec l'Allemagne, les Pays-Bas, la Belgique et le Haut-Rhin, c'est-à-dire par le Rhin. Il existe pourtant une liaison navigable au nord de Strasbourg vers la Lorraine par le canal de la Marne au Rhin (voir Annexe 6/II), mais elle n'est utilisée que pour le tourisme fluvial et non pas pour le transport de marchandises, le gabarit Freycinet¹²⁵ de ce canal étant assez contraignant. C'est aussi certainement ce qui explique que malgré la présence d'une voie navigable dans le Sud-ouest (au même gabarit, voire inférieur à certains endroits), le trafic fluvial de marchandises y est nul ou presque (Figure 30).

2) Amélioration des procédés et de la circulation des flux

Les données utilisées ne nous amènent pas de détails sur les procédés opérant sur le territoire portuaire. Toutefois, elles nous permettent de mettre en avant des potentialités pour un transport plus doux de marchandises. Notamment, le transport de bois et papier est inexistant par cette voie navigable. Or, dans les espaces portuaires, il y a par exemple Blue Paper, une papeterie qui produit près de 400 000 tonnes de papier par an. Il est possible qu'une partie de ce papier transite dans des conteneurs, sans que nous puissions ainsi l'observer au travers des statistiques, mais cela mériterait une attention particulière pour voir si c'est vraiment le cas. Autre sujet : l'approvisionnement en bois des chaufferies biomasse de la métropole. Sur les

¹²⁴ Dont il est toutefois compliqué de connaître le contenu.

¹²⁵ Le gabarit Freycinet est une norme de 1879 pour le dimensionnement des écluses : 39 m de long x 5,20 m de large, avec 1,80/2,20 m de tirant d'eau.

espaces portuaires, une telle chaufferie existe consomme près de 5 000 t/an de bois pour alimenter en chauffage 27 000 logements. Ce bois n'étant pas disponible très localement, il est actuellement acheminé par camion depuis les Vosges à une cinquantaine de kilomètres à l'ouest (voir Annexe 6/III). Or, l'existence d'une connexion fluviale via Saverne, ou ferroviaire plus au sud, pourrait être favorable à un transfert modal de cette marchandise.

De manière générale, le transport par voie navigable connaît un renouveau. Notamment, malgré les contraintes, un projet de transport fluvial de conteneurs en gabarit Freycinet est en train de voir le jour entre Saverne et Strasbourg, sous l'impulsion de quelques groupes industriels du bassin de vie (PAS 2018).

3) Les projets de constructions et d'aménagement

a) Les matériaux du BTP

En raison de la densité, des quantités et du faible prix des matériaux mobilisés par le BTP (particulièrement sable et graviers), ce secteur reste l'un des derniers à utiliser régulièrement les voies fluviales pour leur transport. Cela se remarque particulièrement Figure 31 (page 185) et Figure 35 (page 188). Les caractéristiques géologiques de la plaine d'Alsace en font une région favorable à l'extraction de ces matières. Si leur trafic par le Rhin vers l'aval est important (~ 3 millions t/an), seules quelque 75 000 t transitent de manière interne (dans le département) par voie fluviale (~ 1% de ce qui transite sur la route), et ce malgré un réseau fluvial assez favorable en termes de couverture du territoire et de la proximité de quelques gravières à la voie d'eau.

Il semble ainsi y avoir un potentiel d'intensification du transport de ces matériaux par voie navigable, notamment pour le BTP. Plusieurs chantiers à venir dans la métropole sont d'ailleurs en proximité directe des voies d'eau (Annexe 6/IV).

b) Les axes logistiques

La logistique est une « fonction vitale » du métabolisme d'un territoire (Laubard & Lissorgues 2010) mais elle est elle-même contrainte par l'aménagement du territoire, en particulier les axes de transport de marchandises et des zones d'activité. Dans cette logistique, le PAS joue un rôle important pour le territoire. Les propos suivants sont appuyés par la carte en Annexe 6/II.

D'une part, il constitue un nœud trimodal permettant le passage de différents types de marchandises (conteneurs, graviers, céréales...) entre la route, le fer et le fleuve, ce qui est un atout majeur pour un port (Beyer & Verhaeghe 2015). En considérant la ligne de tramway qui le traverse, le port pourrait même devenir quadrimodal en accueillant un terminal de tram-fret, en appui à une logistique urbaine plus douce. Des expérimentations d'un transport plus doux de marchandises en milieu urbain sont en également en cours, notamment avec l'installation par VNF d'une plateforme fluviale de logistique urbaine au cœur de la métropole (VNF 2019).

D'autre part, de nombreuses activités industrielles et entreprises logistiques sont implantées sur les terrains du PAS, profitant parfois plus de la disponibilité des terrains que de l'intermodalité. Le transport routier reste ainsi privilégié pour ces entreprises, générant également des nuisances pour les habitants de Strasbourg, notamment autour des principaux axes routiers qui traversent la ville : l'A4/A35 et l'Avenue du Rhin (E52). Le projet de Contournement Ouest de Strasbourg (COS) est supposé venir pacifier la circulation sur ces axes, il vient cependant éloigner le port des axes de transport routier, allongeant son accès d'une douzaine de kilomètres depuis le nord de la ville.

En regardant les réseaux de transport, et la nouvelle configuration amenée par la perspective du COS, nous voyons apparaître des seuils d'entrée dans l'Eurométropole. Au nord, la zone autour de Vendenheim est également un nœud trimodal, à la croisée de nombreux axes. Au Sud, le seuil est moins flagrant. Il s'étend de l'aéroport, jusqu'au Port de Strasbourg. Bien qu'étalée, cette zone offre des interfaces routières, ferroviaires, fluviales et aérienne, elle est aussi l'interface vers l'ouest du territoire, pour des dessertes plus locales jusque dans le massif vosgien. À l'Ouest, l'ouverture de l'Eurométropole vers l'Allemagne se concrétise autour de deux seuils : à Kehl et au sud du PAS.

En suivant les axes de transport, ce sont même des seuils d'entrée du bassin de vie qui émergent : à Saverne, située à proximité des principaux axes routier, ferroviaire et fluvial pour l'accès par le nord-ouest, et à Sélestat pour l'accès par le Sud, voir Colmar pour une connexion fluviale. Lautebourg, au nord-est, dispose également d'une position intéressante et d'un accès trimodal. Ces différents seuils nous semblent constituer des positions logistiques stratégiques : ils permettraient de connecter l'essentiel du territoire de manière cohérente et avec des modes de transport doux.

c) Les réseaux énergétiques

Le Port constitue également un nœud important du réseau énergétique. D'une part, les activités industrielles sont fortement consommatrices d'énergie. D'autre part, la proximité avec un grand fleuve y a permis l'implantation d'un barrage hydroélectrique de 150 MW, d'autres barrages étant situés tout le long du Rhin¹²⁶. Enfin, une chaufferie à biomasse est également présente sur le port et vient alimenter le réseau de chaleur urbain. Les servitudes de réseaux (électriques, gaz, chaleurs, routes...) constituent ainsi des éléments contraignants mais nécessaires au fonctionnement des espaces industrialo-portuaires. De fait, le port et certaines de ses entreprises sont des acteurs clés du réseau et de la transition énergétique, malgré une coordination avec les pouvoirs publics encore insuffisante et qui mériterait d'être mieux formalisée (Beyer & Lacoste 2017).

4) Des questions de résilience et d'anticipation de l'épuisement des ressources

Le port a été affecté par une période d'étiage pendant l'été 2018, cela se voit notamment à la réduction des volumes de marchandises transitant par le port (Figure 38 et Figure 39 pages 191-192). Ce phénomène a concerné une portion du Rhin en Allemagne (en aval), limitant la capacité des bateaux pouvant circuler à cet endroit. Au niveau du port, les processus les plus affectés ont été le transport de conteneurs, l'approvisionnement en produits pétroliers et les exportations de matériaux de construction.

Cet épisode vient donc interroger la résilience du port face à ces phénomènes d'étiage, bien que les représentants du PAS soulignent le caractère exceptionnel de cet épisode. Les changements climatiques à venir pourraient toutefois venir relativiser ce caractère exceptionnel. Dans une optique de renforcement de la résilience, deux approches nous paraissent intéressantes :

- Renforcer le réseau ferroviaire en parallèle du Rhin, surtout sur le tronçon bas-rhinois reliant Strasbourg à Lauterbourg puis à l'Allemagne, pour permettre bonne complémentarité entre la voie fluviale et la voie ferroviaire. Cela faciliterait le transfert de charge de l'une à l'autre en cas de besoin et peut également améliorer l'efficacité

¹²⁶ https://www.edf.fr/sites/default/files/Hydraulique/Alsace-Vosges/documents/les_amenagements_hydro_electriques_du_rhin_franco-allemand.pdf

économique et logistique pour le transport de certaines marchandises (Beyer & Verhaeghe 2015).

- Développer le rôle local du PAS, pour l'instant très limité (nous le voyons au trafic intérieur fluvial Figure 33 p. 187), pour réduire son exposition aux risques extérieurs.

5) Autres applications non abordées

Certaines des applications présentées dans le Chapitre 2 n'ont pas pu être abordées avec les quelques jeux de données que nous avons exploités.

Tout d'abord, l'évaluation d'impacts environnementaux n'a pas pu être menée dans la temporalité de notre travail. Dans des travaux futurs, nous envisageons d'exploiter la base IREP (Registre des émissions polluantes) qui rassemble des informations d'émission et de production de déchets pour les sites industriels les plus importants. 22 des 270 entreprises de la zone d'activité du PAS à Strasbourg sont identifiées dans ce registre.

Également impactant l'environnement étant donné leur dispersion, les déchets sauvages sont aussi une problématique du port. Ce sont environ 165 tonnes qui sont collectées chaque année¹²⁷, mais l'absence de données structurées à ce sujet ne permet pas de construire une information plus élaborée.

Concernant le suivi de la réglementation applicable, nous n'avons pas de données permettant d'aller examiner le respect d'un quelconque texte de loi. Toutefois, certains textes de loi, en particulier des arrêtés préfectoraux, pourraient nous aider à rassembler des informations sur les flux de certaines installations, à l'image des 400 000 tonnes annuelles de papier produite par Blue Paper dont nous retrouvons la mention dans les textes réglementaires.

Enfin la démarche d'écologie industrielle entre les entreprises du PAS mérite un intérêt particulier, ne serait-ce parce qu'elle est unique dans le département (Beyer & Lacoste 2017). Cette démarche n'est toutefois pas productrice de données structurées que nous pourrions analyser à travers le SINAMET, les entreprises étant particulièrement prudentes quant à la transmission de leurs données sur les flux entrants et sortants, car elles reflètent leur niveau

¹²⁷ <https://www.strasbourg.port.fr/port-responsable/responsabilite-societale/presentation/>

d'activité. Les quelques données disponibles sont plutôt des éléments de communication¹²⁸, ou la liste des synergies entre les entreprises, mais qui ne sont pas quantifiées.

V. Conclusion du sixième chapitre

Nous avons étudié les espaces fluviaux et portuaires dans le bassin de vie de Strasbourg à travers les flux de marchandises qui y transitent. Dans un premier temps, il s'agissait d'analyser des données statistiques avec l'aide du SINAMET. Dans un second temps, nous avons utilisé la liste d'applications des analyses de flux pour construire et structurer une meilleure connaissance du territoire et de sa dynamique matérielle. Cette seconde partie s'appuyait sur les résultats obtenus grâce au SINAMET, mais aussi sur d'autres éléments de compréhension notamment des entretiens avec les responsables du PAS et un travail cartographique autour des espaces fluviaux et portuaires du territoire.

Nous discuterons des apports et limites de notre approche méthodologique dans la conclusion générale de cette thèse. En ce qui concerne ce cas d'étude plus spécifiquement, nous notons la place importante des installations fluviales pour le métabolisme global du territoire. Sous certaines conditions (coordination des acteurs, constructions des infrastructures idoines...), ces installations pourraient jouer un rôle significatif pour la transition écologique. Nous avons d'ailleurs identifié quelques potentiels de développement qui pourraient participer à modifier le fonctionnement métabolique du territoire pour qu'il soit moins impactant sur l'environnement.

Toutefois, ce développement territorial est aussi le résultat d'un jeu d'acteurs entre les gestionnaires d'infrastructures (PAS, VNF...), les collectivités (Ville et Eurométropole de Strasbourg, Bas-Rhin, Grand Est...), l'État et les acteurs industriels (Blue Paper, Électricité de Strasbourg...) pour ne citer que les principaux. L'élaboration d'une stratégie commune et cohérente pour une véritable écologie territoriale est ainsi un processus complexe. Cela risque donc de prendre encore du temps même si certaines actions qui sont menées actuellement

¹²⁸ <https://www.cles-ports-de-strasbourg.eu/>

(écologie industrielle, péniches au gabarit Freycinet, plateforme d'observation multi-acteurs...)
révèlent une évolution déjà en cours.

Chapitre 7 Sensibilité à l'échelle d'indicateurs de la comptabilité des flux

Ce chapitre a été soumis comme article scientifique en juin 2020 dans la revue Journal of Industrial Ecology sous le titre « Scale Sensitivity of Import and Export Indicators in the Economy-Wide Material Flow Accounts Methodology » par B. Ribon. Nous le présentons ici dans sa version française et légèrement adaptée au format de cette thèse.

I. Introduction

La comptabilité des flux de matières pour une économie entière (Economy-Wide Material Flow Accounts – EWMFA) proposée par Eurostat (2001), puis adaptée à des régions (Barles 2009 ; CGDD 2014) vise à étudier l'usage des matériaux du système socio-économique sur un territoire à travers différents indicateurs. Nous avons recensé plus d'une quinzaine dans la littérature (Bringezu et al. 2003 ; Barles 2014) et nous en présentons une partie Tableau 17 page suivante.

Ces indicateurs sont utilisés dans certaines études pour comparer différents territoires (par exemple Barles 2014). Bien sûr, ces indicateurs, qui reposent sur l'agrégation de masses de flux de matières, peuvent être questionnés sur leur signification, les impacts environnementaux ne sont pas « juste une question de poids » (Van der Voet et al. 2004). La présence de plateforme logistique comme des ports peut aussi augmenter significativement les entrées et sorties de matières d'un territoire sans que cela ne signifie des besoins accrus des habitants du territoire. Pour finir, la qualité des données, le périmètre et les hypothèses diffèrent souvent d'une étude à l'autre. Tous ces éléments rendent la comparaison de ces indicateurs territoriaux complexe.

Notre article se concentre toutefois sur une autre problématique. Afin de compenser la différence de taille entre les territoires, ces indicateurs sont souvent exprimés sous forme de ratios « par habitant ». Toutefois, certains indicateurs, comme la masse des importations ou des exportations, sont influencés par la taille du territoire étudié, même en utilisant ces ratios par

habitant. Les plus grands territoires importent ou exportent mathématiquement moins *par habitant* que les plus petits. Ce phénomène est une manifestation du problème d'agrégation spatiale (Modifiable Areal Unit Problem – MAUP) formalisé par Openshaw & Taylor (1979). Ils illustrent comment la manière de délimiter et grouper les territoires peut avoir une influence sur les données statistiques. Dans notre étude, nous allons démontrer que la masse totale des importations d'un groupe de territoires est inférieure à la somme des masses des importations de chacun de ces territoires. Cela s'explique par les flux échangés entre les territoires du groupe qui ne sont plus considérés lorsqu'il s'agit d'étudier les importations du groupe.

Indicateur	Direction	Formule	Sensible* / Non additif	Insensible* / Additif
(I) Importations	Entrée		X	
(E) Exportations	Sortie		X	
(DE) Extraction intérieure (utilisée)	Entrée			X
(DEU) Extraction intérieure inutilisée	E & S			X
(DPO) Émissions vers la nature	Sortie			X
(DMI) Entrée directe de matière	Entrée	DE + I	X	
(TMI) Entrée totale de matière	Entrée	DMI + DEU	X	
(TMO) Sortie totale de matière	Sortie	DEU + DMO	X	
(DMO) Sortie directe de matière	Sortie	DPO + E	X	
(TDO) Émissions totales vers la nature	Sortie	DPO + DEU		X
(DMC) Consommation intérieure apparente de matières	Mix	DMI - E		X
(PTB) Balance commerciale physique	Mix	E - I		X
(NAS) Addition nette au stock	Mix	Entrées – sorties		X

Tableau 17 : Principaux indicateurs de la comptabilité des flux de matières pour une économie entière (EWMFA), d'après Eurostat 2001, OECD 2008 p.84. Les indicateurs basés sur des équivalents de matière brute ne sont pas mentionnés. * : Indicateur sensible ou insensible à l'échelle.

Nous n'avons pas trouvé de recherche antérieure qui lie directement le MAUP aux méthodes de comptabilité matière et à leurs indicateurs et en approfondissant cette question. Le problème est parfois mentionné comme la « non-additivité sur les territoires » (OECD 2008) des indicateurs d'importations, d'exportations et de leurs dérivés (Tableau 17). En effet, il n'est pas possible de simplement additionner ces indicateurs pour obtenir celui d'un groupe. Nous montrons ici qu'il ne s'agit pas seulement d'un problème d'addition, mais aussi de division lorsque ces indicateurs sont exprimés sous la forme d'un ratio, notamment en étant rapportés au nombre d'habitants. Aussi, nous préférons parler de « sensibilité à l'échelle » plutôt que de « non-additivité ».

Pour appuyer notre démonstration, nous utiliserons deux moyens : un cas d'étude réel et un modèle mathématique. Dans un second temps, nous discuterons de cette sensibilité à l'échelle et nous proposerons des approches permettant d'aborder le problème différemment. Afin de gagner en lisibilité, nous ne mentionnerons que le cas des importations, mais le même raisonnement peut être appliqué aux exportations.

II. Cas réel : les importations selon la structure administrative française

Pour illustrer nos propos avec un exemple réel, nous nous appuyons sur le « mille-feuille » administratif français. Entre 1970 et 2015, la France¹²⁹ était subdivisée en 22 régions, chacune divisée en deux à cinq départements, pour un total de 96 départements. En 2016, la répartition des régions a changé et il y a maintenant 13 nouvelles régions qui rassemblent et remplacent une, deux ou trois anciennes régions, le nombre de départements restant le même.

Il n'est pas possible ici de montrer toutes les données pour les 132 territoires : 1 pays, 13 nouvelles régions, 22 anciennes régions, 96 départements. Aussi, nous allons présenter les données uniquement pour la région Grand Est. Cette nouvelle région rassemble trois anciennes régions (Alsace, Champagne-Ardenne, Lorraine) et dix départements (voir Tableau 18 p. 205, colonne 0).

Dans un premier temps du processus de traitement des données, nous avons calculé, à l'aide du SINAMET (Chapitre 4), les masses des matériaux importés pour les 132 territoires

¹²⁹ Il s'agit ici uniquement de la France métropolitaine

français grâce aux données SitraM (Tableau 13 p. 180). Il s'agit des données de transport de marchandises par camion et voie navigable entre les départements français, et à l'étranger en 2010, les données relatives aux autres moyens de transport n'étant pas utilisables. C'est évidemment un biais dans l'étude mais : 1) Cela a peu d'influence sur ce que nous voulons démontrer, c'est-à-dire la sensibilité à l'échelle des importations, 2) En 2010, près de 84,6 % du transport de marchandises s'est fait par la route et 2,2 % par voie navigable (MEDDTL 2012). Ainsi, les données analysées et calculées sont assez bien représentatives.

La masse des importations d'un territoire est calculée en additionnant tous les flux dont l'origine n'est pas incluse dans le territoire, mais la destination l'est, conformément à la méthodologie du CGDD (2014). Cela permet ensuite de calculer le ratio *par habitant*. Nous présentons les valeurs obtenues pour le Grand Est dans les colonnes 2 et 3 du Tableau 18. Nous montrons également la distribution de tous les ratios d'importations pour différentes échelles administratives en France (Figure 42).

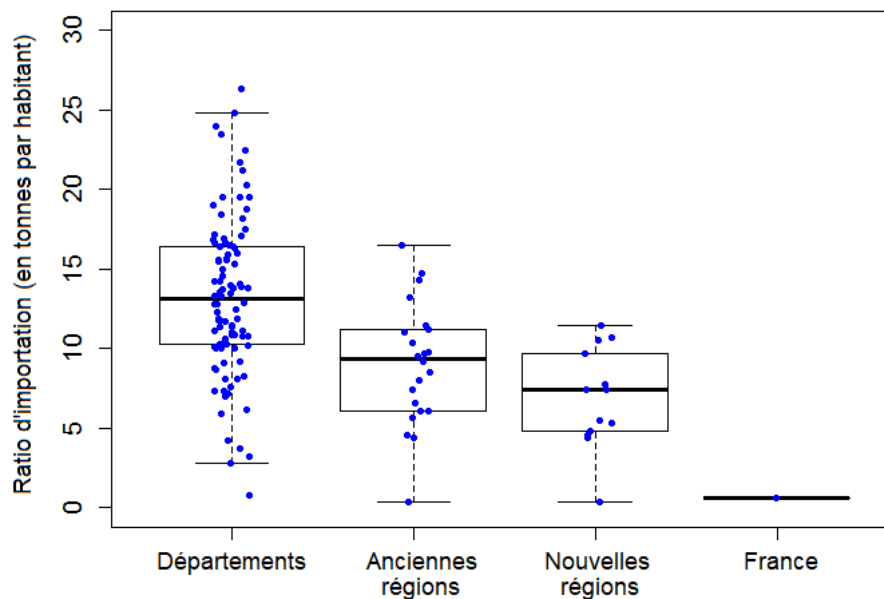


Figure 42 : Ratios d'importation (en tonnes par habitant) de différents territoires administratifs en France. Marchandises transportées par camion et voie fluviale uniquement.

Sources : SitraM (2010), population Insee (2010)

Sur le tableau et sur la figure, nous observons que le ratio d'importation des grands territoires est globalement plus faible que celui des petits territoires. Dans la section suivante, nous présentons un modèle mathématique qui va nous permettre de prouver et de quantifier cette sensibilité à l'échelle dans un monde simplifié.

Structure administrative		Importations		Dissociation de l'origine des importations		Inclusion des flux internes		
(0)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	P	M	m			M_I	$M + M_I$	m_{ind}
Nom du territoire	Population	Masse des importations (mt)	Ration d'importation (t/h)	Ration des importations depuis le GE (t/h) (% des importations totales)	Ratio des importations hors du GE (t/h)	Masse des flux internes (mt)	Flux internes + Importations (mt)	Importation individuelle (t)
Grand Est (GE)	5 532 530	43,1	7,8	-	7,8	193,6	236,7	42,8
Lorraine	2 350 920	26,3	11,2	4,3 (38%)	6,9	63,0	89,3	38,0
Meurthe-et-Moselle (54)	732 207	11,4	15,6	8,7 (56%)	6,8	14,7	26,1	35,6
Meuse (55)	193 923	4,4	22,5	16,8 (75%)	5,7	3,8	8,1	42,0
Moselle (57)	1 045 066	17,1	16,4	9,0 (55%)	7,3	23,1	40,2	38,5
Vosges (88)	379 724	5,8	15,3	8,7 (57%)	6,6	9,0	14,9	39,1
Alsace	1 845 687	16,9	9,2	2,5 (28%)	6,6	53,9	70,8	38,4
Bas-Rhin (67)	1 095 905	11,2	10,2	4,8 (47%)	5,4	31,1	42,3	38,6
Haut-Rhin (68)	749 782	10,2	13,6	5,2 (38%)	8,4	18,3	28,4	37,9
Champagne-Ardenne	1 335 923	19,1	14,3	3,4 (24%)	10,9	57,6	76,6	57,4
Ardennes (08)	283 250	3,3	11,7	5,2 (44%)	6,5	6,8	10,1	35,5
Aube (10)	303 327	5,8	19,0	7,4 (39%)	11,7	16,3	22,1	72,8
Marne (51)	565 307	13,3	23,5	11,0 (47%)	12,4	22,0	35,2	62,3
Haute-Marne (52)	184 039	4,6	24,8	13,2 (53%)	11,6	4,7	9,3	50,3

Tableau 18 : Structure administrative (2015) et importations par camion et voie navigable (2010) pour la région Grand Est et ses subdivisions.

Sources : Population Insee (2010), Importations SitraM (2010). mt : millions de tonnes ; t/h : tonne par habitant

III. Modèle mathématique

Le développement d'un modèle mathématique de la sensibilité à l'échelle des importations vise à quantifier théoriquement le phénomène, et ainsi le prouver, et mettre en avant les facteurs qui ont une influence. Afin de garder le modèle simple, nous formulons quelques hypothèses. Nous terminerons en comparant les résultats du modèle avec ceux du cas d'étude réel.

1) Description du modèle

Dans la suite : $n \in \mathbb{N}^*$, $\mathbf{I} = \{1, 2, \dots, n\}$ et $i \in \mathbf{I}$ et $j \in \mathbf{I}$ tel que $i \neq j$.

Nous considérons le monde Ω comme une surface. Nous partitionnons Ω en un ensemble de n surfaces $T_1, T_2, T_3, \dots, T_n$ appelées « territoires élémentaires ». Par définition : $\Omega = \bigcup_{k=1}^n T_k$ et $T_i \cap T_j = \emptyset$.

R désigne une « région », c'est-à-dire une surface constituée par un ou plusieurs territoires élémentaires. \mathbf{E} réfère à l'ensemble des régions possibles. Par définition $T_i \in \mathbf{E}$. Une définition plus formelle est : $R \in \mathbf{E} \Leftrightarrow$ il existe $a, b, c, \dots, \in \mathbf{I}$ distincts tel que $R = \bigcup_{k=a,b,c,\dots} T_k$. Par l'usage d'un indice numérique R_x nous signifions que la région rassemble les x -premiers territoires, c'est-à-dire $R_x = \bigcup_{k=1}^x T_k$ ($x \in \mathbf{I}$).

$P(R): \mathbf{E} \rightarrow \mathbb{N}$ est une application qui retourne le nombre d'habitants d'une région R .

Les territoires échangent de la matière sous forme de biens qui peuvent être quantifiés par leur masse. Si nous définissons une période (par exemple un an) et une unité de masse, nous pouvons définir les applications suivantes :

- $f(R_A, R_B): \mathbf{E} \times \mathbf{E} \rightarrow \mathbb{R}$ retourne la masse totale des matières échangées de R_A vers R_B .
- $N(R): \mathbf{E} \rightarrow \mathbb{N}$ retourne le nombre de territoires élémentaires constituant R , par exemple : $N(T_i) = 1$; $N(R_x) = x$; $N(\Omega) = n$.
- $M(R): \mathbf{E} \rightarrow \mathbb{R}$ retourne la masse de tous les matériaux importés par R : $M(R) = f(\bar{R}, R)$.
- $m(R): \mathbf{E} \rightarrow \mathbb{R}$ retourne le ratio d'importation de R , c'est-à-dire : $m(R) = M(R)/P(R)$.

2) Définitions de l'additivité et de la sensibilité à l'échelle

Nous devons encore définir mathématiquement les notions d'additivité sur les territoires ou de sensibilité à l'échelle. Soit Q une caractéristique territoriale numérique, c'est-à-dire une application $Q: \mathbf{E} \rightarrow \mathbb{R}$.

- Q est **directement additif** sur les territoires si $Q(\cup_i T_i) = \sum_i Q(T_i)$. Par exemple, les applications qui retournent la population, le produit intérieur brut (PIB) ou la surface d'un territoire sont directement additives.
- Q est **indirectement additif** sur les territoires s'il existe une application directement additive D tel que : $D(\cup_i T_i) \cdot Q(\cup_i T_i) = \sum_i D(T_i) \cdot Q(T_i)$. Différents ratios tels que la densité d'habitants ou le PIB par habitant sont indirectement additifs.
- Q est **non additif sur les territoires ou sensible à l'échelle** s'il n'est pas additif (directement ou indirectement). Dans ce cas, nous avons $Q(\cup_i T_i) < \sum_i Q(T_i)$ et $Q'(\cup_i T_i) \cdot Q(\cup_i T_i) < \sum_i Q'(T_i) \cdot Q(T_i)$ pour toute application Q' qui soit une caractéristique territoriale numérique. Les importations, les exportations et la plupart des indicateurs dérivés sont sensibles à l'échelle.

3) Hypothèses

Afin de simplifier fortement la démonstration, nous formulons quatre hypothèses :

- (H1) La masse des matériaux importés est proportionnelle à la population de chaque territoire élémentaire par un facteur α . Cela signifie que $M(T_i) = \alpha \cdot P(T_i)$.
- (H2) Les territoires ont une densité homogène et le même nombre $h \in \mathbb{N}$ d'habitants. Ainsi $P(T_i) = h$.
- (H3) Les flux sont répartis avec une parfaite homogénéité entre les territoires : chaque territoire importe la même masse de matière de chacun des $n - 1$ autres territoires. Il n'y a pas de facteur de préférence comme la proximité. Ainsi $f(T_i, T_j) = \frac{\alpha h}{n-1}$.
- (H4) Le monde est un espace fermé. Il n'y a pas d'échange de matière avec l'extérieur. Il n'y a pas non plus d'extraction de matière, ni de rejet dans l'environnement, ni de capacité de stockage. La loi de conservation de la masse est vérifiée, c'est-à-dire que la masse des entrées égale la masse des sorties pour tous les territoires. Il s'agit en quelques sortes d'une économie circulaire idéale.

4) Preuve de la sensibilité à l'échelle

Afin de quantifier la sensibilité à l'échelle sur le ratio d'importation, nous pouvons le calculer pour un territoire élémentaire T_i et pour un groupe de territoires R_x .

Ratio d'importation du territoire élémentaire (selon l'hypothèse H1) :

$$m(T_i) = \frac{M(T_i)}{P(T_i)} = \alpha$$

Ratio d'importation d'un groupe de territoires :

$$m(R_x) = \frac{M(R_x)}{P(R_x)}.$$

La population est directement additive, ainsi $P(R_x) = \sum_{i=1}^x P(T_i) = xh$. Pour calculer les importations, nous additionnons la masse de tous les flux venant de l'extérieur de R_x (c'est-à-dire tous les T_i tels que $i > x$), vers ceux à l'intérieur ($i \leq x$). Cela peut être mis en équation :

$$M(R_x) = \sum_{j=1}^x \sum_{i=x+1}^n f(T_i, T_j) = xh \frac{n-x}{n-1} \alpha$$

Finalement, nous obtenons $m(R_x) = \frac{n-x}{n-1} \cdot m(T_i) < m(T_i)$ pour $x > 1$. Cela prouve que, dans un monde homogène, le ratio d'importation d'un groupe de territoires est inférieur à celui des territoires de ce groupe.

Pour illustrer cet effet avec un exemple numérique, nous posons¹³⁰ :

$$n = 96 ; x = 10 \text{ et } m(T_i) = \alpha = 20 \text{ tonnes par habitant par an (t/h.an).}$$

Ainsi, nous obtenons $m(R_{10}) \approx 18,1$ t/h.an, soit 9,5 % de moins que $m(T_i)$. Les 1,9 t/h.an qui manquent correspondent aux échanges entre les territoires dans R_{10} .

¹³⁰ Ces nombres sont inspirés de la France qui est divisée en 96 départements, la région Grand Est en réunissant 10 d'entre eux.

5) Discussion du modèle

Ces quatre hypothèses nous éloignent de la réalité, mais sont nécessaires pour simplifier le développement mathématique. Si nous supprimons l'une d'elles, les calculs deviendraient significativement plus complexes. Cet aspect pourrait être exploré dans de prochaines recherches.

Néanmoins, ce modèle nous permet de remarquer une importante différence entre le modèle mathématique et les données réelles. Dans le premier cas, le ratio d'importation du groupe de dix territoires est 9,5 % moins que pour un territoire élémentaire. Dans le second cas, ce ratio d'importation diminue d'au moins 24 % lorsque le système spatial de référence passe d'un département à la région Grand Est (Tableau 18, colonne 3). Il nous semble que l'hypothèse H3 qui stipule d'une répartition homogène des flux est la plus restrictive et peut expliquer l'essentiel de la différence. En réalité, nous observons une préférence significative pour des matériaux ayant leur origine dans les territoires proches. En région Grand Est, entre 38 % et 75 % des importations des départements viennent de la même région (Tableau 18, colonne 4). Cela a pour conséquence une importante différence entre les ratios d'importation des départements et celui de la région. Cela constitue également un bon rappel de la première loi de la géographie : « Everything is related to everything else, but near things are more related than distant things »¹³¹ (Tobler 1970),

IV. Discussion et autres approches

La sensibilité à l'échelle peut être perçue comme contre-intuitive. Il est courant de penser en première lecture que le ratio d'importation d'une région est la moyenne pondérée de ceux des territoires qui la composent, ce qui n'est pas le cas. Pour cette raison, nous recommandons de ne pas utiliser de tels ratios : ils ne sont pas adaptés pour comparer des territoires de différentes tailles. Leur valeur peut en effet être réduite simplement en élargissant le système spatial de référence. Cela peut aussi mener à la construction d'affirmations en apparences absurdes, mais numériquement vérifiées, comme « un habitant français importe moins qu'un habitant d'un département français ». Près de la moitié des indicateurs dans la méthode EWMFA sont sensibles à l'échelle. Pour construire des indicateurs additifs, comme la

¹³¹ « Tout est lié à tout le reste, mais les choses proches sont plus liées que les choses distantes. »

« Balance commerciale physique » ou l'« Addition nette au stock », il est souvent nécessaire de soustraire les exportations des importations ou inversement (voir Tableau 17 p. 202).

Cela ne veut pas dire que les indicateurs sensibles à l'échelle sont inutiles. Ils restent valides comme indicateurs historiques, permettant l'observation de tendances pour un territoire donné. Ils fournissent également un ordre de grandeur et une vue globale des besoins territoriaux (Kleijn 2000). Aussi, ces indicateurs sont un bon point de départ d'une analyse approfondie sur les dynamiques du métabolisme d'un territoire.

Dans certaines circonstances, ou à petite échelle, la sensibilité à l'échelle peut ne pas avoir d'effet, même pour les importations ou exportations. Cela concerne les territoires qui ont peu de moyens de production. C'est le cas pour la majorité des villes pour lesquelles la plupart des biens sont importés de l'extérieur (Chrysoulakis et al. 2015). L'indicateur d'importation d'une ville est ainsi probablement directement additif sur ses quartiers, et donc le ratio par habitant est indirectement additif.

Cette additivité du ratio d'importation peut aussi s'appliquer à de plus grands territoires, mais pour des biens spécifiques : ceux qui ne sont pas produits sur le territoire. Une illustration avec un exemple simple : la masse des bananes importées est très certainement additive dans les pays nordiques, là où le climat ne permet pas leur culture. Il faut toutefois être vigilant car les chargements et déchargements logistiques successifs peuvent venir masquer l'origine réelle des produits dans les données et donc rendre non vérifiable cette assertion.

Néanmoins, dans les autres contextes où cette sensibilité à l'échelle produit des effets, il est légitime de vouloir disposer d'indicateurs d'importations pour comparer des territoires. Aussi, nous présentons deux approches qui permettent d'aborder plus adéquatement cette sensibilité à l'échelle.

1) Dissocier l'origine d'importation

La première approche consiste simplement à dissocier l'origine des matériaux importés par une région R_A entre ceux provenant d'un territoire plus large R_B , et ceux venant de l'extérieur. Cela suppose que R_A est inclus dans R_B .

Dans notre exemple, nous comparons les importations de la région Grand Est avec celles des territoires qui la constituent. Ainsi, pour chaque département, nous dissocions les importations venant du Grand Est de celles venant de l'extérieur. Les valeurs correspondantes

sont présentées colonnes 4 et 5 du Tableau 18. Dans ce cas, les ratios d'importation hors du Grand Est sont mieux comparables car ils sont indirectement additifs.

2) Calculer l'importation individuelle

Une approche souvent mobilisée pour éviter le MAUP est de travailler à l'échelle micro, c'est-à-dire à l'échelle des individus (Dusek 2005). Considérant notre modèle mathématique, cette approche signifierait de diviser le monde Ω en autant de territoires que d'habitants. T_i devient ainsi un groupe de h territoires d'un seul habitant, et $P(\Omega) = N(\Omega) = nh$. Nous pouvons calculer la masse d'importation théorique pour un unique habitant pour une région R donnée. Nous l'appellerons l'*importation individuelle* et nous la noterons $m_{ind}: \mathbf{E} \rightarrow \mathbb{R}$. Nous pouvons l'obtenir par la formule suivante :

$$m_{ind}(R) = \frac{P(\Omega) - 1}{P(\Omega) - P(R)} \cdot m(R)$$

En application numérique, nous réutilisons les valeurs précédentes et en complétant avec de nouveaux paramètres :

$$n = 96 ; x = 10 ; m(T_i) = \alpha = 20 \text{ et } m(R_{10}) \approx 18.1,$$

$$P(T_i) = h = 675\,000, \text{ donc } P(R_{10}) = xh = 6\,750\,000 \text{ et } P(\Omega) = 64\,800\,000^{132}.$$

Dans ce cas, nous obtenons : $m_{ind}(T_i) \approx 20,2$ et $m_{ind}(G) \approx 20,2$. Ils sont tous les deux égaux, comme attendu d'une application indirectement additive dans un monde homogène et où l'individu étudié est identique dans T_i et dans R_{10} .

Toutefois, l'application de cette approche mathématique avec des données réelles est plus complexe. Notre modèle décrit en effet un monde Ω fermé dans lequel les flux sont distribués de manière homogène. De plus, un habitant unique ne peut pas constituer une économie à lui seul, et donc la notion d'importation à cette échelle individuelle n'a probablement pas beaucoup de sens. Il n'en reste pas moins que cette notion d'importation individuelle a le grand avantage d'être additive, et donc plus adéquate pour comparer des territoires.

D'un point de vue théorique, et considérant une région R comme homogène, l'importation individuelle peut aussi être obtenue en additionnant les importations $M(R)$ avec

¹³² Ce qui est proche de la population réelle en France métropolitaine.

les flux internes $M_I(R)$, c'est-à-dire tous les flux qui ont leur origine et destination dans R , puis en divisant par le nombre d'habitants $P(R)$:

$$m_{ind}(R) = \frac{M(R) + M_I(R)}{P(R)}$$

Nous soulignons au passage que l'application $[M + M_I]: \mathbf{E} \rightarrow \mathbb{R}$ est directement additive.

Nos données statistiques incluent les flux internes qui ont le même département d'origine et de destination. La masse M_I de ces flux internes est même souvent plus importante que la masse M des importations pour la plupart des départements (Tableau 18, colonnes 2 et 6). Finalement, nous calculons $[M + M_I]$ (colonne 7) pour obtenir l'importation individuelle m_{ind} (colonne 8). Cet indicateur peut être utilisé pour comparer les différents territoires, indifféremment de leur taille.

Particulièrement, la Corse peut avoir l'air très singulière lorsque nous regardons son ratio d'importation : $m(Corse) = 0,4$ t/hab en 2010, c'est le point solitaire en bas de la répartition des ratios d'importations régionaux (Figure 42 p. 204). À cause du manque de données, nous n'avons pas pu prendre en considération le transport maritime de marchandises ce qui est un problème pour étudier une île. Cela étant, l'importation individuelle calculée d'un Corse est $m_{ind}(Corse) = 40,2$ tonnes, c'est proche de la moyenne française de 37,1 tonnes constatée dans le paragraphe suivant. De ce point de vue, la Corse ne semble donc pas si singulière, ni plus efficace dans l'usage des matériaux comparé au reste de la France. Les Corses gèrent peut-être simplement mieux leurs ressources pour eux-mêmes, ce qui est attendu d'un territoire isolé comme une île. Nous réservons l'analyse de la typologie des produits circulant sur ce territoire pour de futurs travaux.

Afin d'explorer le concept d'importation individuelle, nous représentons ses composantes (population et $[M + M_I]$) dans un graphique x-y pour tous les départements en Figure 43.

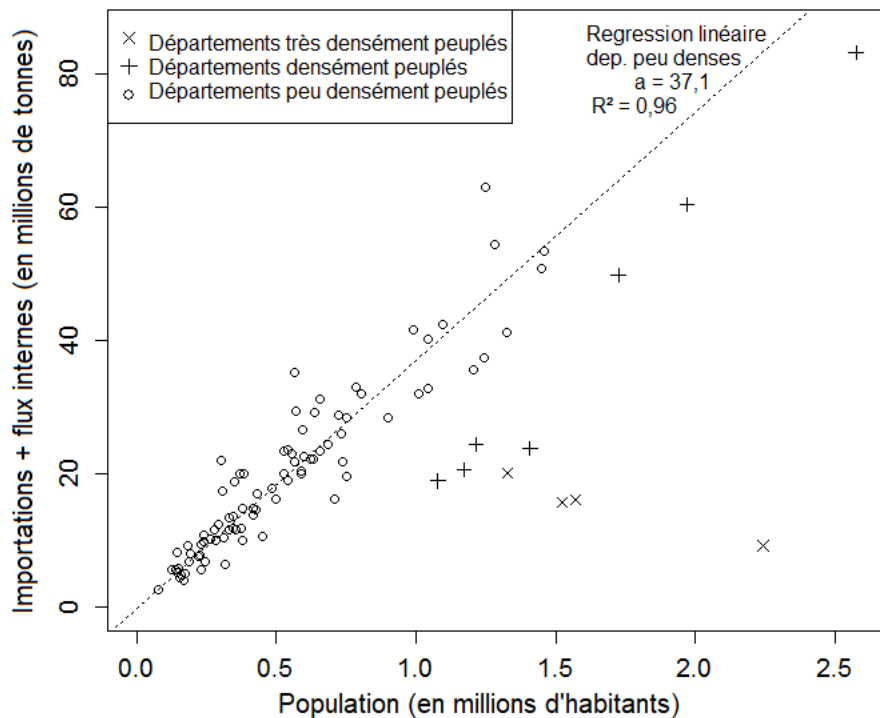


Figure 43 : Masse des importations + flux internes de marchandises en fonction de la population pour les 96 départements français. Marchandises transportées par camion et voie fluviale uniquement.

Très densément peuplé : $>1000 \text{ hab/m}^2$, densément peuplé : $>250 \text{ hab/m}^2$.

Sources : SitraM (2010). Population Insee (2010).

Cette figure nous permet de remarquer certains éléments intéressants. Tout d'abord, il y a une bonne corrélation linéaire pour la plupart des départements entre les deux composantes. Cela nous donne une valeur moyenne pour ces départements $m_{ind} \approx 37,1$ tonnes. Considérer seulement M , sans M_I , ne donne pas une aussi bonne corrélation. De toute façon, la sensibilité à l'échelle de M fait d'une éventuelle corrélation linéaire un objet sans signification réelle.

Nous remarquons également que certains départements ne suivent pas la tendance majoritaire, et s'écartent sur la droite. Ces départements ne sont pas aléatoires : ils sont les plus densément peuplés en France. Nous identifions deux explications possibles à cette différence. D'une part, les départements densément peuplés accueillent probablement moins de moyens de production. Aussi, une large partie des matériaux, utilisés et relâchés durant le processus de fabrication (les flux indirects) ne sont pas importés par ces territoires. Seulement les produits finis sont importés et donc une masse directe de matière moindre. Une autre explication serait que ces territoires densément peuplés ont atteint une saturation dans leur urbanisation, soit ont une densité d'usage plus importante de matériaux de construction. Dans les deux cas, cela

résulte en des besoins moindres de ces matériaux qui représentent une part importante des flux en termes de masse.

V. Conclusion du septième chapitre

La comparaison est une approche essentielle dans n'importe quelle science. Elle permet de souligner des particularités ou des constantes dans les dynamiques des systèmes (des territoires dans notre étude). Cela permet ainsi une meilleure compréhension de leur fonctionnement. Mais la comparaison pose aussi des limites conceptuelles qu'il faut connaître pour éviter les mauvaises interprétations.

C'était dans cet objectif que nous avons montré le lien entre le problème d'agrégation spatiale (MAUP) et certains indicateurs des méthodes de comptabilité matière, notamment la EWMFA. Grâce à un modèle mathématique simplifié et une étude de cas, nous avons montré que certains indicateurs sont ainsi sensibles à l'échelle et que leurs valeurs tendent à décroître quand le système spatial de référence s'élargit. Ainsi, la comparaison de territoires sur l'usage de ces indicateurs devrait être évitée. Des exceptions peuvent être notées pour les territoires aux dynamiques très linéaires, comme des villes, ou pour l'étude de produits particuliers et exotiques.

Deux approches ont été présentées pour construire des indicateurs additifs, c'est-à-dire qui ne soient pas sensibles à l'échelle : la dissociation des origines des importations, et le calcul de l'importation individuelle. Cette dernière notion nous permet de construire un assez bon modèle linéaire et d'observer l'influence de la densité de population sur ces indicateurs de masse. Pour ces raisons, il nous semble intéressant d'ajouter l'importation (et/ou exportation) individuelle dans le panel des indicateurs additifs de la méthode EWMFA lorsque cela est possible.

La compréhension de ce nouvel indicateur reste toutefois à approfondir, notamment en étudiant distinctement les différents types de matières. Le développement d'un modèle mathématique plus complexe, qui rapproche les hypothèses de la réalité, peut également aider à mieux identifier les facteurs influençant cette sensibilité à l'échelle et servir ainsi d'approche exploratoire. La question peut également être adressée à une audience plus large pour laquelle les notions de flux territorialisés, d'importation et d'exportation font sens, par exemple en économie avec les flux monétaires.

Chapitre 8 Le circuit de la matière alimentaire à Strasbourg

Durant cette thèse, nous avons travaillé sur la question des flux alimentaires à l'échelle de l'Eurométropole de Strasbourg. Ces travaux se sont inscrits dans le cadre d'une coopération avec la Zone Atelier Environnementale Urbaine. La ZAEU vise à favoriser les échanges et la coopération entre des laboratoires, des services de la collectivité et d'autres partenaires territoriaux pour l'étude et la compréhension des relations entre la société et son environnement dans la métropole strasbourgeoise.

Derrière la matière alimentaire, il y a des enjeux sanitaires et écologiques. Il s'agit en effet de permettre l'accès à chacun à une nourriture en quantité et qualité suffisante, tout en réduisant les impacts liés à la production, notamment l'épuisement des sols et le réchauffement climatique.

Le travail présenté dans ce chapitre n'est pas abouti, la collecte des données s'étant avérée particulièrement difficile, comme souvent dans les études de métabolisme. Nous tenons toutefois à partager notre travail en l'état car il nous a permis de rassembler différents éléments qui nous semblent intéressants ; il a notamment permis d'utiliser le SINAMET sur d'autres jeux de données.

Nous commençons par présenter notre démarche sur cette question, et les quelques résultats obtenus. Nous discuterons ensuite des apports et des limites que l'étude peut révéler.

I. Présentation de la démarche

1) Partir d'un modèle du métabolisme humain

Notre essai part d'une quantification du métabolisme d'un être humain. À la différence de Santorio (Chapitre 2/II), nous ne disposons pas de patient pour les mesurer, la littérature à ce sujet est heureusement suffisante pour ne pas devoir mener cet exercice. Évidemment, ces entrées et sorties dépendent fortement de nombreux paramètres tout aussi bien endogènes qu'exogènes : physiologies des individus, habitudes culturelles, activités physiques, catégories socio-professionnelles, conditions climatiques... (MacElroy et al. 1992 ; EFSA 2010 ; Anses

2017). Il faut donc s'attendre à de fortes variations selon les contextes. Nous retenons toutefois les valeurs moyennes présentées Figure 44 et nous justifions ces valeurs en Annexe 7.

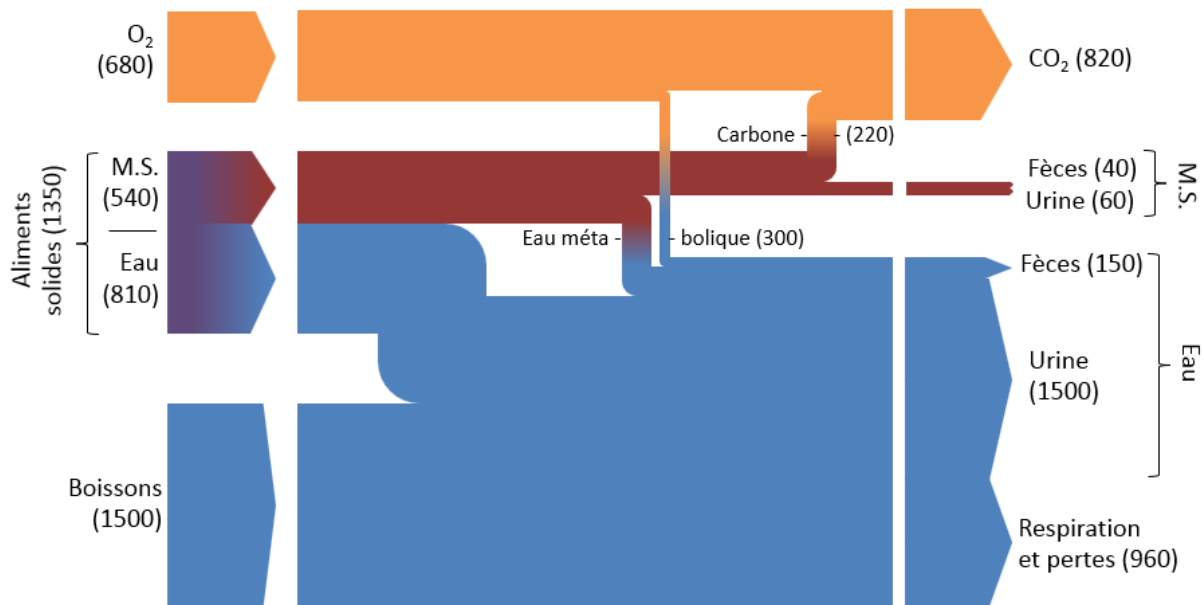


Figure 44 : Métabolisme moyen d'un corps humain (quantités en grammes par jour).

M.S. = Matière Sèche. Sources : voir Annexe 7

À partir de ce modèle, à multiplier par le nombre d'habitants de l'Eurométropole (environ 500 000), nous estimons qu'environ 246 kt d'aliments solides sont consommées chaque année sur le territoire. Dans la suite, nous tentons de déterminer en amont et en aval les flux engendrés pour satisfaire cette consommation.

2) Flux en amont : Production, transformation, distribution, préparation

En amont de la consommation d'aliments, nous identifions les activités de production, de transformation, de distribution et de préparation par un restaurateur ou dans les ménages.

a) Production locale

La production agroalimentaire est décrite à l'échelle départementale, à travers les données statistiques « Agreste » diffusées par le Ministère de l'Agriculture et de l'Alimentation. Nous pourrions réduire l'échelle (Chapitre 3/I.1) à l'aide des données du Registre Parcellaire Graphique (RPG) qui nous permettent de connaître les surfaces affectées à la production agricole, le type de production et leur localisation. Cela nécessite toutefois un travail sur la conversion de nomenclature entre les données Agreste et celles du RPG qui n'a

pas encore été mené. De manière plus macro, nous disposons toutefois de l'information d'une surface agricole de 10 400 ha au sein de la métropole (Speich 2016). Sur une base de 2000 m² nécessaires par habitant (ADEUS 2017), cela permettrait de satisfaire les besoins de 52 000 habitants, soit 10 % des besoins totaux, si tant est que l'intégralité de la production corresponde aux besoins nutritionnels et soit consommée sur place. En considérant 25 % de pertes entre la production et la consommation (ADEUS 2017), et sur la base de 1,35 kg/jour/personne de nourriture consommée (Figure 44), nous évaluons la production alimentaire métropolitaine à environ 34 000 tonnes annuelles.

b) Industrie agroalimentaire, distribution, restauration

En ce qui concerne la transformation des produits alimentaires, nous identifions sur le territoire 391 entreprises du secteur de l'industrie agroalimentaire à l'aide de la base SIRENE. Nous sommes toutefois en peine de préciser les volumes et la nature des produits mobilisés par ces industries. Il en est de même pour la distribution et la préparation des repas : les données SIRENE peuvent nous aider à identifier et localiser les entreprises du secteur, sans toutefois nous apporter plus d'informations. En consultant l'enquête de l'Anses (2017), nous arrivons à estimer qu'environ un tiers des repas est pris hors du domicile, mais cela ne dit pas s'il s'agit de restauration professionnelle ou non : cela peut concerner des repas chez des amis ou la préparation à domicile de panier-repas consommés sur le lieu de travail.

c) Associations caritatives

Parmi les composantes du circuit de la matière alimentaire, les associations caritatives ont un rôle particulier. Elles peuvent jouer un rôle de distributeur, voire de préparateur des denrées alimentaires pour les ménages précaires. Elles peuvent également avoir un rôle dans la récupération des invendus alimentaires avant leur date de péremption auprès des industriels ou des distributeurs, avant de le redistribuer évitant ainsi qu'ils ne deviennent des déchets dans le respect des règles d'hygiène (DILA 2011).

Au niveau du département, la Banque Alimentaire du Bas-Rhin (BA67), créé en 1985, joue ce rôle et distribue chaque année près de 2 000 tonnes de nourritures selon les quelques données qui nous ont été partagées par l'association. La BA n'est pas en relation directe avec les bénéficiaires, ce sont différentes associations qui assurent la distribution des denrées collectées. Ces denrées proviennent de dons des particuliers, notamment lors de la collecte nationale (346 t en 2017), de tournées de récupération quotidiennes dans les grandes et moyennes surfaces (1111 t), de dons des industries agroalimentaires, des producteurs et des

fournisseurs (436 t), et de dotations en nourriture dans le cadre du Fond Européen d'Aide aux Démunis et du Plan National d'Aide Alimentaire (669 t).

Nous n'avons toutefois pas pu obtenir de données plus détaillées sur l'origine et la destination de ces aliments. Il nous est ainsi difficile de distinguer les flux intra-métropolitains de ceux avec le reste du département. Le rôle de la Banque Alimentaire reste toutefois quantitativement marginal sur le territoire, mais n'est pas non plus anodin pour un seul acteur. En plus de la Banque Alimentaire, les Restos du Cœur est l'autre grande organisation qui assure une collecte et une distribution caritative de nourriture. Nous n'avons toutefois pas d'information plus précise pour quantifier leur activité.

3) Flux en aval : Déchets alimentaires et eaux usées

a) Eaux usées

Une partie des sorties physiologiques humaines se retrouvent dans les eaux usées, nous estimons cela à 100 g de matière sèche par jour et par personne, soit environ 18 250 t par an pour l'Eurométropole. Nous avons pu récupérer les données de fonctionnement de la principale station d'épuration (STEP) de Strasbourg (Tableau 19) qui assure le traitement de 95 % des eaux usées.

Nom de la source : Données STEP Strasbourg	Producteur : Agence de l'Eau Rhin-Meuse / SIERM
Détails :	Données mesurées de fonctionnement de la Station d'Épuration (STEP) de Strasbourg
Unité de quantification : kg / kg eq. N / kg eq. P / kg eq. O ₂ selon produits	
Période : 2009 – 2017	Pas de temps : Annuel
Territoire couvert : STEP Strasbourg	Niveau de détail : /
Classification des produits :	Phosphore total ; Nitrates ; Nitrites ; Ammonium ; Azote Kjeldahl ; DCO ; DBO5 ; Matières en suspension ; DCO ad2 ; DBO5 ad2
Format des données : 1 fichier .csv zippé	
Obtention :	http://rhin-meuse.eaufrance.fr/resultats-EPU?perimetre[026751900405]

Tableau 19 : Données mesurées sur la station d'épuration de Strasbourg

À l'aide du SINAMET, nous représentons l'évolution des principaux paramètres en termes de masse : matière en suspension, phosphore, ammonium et azote (Figure 45). Au total, cela représente environ 19 000 t par an, ce qui est très proche des estimations. Il faut toutefois également inclure d'autres produits que ceux issus du métabolisme humain qui peuvent se retrouver également dans les eaux usées, comme environ 5 kg par an et par personne de papier toilette (Rose et al. 2015).

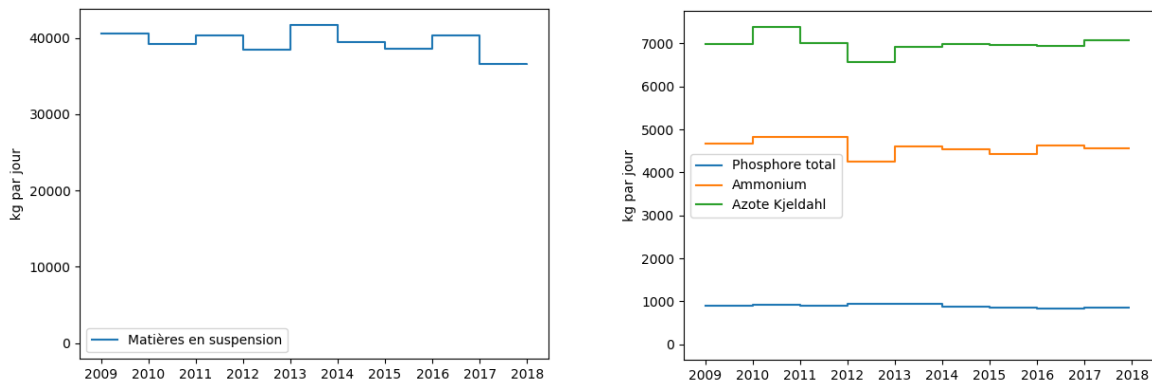


Figure 45 : Mesures de certains composés des eaux usées entrantes dans la station d'épuration Strasbourg Wantzenau (données SIERM).

Si nous nous intéressons plus spécifiquement au phosphore, la littérature évalue la quantité évacuée à environ 1,5 gramme par jour et par personne, soit 750 kg/jour à l'échelle de la métropole. C'est dans le même ordre de grandeur que la valeur mesurée au niveau de la STEP, d'un peu moins de 1000 kg par jour (Figure 45).

Après traitement de ces eaux usées, une partie du carbone résiduel sera relâché, il reste ainsi environ 12 000 tonnes de boues produites chaque année par la STEP, et dont il faut s'occuper, par exemple en pratiquant l'épandage si leur qualité est satisfaisante.

b) Déchets alimentaires

Nous avons déjà esquissé quelques problématiques au sujet des déchets alimentaires, en application de l'« octomètre » du métabolisme territorial (Chapitre 3/III, notamment Tableau 5 p. 115). Les données que nous avons pu récupérer restent toutefois limitées pour répondre à toutes ces problématiques. Deux d'entre elles peuvent toutefois être abordées : Quelle est la quantité de déchets alimentaires produits par les ménages ? Quelle est l'accessibilité des composteurs publics dans la ville ?

Selon l'ADEME (2016), la quantité de déchets alimentaires produits par les ménages est évaluée au niveau national à 20 kg de déchets alimentaires par an et par personne¹³³, mais certaines sources indiquent jusqu'à quatre fois cette valeur¹³⁴. Pour savoir ce qu'il en est dans l'EMS, nous avons pu récupérer les données quotidiennes relatives à la collecte des déchets des ménages et assimilés (Tableau 20). Ces déchets sont destinés à être incinérés.

Nom de la source : Collecte des déchets EMS Producteur : EMS	
Détails :	Poids des ordures ménagères collectées pour les différents circuits de collecte des déchets en porte à porte
Unité de quantification : kg	
Période : 2018	Pas de temps : Quotidien
Territoire couvert : EMS	Niveau de détail : Circuit de collecte
Classification des produits : Déchets triés / déchets non triés	
Format des données : 1 fichier .xls par trimestre	
Obtention : Communiqué par les services de la collectivité	

Tableau 20 : Données de collecte des déchets en porte à porte (ménages et assimilés)

Les données collectées fournissent assez peu d'information pour notre sujet, nous remarquons toutefois Figure 46 la stabilité dans les quantités de déchets ménagers collectées : entre 2 000 et 2 500 tonnes par semaine, avec quelques variations saisonnières comme un mois de mai marqué par les jours fériés et une très légère baisse en juillet-août.

¹³³ <https://www.ademe.fr/expertises/dechets/passer-a-laction/eviter-production-dechets/dossier/reduire-gaspillage-alimentaire/enjeux>

¹³⁴ <https://zero-gachis.com/fr/quelques-chiffres>

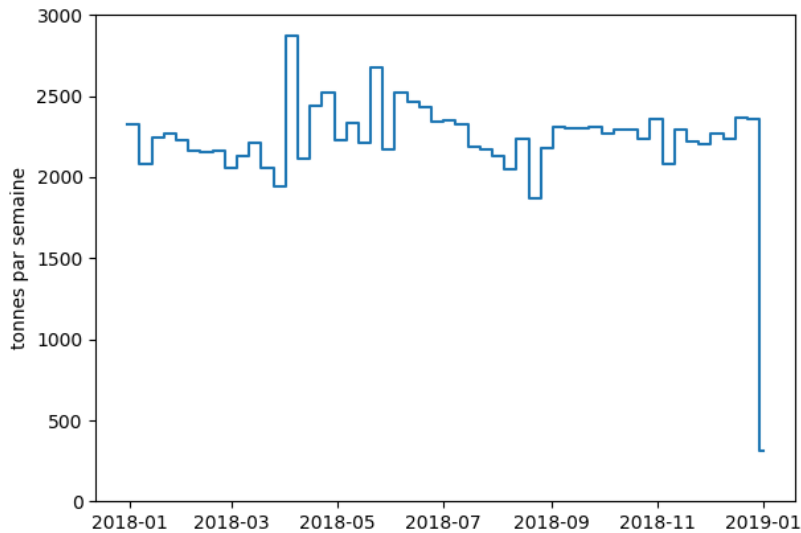


Figure 46 : Collecte des déchets ménagers (non triés) par le service de collecte de l’Eurométropole de Strasbourg

Au total, ce sont près de 118 000 tonnes de déchets non triés collectés en 2018 (Figure 47). Selon un travail d’échantillonnage réalisé sur les déchets de l’EMS en 2012, on estime qu’environ 20 % des déchets seraient des aliments, ce qui représenterait environ 50 kg/an/personne. Dans ces déchets, il faut toutefois remarquer qu’il y a certains points de collecte assimilés aux ménages, c’est-à-dire possiblement des restaurateurs. Ce chiffre de 50 kg ne concerne donc pas strictement que les ménages. La valeur de 20 kg/an/personne annoncée par l’ADEME nous paraît toutefois sous-évaluée au regard du cas strasbourgeois.

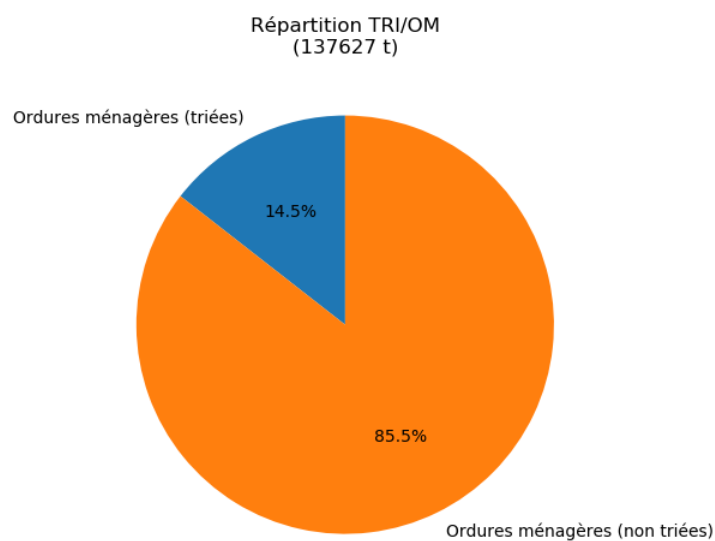


Figure 47 : Masse des ordures ménagères collectées en porte-à-porte en 2018 dans l’Eurométropole de Strasbourg

Une partie des déchets alimentaires des particuliers n'est pas collectée par les services de ramassage de l'EMS mais est compostée. Il est cependant difficile d'en évaluer les quantités, peu de données étant disponibles. L'EMS subventionnant l'achat de (lombricomposteurs, il aurait été intéressant de pouvoir disposer au moins du nombre de ménages qui ont sollicité cette subvention, pour en suivre l'évolution. À défaut de ces données, nous avons pu disposer d'une liste de composteurs collectifs disposés sur la voie publique et gérés par des associations. Cela nous a permis de construire la carte des composteurs sur la voie publique et des zones à prioriser pour de futures installations (Figure 48), ainsi que des indicateurs associés (Tableau 21).

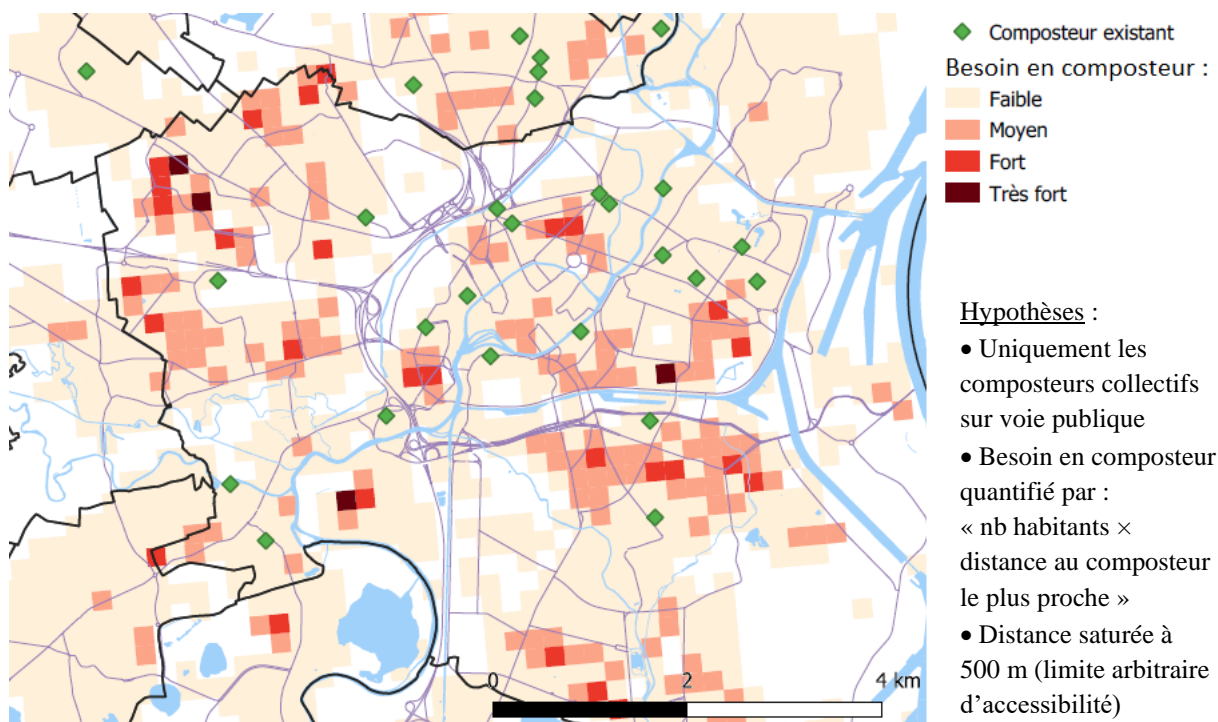


Figure 48 : Localisation des composteurs collectifs sur voie publique de l'Eurométropole de Strasbourg et zones à prioriser pour l'installation de nouveaux composteurs.

Sources : Insee (données de population carroyées) et EMS (localisation des composteurs), 2018

Indicateur	Valeur pour l'EMS
Part de la population à moins de 200 m d'un site de compostage (bonne accessibilité / limite de détail carroyage Insee)	4,7 %
Part de la population à moins de 500 m d'un site de compostage (distance considérée rédhibitoire)	24,5 %
Distance médiane au composteur le plus proche	1012 m

Tableau 21 : Propositions d'indicateurs d'accessibilité des composteurs collectifs (données 2018)

Ce que nous constatons est la plutôt mauvaise couverture du compostage sur voie publique dans l'EMS. En fait, le déploiement de composteurs sur le territoire repose pour l'instant sur le volontariat des associations qui acceptent de les gérer. Selon les services de l'EMS, la plupart de la trentaine de composteurs installés sont déjà saturés en raison d'une forte attente des habitants. En faisant l'hypothèse d'un fonctionnement à régime maximum, soit une tonne par site par semaine (AMORCE 2018), cela représente tout au plus 1500 tonnes de déchets traités annuellement par cette voie. À cela, s'ajoutent les efforts des particuliers et des copropriétés qui disposent pour certains de leur propre bac à compost, non recensé ici.

La Figure 48 fait également apparaître certains quartiers comme déficitaires en composteurs publics, et nous constatons qu'il s'agit en général de quartiers défavorisés de la ville. Notre carte reflète ainsi probablement et indirectement une vie associative moins importante dans ces quartiers ou moins engagée sur ces questions écologiques.

L'approche spatiale présentée ici ne traduit pas adéquatement le mode de gestion des points de collecte. En revanche, elle pourrait être plus adéquate sur d'autres thématiques, par exemple celle de la collecte volontaire de verre qui ne repose pas sur une gestion associative.

4) Synthèse

Nous avons présenté différents éléments autour de la question du circuit de la matière alimentaire dans l'Eurométropole de Strasbourg. Ces éléments sont malheureusement partiels, et ne permettent pas de construire une vision globale des flux. Ils contribuent néanmoins à apporter des éléments d'éclairage et des ordres de grandeur pour certains flux. Nous précisons ainsi Figure 49 un circuit global de la matière alimentaire et les quelques quantités que nous avons pu évaluer.

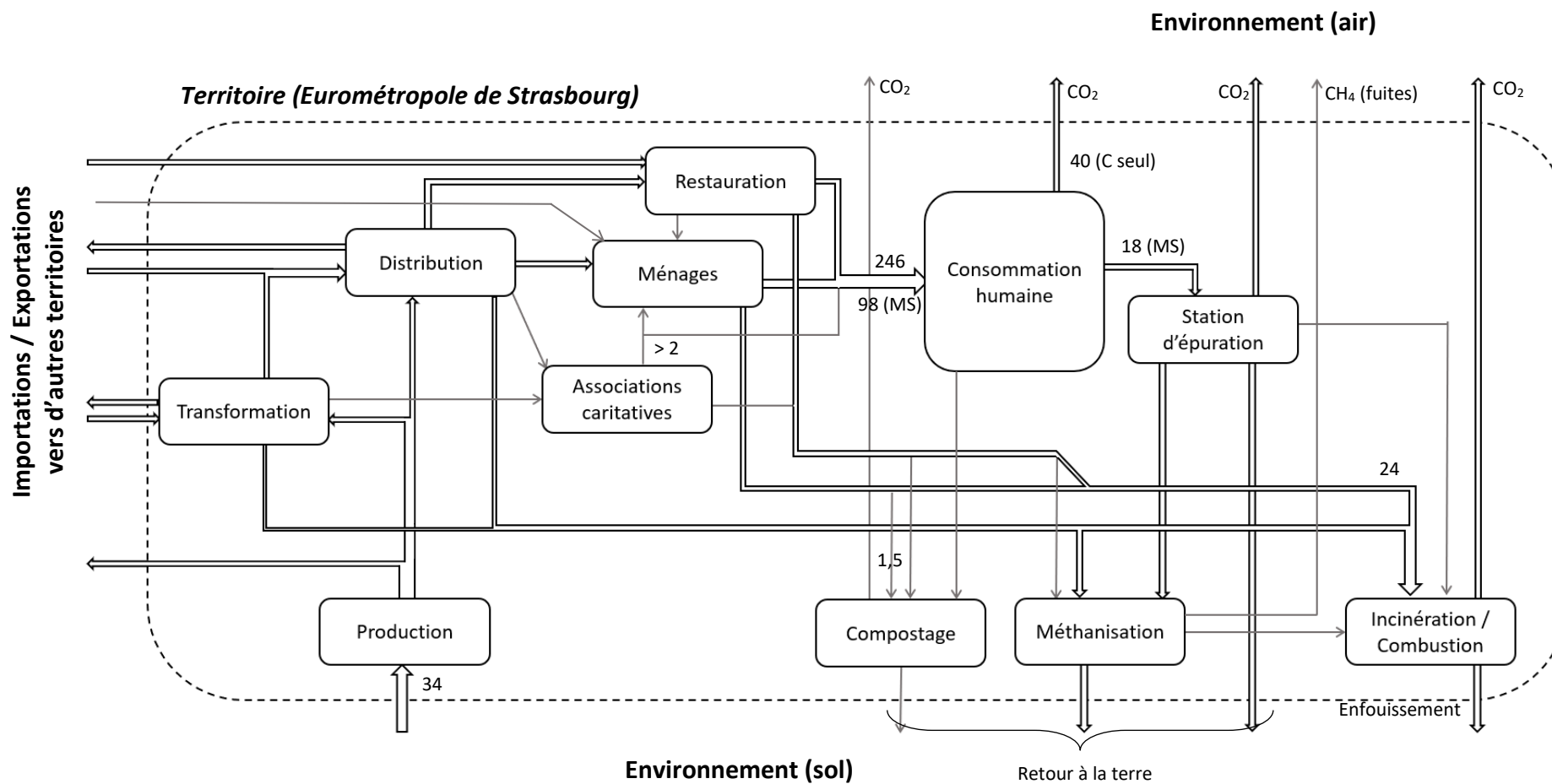


Figure 49 : Circuit de la matière alimentaire (hors boissons) dans une agglomération (valeurs en kt/an estimées pour l'Eurométropole de Strasbourg).

MS : Matière Sèche ; C seul : Composé carbone uniquement. Note : la taille des flèches n'est pas proportionnelle.

II. Discussions et conclusion du huitième chapitre

Cette étude avait pour objectif de dresser une vision d'ensemble du circuit de la matière alimentaire à l'échelle d'une métropole. Nous avons réussi à rassembler différentes pièces d'un puzzle, mais il manque encore un certain nombre d'informations pour construire une vision d'ensemble satisfaisante. Nous pouvons toutefois déjà remarquer des cohérences dans les quelques données collectées, notamment entre des valeurs issues de la littérature, et celles mesurées sur le terrain, par exemple sur les rejets humains vers la station d'épuration. Notre approche reste également très générale, puisqu'elle considère des tonnages globaux et n'entre pas dans le détail des produits alimentaires.

Le SINAMET a été mobilisé sur deux jeux de données : la collecte de déchets, et les données de la station d'épuration. Ces deux jeux structurés contiennent en effet suffisamment de données pour que leur manipulation et leur analyse à travers le SINAMET soient pertinentes. D'autres données sont en revanche isolées (production moyenne de déchets alimentaires, part des associations caritatives, ou de la production alimentaire globale...), et les agréger au sein de SINAMET ne serait pertinent que si ce travail était réutilisé par la suite, ce qui n'est pas certain. Ainsi, dans le cadre de nos travaux, nous avons jugé que ce travail de préparation des données était trop important pour être mené entièrement.

Une limite importante du SINAMET mise en valeur par ce cas d'étude est la difficulté à mobiliser et articuler des modèles, notamment de transformation. Dans cette étude, notre réflexion part d'un modèle du métabolisme humain, auquel viennent s'articuler d'autres modèles comme celui de la production de déchets lors de la préparation des repas. Nous n'avons cependant pas encore trouvé de moyen élégant d'articuler la notion de transformation avec les autres notions du métabolisme que sont les territoires, les acteurs, les produits, les flux et les stocks (Chapitre 4/III.5). Ainsi, en dehors de l'analyse des deux jeux mentionnés précédemment, le prototype du SINAMET n'a pas été utile dans cette étude.

Malgré un travail à poursuivre, cette étude nous apporte tout de même quelques éléments de compréhensions de la thématique, débordant même sur des thématiques connexes (la collecte de déchets ménagers, la vie associative des quartiers...) montrant ainsi la transversalité de l'approche par les flux. Cette approche mobilise différentes typologies d'analyse, notamment comptable, spatiale ou basée sur une modélisation, qui viennent se compléter, mais qui nécessitent encore leurs outils propres comme un SINAMET pour l'approche comptable ou un SIG pour l'approche spatiale.

Conclusion générale

I. Apports et limites identifiés à travers les applications

Nous avons présenté quatre exemples d'applications des différents éléments théoriques développés pendant nos travaux (introduction Partie III). Ces quatre applications ne reposent toutefois pas sur la même approche.

Dans l'étude sur les consommations d'énergie (I) et dans celle du transport de marchandises par voie fluviale (II), nous avons directement travaillé sur l'exploration de données disponibles à l'aide SINAMET pour constituer une preuve de l'intérêt de notre démarche. L'étude sur la sensibilité à l'échelle des indicateurs (III) a également été réalisée à partir de données manipulées à travers le SINAMET. Nous avons cependant mis la focale sur l'interprétation des résultats et les implications dans la construction d'indicateurs, plutôt que sur le traitement des données lui-même. Enfin, dans l'étude sur le circuit de la matière alimentaire (IV), nous sommes partis d'un objectif global et avons tenté de rassembler les données qui s'y rapportent.

Dans tous ces cas, nous avons montré que l'usage du SINAMET nous offre une grande dextérité pour la manipulation de données comptables, massives et structurées. Il nous permet de facilement mettre en œuvre des méthodes d'analyses de données, notamment les constructions graphiques sur lesquelles nous nous sommes largement appuyés. Il permet également de capitaliser le travail de construction algorithmique, facilitant ainsi la réutilisation des données dans d'autres analyses, ou des mêmes analyses avec d'autres données. Nous avons toutefois identifié deux limites principales dans l'utilisation du SINAMET. D'une part, la

préparation d'un jeu de données à travers le logiciel prend tout de même du temps. Lorsqu'il s'agit de gérer de grands jeux de données, le retour sur le temps investi ne fait pas de doute, mais lorsqu'il s'agit de compiler des données isolées, ce retour sur investissement n'est pas évident, surtout pour une itération unique. D'autre part, si le SINAMET permet de faciliter les approches comptables, ce n'est pas le cas pour l'implémentation d'approches modélisatrices qui peuvent pourtant être nécessaires pour mieux comprendre la circulation des flux.

C'est d'ailleurs un deuxième enseignement de ces exemples d'application : l'approche comptable est complémentaire à d'autres approches, comme la modélisation ou la cartographique que nous avons également mobilisées. La comptabilité des flux contribue bien à la compréhension de certaines dynamiques d'un territoire en révélant différents phénomènes (saisonnalité, anomalies, intensité d'échange avec d'autres territoires...), mais ces phénomènes sont souvent difficiles à expliquer avec les seules données comptables.

En combinant les approches, nous avons pu analyser différents jeux de données pour en extraire une information mieux assimilable, et même opérationnelle dans certains cas. La liste d'applications et l'explicitation de différentes dimensions du métabolisme nous ont aidés dans la démarche en facilitant la formulation des interrogations et la structuration de l'information.

Nous nous sommes également appuyés sur le principe de l'exploration de données, notamment sous forme graphique en particulier dans les deux premières applications. Cette approche se révèle utile pour mieux comprendre les dynamiques territoriales en mettant en lumière des phénomènes parfois inattendus. Elle suppose toutefois d'avoir une grande liberté dans la manière d'explorer les données, afin de pouvoir analyser sous différents angles ces phénomènes. Nous avons pu avoir cette liberté grâce au SINAMET.

Un dernier point que nous pouvons souligner à travers les applications est la difficulté à collecter et utiliser les données. Cela demande souvent une coopération active avec les acteurs producteurs des données pour pouvoir mener des analyses plus fines et pertinentes. Néanmoins, ces coopérations, bien qu'existantes, manquaient un peu de teneur. Cela nous a obligé à nous contenter parfois de données partielles, peu utilisables ou de qualité discutable alors qu'elles existaient probablement sous une meilleure forme mais ne nous ont pas été communiquées.

En lien avec une meilleure utilisabilité des données, il nous a semblé également important de décrire de façon synthétique et systématique les données que nous avons mobilisées dans les différentes applications, et même les difficultés liées à leur préparation pour certains. Cela nous paraît être une bonne pratique pour faciliter leur réemploi éventuel dans de futurs travaux.

II. Contributions et perspectives

À travers cette thèse, nous avons interrogé le paradigme du métabolisme territorial à travers la question du traitement des données qui en permet l'étude. Cette approche transdisciplinaire interroge tant les moyens que les finalités du paradigme, d'un point de vue aussi bien général qu'opérationnel. Notre démarche fait émerger différents aspects, à la fois sur le traitement de données et sur les façons de comprendre et d'appliquer cette notion de métabolisme territorial. À partir de ces éléments, nous avons pu conceptualiser un système d'information pour traiter les données relatives à cette notion de métabolisme.

Dans notre approche, nous avons essentiellement considéré le traitement des données comme le moyen d'étude de la thématique du métabolisme territorial. C'est une posture assez inédite qui permet pourtant de mettre en lumière un certain nombre d'éléments qui ont une incidence forte sur la faisabilité et la complexité des études de métabolisme. Certains aspects du traitement de données dépassent largement la seule thématique du métabolisme, nous les avons présentés dans le Chapitre 1. C'est ainsi que nous identifions différentes étapes (collecte, préparation, stockage, analyse de données), différentes techniques (structures, représentations, concepts algorithmiques et de base de données...) et différents enjeux sociétaux (scientifiques, économiques, politiques, règlementaires...) liés à l'usage et au traitement de données. Nous soulignons en particulier la dépendance des études aux données sources, et l'importance de rendre ces dernières plus *utilisables*. En effet, une part importante de notre énergie fut consacrée à rechercher, collecter et préparer des données. Il s'agit là d'un travail à la fois chronophage, mais également ingrat car la valeur ajoutée de ces étapes est souvent faiblement perçue. Rendre les données plus utilisables réduirait ainsi les efforts nécessaires dans les études. Nous n'avons toutefois que peu de marge de manœuvre pour cela, ce sont surtout les producteurs des données qui doivent agir en ce sens. Pour apporter une contribution pour renforcer l'utilisabilité des données à notre niveau plus modeste, nous avons réalisé une synthèse des jeux de données qui peuvent être mobilisés dans les études de métabolisme. Par ailleurs, nous avons choisi de bien décrire les principaux jeux de données que nous avons utilisés afin de les rendre plus utilisables dans d'éventuels futurs travaux.

En ce qui concerne la thématique du métabolisme proprement dite, nous proposons deux grilles de lecture, plus exactement une grille d'analyse selon huit dimensions, et une grille d'applications. Ces deux grilles aident à questionner, à structurer et à représenter de façon plus systématique la notion de métabolisme territorial afin de produire une information qui peut être

aussi bien générale qu'opérationnelle. Ainsi, bien que la notion soit complexe (dynamique, multiscalaire, polysémique...), il est possible de la rendre pragmatique pour éclairer certains aspects sociétaux (impacts environnementaux, respect de la réglementation, efficacité économique...) et servir de base à la construction d'un territoire commun et efficace dans la gestion de ses ressources. Toutefois, plusieurs obstacles restent à surmonter : les problématiques de traitement de données sont très présentes dans les études de métabolisme, et leur faisabilité est très dépendante de choix politiques. Ces choix résultent en des dispositions réglementaires et en des moyens territoriaux qui peuvent favoriser, ou au contraire limiter, les possibilités des études. Certainement, la sensibilisation d'une grande diversité d'acteurs aux possibilités offertes par l'analyse des flux de matières et d'énergie à l'échelle des territoires est encore à développer. Dans cette ambition, il nous semble que l'approche graphique est essentielle, car elle est la plus à même de transmettre une information complexe à un public large. Un graphique étant toutefois limité dans ses capacités de représentations, nous proposons de les multiplier et de les rendre interactifs pour diversifier l'information accessible.

La formalisation de ces différents éléments relatifs au traitement de données et au métabolisme territorial nous permet de conceptualiser un Système d'Information pour l'Analyse du Métabolisme des Territoires (SINAMET). Son objectif est de permettre un traitement plus simple et plus automatisé des données dans ce type d'étude. L'innovation technologique sur laquelle repose ce système d'information est la définition d'une ontologie opérationnelle du métabolisme territorial. Cette ontologie permet d'assimiler les données de la thématique à des propriétés d'instances de six types : territoire, acteur, produit, flux, stock et transformation. Cela permet la constitution d'une base de données pour centraliser et mettre en lien des données issues de diverses sources hétérogènes. En plus d'une base de données centrale, le SINAMET implémente un certain nombre de fonctionnalités qui permettent de gérer plus facilement l'importation des données et les liens entre les différents concepts. Ainsi, la hiérarchisation des concepts (territoires, acteurs, produits...), la quantification dans différentes unités et la gestion des coefficients, la structuration des produits dans différentes nomenclatures et leur conversion, les références par des noms et des codes multiples ou encore la gestion des références temporelles sont des fonctionnalités régulièrement mobilisées dans les études de métabolismes. En étant disponibles à travers le SINAMET, elles permettent à l'analyste une plus grande productivité, car il n'a pas besoin de développer à nouveau le modèle de données et les fonctionnalités associées. Enfin, le SINAMET a été conçu pour une utilisation modulaire. Cela permet de réutiliser les modules développés par d'autres analystes, permettant ainsi de

capitaliser une information fragmentée par les diversités des sources et un gain de temps lié à la nécessaire préparation des données. De plus, ces modules peuvent être facilement rendus paramétrables via une interface graphique. Cela permet aux utilisateurs qui sont moins familiers avec la notion de traitement de données d'obtenir quand même et facilement des informations personnalisées pour leur activité. Les différents exemples d'applications montrent la polyvalence de notre approche et l'intérêt de mobiliser le SINAMET sur différents cas d'étude : il permet d'analyser facilement des données structurées et massives (plusieurs milliers de lignes) décrivant des flux territorialisés.

Le SINAMET permet avant tout de gérer une information comptable des flux de matières et d'énergie, dans la lignée des méthodologies européenne (Eurostat) ou française (CGDD). Si cette approche permet en effet d'apporter un éclairage différent sur les dynamiques du territoire, elle ne se suffit pas à elle-même. Elle vient plutôt en complément d'autres approches, par exemple la cartographie ou la modélisation. C'est bien la combinaison de ces approches qui permet de construire une connaissance plus fine du territoire.

Par ailleurs, nous avons également rencontré des difficultés dans la conceptualisation de la notion de transformation. Si elle est essentielle pour modéliser la dynamique de certains flux, comme le montre notre étude sur les flux de matières alimentaire, nous n'avons pas trouvé de solution élégante pour structurer l'information correspondante, en lien avec les autres concepts sous-jacents du métabolisme. Réussir cette harmonisation permettrait sans aucun doute d'augmenter significativement le potentiel applicatif du SINAMET.

L'application des différents éléments présentés dans cette thèse sur les cas d'étude réels a permis de valider leur pertinence. Le lecteur pourrait regretter que l'analyse de ces cas reste superficielle, il ne s'agit en effet que de premières analyses qui mériteraient d'être approfondies. Il aurait été également intéressant d'étudier comment les acteurs concernés peuvent s'emparer de ces analyses pour entreprendre des changements effectifs dans leur manière de gérer leurs flux. Toutefois, cela n'était pas possible dans le cadre de notre travail qui se focalisait plus sur le moyen d'étude du territoire, que sur le territoire lui-même. À partir des outils développés, nous espérons toutefois pouvoir approfondir ces analyses dans de futurs travaux. Il faudra toutefois que les acteurs, en premier lieu les collectivités, s'approprient mieux la notion de métabolisme territorial. Cela est en cours, mais les moyens qui en permettent l'étude sont encore limités. Nous espérons que nos travaux viendront ainsi renforcer les moyens disponibles, et permettront d'accompagner les territoires dans une gestion plus sobre et plus efficiente de leurs ressources.

Bibliographie

Tous les liens URL cités ici ont été vérifiés le 24 juin 2020.

- ADEME, 2010, Bilan Carbone® Entreprises - Collectivités - Territoires : Guide Méthodologique - Version 6.1 - Objectifs et principes de comptabilisation. Agence de l'environnement et de la maîtrise de l'énergie
- ADEME, 2014, Opération foyers témoins pour estimer les impacts du gaspillage alimentaire des ménages. Étude réalisée pour le compte de l'Ademe par Verdicité et Biens Communs. Agence de l'environnement et de la maîtrise de l'énergie
- ADEME, 2016, État des masses des pertes et gaspillage alimentaires : état des lieux par étapes de la chaîne alimentaire. URL : <http://r485633301.racontr.com/index.html> .
- ADEUS, 2017, Le système alimentaire local bas-rhinois - Première analyse, *Les notes de l'ADEUS*. Agence de Développement et d'Urbanisme de l'Agglomération Strasbourgeoise, n°253, 8 p. URL : http://www.adeus.org/productions/les-notes-de-ladeus-ndeg253-economie/files/note-253_systeme_alimentaire_local_67-web.pdf .
- Adoue, C., 2004, Méthodologie d'identification de synergies éco-industrielles réalisables entre entreprises sur le territoire français. URL : <http://www.theses.fr/2004TROY0016> .
- AFNOR, 2015, Livre Blanc, Données massives - Big Data : Impact et attentes pour la normalisation. AFNOR Normalisation, 66 p. URL : <https://wikip.e.hassante.fr/WikiPE/PHP/Multimedia.php?Concept=112779&langue=fr> .
- Airparif, 2016, Etude prospective dans le cadre de la création d'une zone à circulation restreinte à Paris. URL : <https://api-site.paris.fr/images/87626> .
- Alphandéry, P., Fortier, A., Sourdril, A., 2012, Les données entre normalisation et territoire : la construction de la trame verte et bleue, *Développement durable et territoires*, Vol. 3, n°2. DOI : [10.4000/developpementdurable.9282](https://doi.org/10.4000/developpementdurable.9282)
- Alterre Bourgogne, 2013, La Bourgogne comptabilise ses flux de matières, *Repères*, Vol. 64

- AMORCE, 2018, Construire une organisation territoriale de compostage partagé, 17 p. URL : <https://www.compos13.fr/wp-content/uploads/2019/05/AMORCE-organisation-territoriale-de-compost-partage.pdf> .
- Anses, 2017, Étude individuelle nationale des consommations alimentaires 3 (INCA 3). Avis de l'Anses et Rapport d'expertise collective., 535 p. URL : <https://www.anses.fr/fr/system/files/NUT2014SA0234Ra.pdf> .
- ARENE IdF, ICEB, 2012, L'énergie grise des matériaux et des ouvrages. Agence régionale de l'environnement et des nouvelles énergies (Île-de-France) et Institut pour la conception environnementale du bâti (Ivry-sur-Seine, Val-de-Marne). URL : http://www.asso-iceb.org/wp-content/uploads/2012/11/guide_bio_tech_l_energie_grise_des_materiaux_et_des_ouvrages.pdf .
- Arrow, K., Dasgupta, P., Goulder, L., Daily, G., Ehrlich, P., Heal, G., Levin, S., Mäler, K., Schneider, S., Starrett, D., Walker, B., 2004, Are we consuming too much?, *Journal of Economic Perspectives*, Vol. 18, n°3, pp. 147-172. DOI : [10.1257/0895330042162377](https://doi.org/10.1257/0895330042162377)
- Assar, S., 2015, Méthodes de recherche empirique en ingénierie des SI. Principes et applications, *Revue des Sciences et Technologies de l'Information - Série ISI : Ingénierie des Systèmes d'Information*. Vol. 20, n°6, pp. 11-33. DOI : [10.3166/isi.20.6.11-33](https://doi.org/10.3166/isi.20.6.11-33)
- Audard, F., Carpentier, S., Oliveau, S., 2014, Les "big data" sont-elles l'avenir de la géographie (théorique et quantitative) ?, *20e Biennale de géographie d'Avignon - Géopoint 2014 - "Controverses et géographies"*, Juin 2014, Avignon, France, pp. 1-4. URL : <https://halshs.archives-ouvertes.fr/halshs-01132689/document> .
- Bahers, J., 2012, Dynamiques des filières de récupération-recyclage et écologie territoriale : l'exemple de la filière de traitement des déchets d'équipements électriques et électroniques (DEEE) en Midi-Pyrénées. URL : <https://tel.archives-ouvertes.fr/tel-00711199/document>
- Bahers, J., Giacchè, G., 2018, Échelles territoriales et politiques du métabolisme urbain : la structuration des filières de biodéchets et l'intégration de l'agriculture urbaine à Rennes, *Vertigo*, n°31 (Hors-série). DOI : [10.4000/vertigo.21609](https://doi.org/10.4000/vertigo.21609)
- Bamford, C., Curran, P., 1991, Data Structures, Files and Databases. Springer, pp. 264. DOI : [10.1007/978-1-349-12615-6](https://doi.org/10.1007/978-1-349-12615-6)
- Barles, S., 2009, Urban Metabolism of Paris and Its Region, *Journal of Industrial Ecology*, Vol. 13, n°6, pp. 898-913. DOI : [10.1111/j.1530-9290.2009.00169.x](https://doi.org/10.1111/j.1530-9290.2009.00169.x)
- Barles, S., 2010, Society, energy and materials: the contribution of urban metabolism studies to sustainable urban development issues, *Journal of Environmental Planning and Management*. Routledge, Vol. 53, n°4, pp. 439-455. DOI : [10.1080/09640561003703772](https://doi.org/10.1080/09640561003703772)
- Barles, S., 2014, L'écologie territoriale et les enjeux de la dématérialisation des sociétés : l'apport de l'analyse des flux de matières, *Développement durable et territoires*, Vol. 5, n°1. DOI : [10.4000/developpementdurable.10090](https://doi.org/10.4000/developpementdurable.10090)

- Barles, S., 2017, Écologie territoriale et métabolisme urbain : quelques enjeux de la transition socioécologique, *Revue d'Économie Régionale & Urbaine*, n°5, pp. 819-836. DOI : [10.3917/rru.175.0819](https://doi.org/10.3917/rru.175.0819)
- Batini, C., Cappiello, C., Francalanci, C., Maurino, A., 2009, Methodologies for data quality assessment and improvement, *ACM computing surveys (CSUR)*, Vol. 41, n°3, pp. 16. DOI : [10.1145/1541880.1541883](https://doi.org/10.1145/1541880.1541883)
- Bedard, Y., 1986, A Study of the Nature of Data Using a Communication-Based Conceptual Framework of Land Information, *The Canadian Surveyor*, Vol. 40, n°4, pp. 449-460
- Bellino, C., 2013, Contribution de l'architecture de l'information à l'utilisabilité informationnelle: le cas des intranets, *Etudes de communication. langages, information, médiations*. Groupe d'Etudes et de Recherche Interdisciplinaire en Information et Communication de l'Université Lille 3, n°41, pp. 71-88
- Beloin-Saint-Pierre, D., Rugani, B., Lasvaux, S., Mailhac, A., Popovici, E., Sibiude, G., Benetto, E., Schiopu, N., 2016, A review of urban metabolism studies to identify key methodological choices for future harmonization and implementation, *Journal of Cleaner Production*, Vol. 163. DOI : [10.1016/j.jclepro.2016.09.014](https://doi.org/10.1016/j.jclepro.2016.09.014)
- Berners-Lee, T., 2006, Linked Data (mis à jour en 2009). URL : <https://www.w3.org/DesignIssues/LinkedData.html> .
- Bertrand, R., 1986, Pratique de l'analyse statistique des données. Presses de l'Université du Québec, 380 p.
- Besse, P., Laurent, B., 2016, De Statisticien à Data Scientist, *Statistique et Enseignement*. Société Française de Statistique, Vol. 7, n°1, pp. 75-93. URL : <https://hal.archives-ouvertes.fr/hal-01205336> .
- Beyer, A., Debrie, J., 2011, Les temporalités frontalières et urbaines du port de Strasbourg. Analyse géohistorique d'une relation fluviale ville-port, *Métropoles*, n°10. DOI : [10.4000/metropoles.4494](https://doi.org/10.4000/metropoles.4494)
- Beyer, A., Lacoste, R., 2017, La transition écologique des territoires urbano-portuaires. Les enseignements de la démarche de reconnexion énergétique engagée entre la ville et le port de Strasbourg, *Revue d'Economie Régionale Urbaine*. Armand Colin, n°5, pp. 857-880. DOI : [10.3917/rru.175.0857](https://doi.org/10.3917/rru.175.0857)
- Beyer, A., Verhaeghe, L., 2015, Rail – Voie d'eau, une combinaison contre-nature ? La desserte ferroviaire des ports franciliens et ses développements potentiels, *Revue d'Économie Régionale & Urbaine*, n°4, pp. 695-718. DOI : [10.3917/rru.154.0695](https://doi.org/10.3917/rru.154.0695)
- Bihoux, P., 2014, L'âge des low tech : vers une civilisation techniquement soutenable. Éditions du Seuil, 336 p.
- Bouroche, J., Saporta, G., 2006, L'analyse des données. Presses Universitaires de France, « Que sais-je ? », 128 p. URL : <https://www.cairn.info/l-analyse-des-donnees--9782130554448.htm> .

- Boydens, I., 2012, L'océan des données et le canal des normes, *Responsabilité et environnement*. Annales des Mines, Vol. 67, n°3, pp. 22-29. DOI : [10.3917/re.067.0022](https://doi.org/10.3917/re.067.0022)
- Brackstone, G., 1999, La gestion de la qualité des données dans un bureau de statistique, *Techniques d'enquête*, Vol. 25, n°2, pp. 159-171. URL : <https://unstats.un.org/unsd/dnss/docViewer.aspx?docID=397#start> .
- Bringezu, S., Schütz, H., Moll, S., 2003, Rationale for and Interpretation of Economy-Wide Materials Flow Analysis and Derived Indicators, *Journal of Industrial Ecology*. MIT Press, Vol. 7, n°2, pp. 43-64. DOI : [10.1162/108819803322564343](https://doi.org/10.1162/108819803322564343)
- Broto, V., Allen, A., Rapoport, E., 2012, Interdisciplinary Perspectives on Urban Metabolism, *Journal of Industrial Ecology*, Vol. 16, n°6, pp. 851-861. DOI : [10.1111/j.1530-9290.2012.00556.x](https://doi.org/10.1111/j.1530-9290.2012.00556.x)
- Brunetière, J., Alexandre, S., d'Aubreby, M., Debieesse, G., Guérin, A., Perret, B., Schwartz, D., 2013, Le facteur 4 en France : la division par 4 des émissions de gaz à effet de serre à l'horizon 2050. Conseil Général de l'Environnement et du Développement Durable, 215 p. URL : http://www.cgedd.developpement-durable.gouv.fr/IMG/pdf/008378-01_rapport-final_cle0aca84.pdf .
- Brunner, P., Rechberger, H., 2003, Practical Handbook of Material Flow Analysis. CRC Press, 318 p.
- Buclet, N., 2015, Essai d'écologie territoriale : l'exemple d'Aussois en Savoie. CNRS éditions, 216 p.
- Bwele, C., 2013, Le cyberespionnage industriel : de la sécurité informatique à la sécurité économique, *Securite globale*. ESKA, n°2, pp. 49-58
- CCNR, 2019, Basses eaux et impact sur la navigation rhénane - Communiqué de presse. Commission Centrale pour le Navigation sur le Rhin. URL : <https://www.ccr-zkr.org/files/documents/cpresse/cp20191218fr.pdf> .
- Cencic, O., Rechberger, H., 2008, Material flow analysis with software STAN, *EnviroInfo*, pp. 440-447
- Cerceau, J., Junqua, G., Gonzalez, C., Laforest, V., Lopez-Ferber, M., 2014, Quel territoire pour quelle écologie industrielle ? Contribution à la définition du territoire en écologie industrielle, *Développement durable et territoires*, Vol. 5, n°1. DOI : [10.4000/developpementdurable.10179](https://doi.org/10.4000/developpementdurable.10179)
- CGDD, 2014, Comptabilité des flux de matières dans les régions et les départements - Guide méthodologique, *Références du Service de l'observation et des statistiques (SOeS)*. Commissariat Général au Développement Durable. URL : <https://www.statistiques.developpement-durable.gouv.fr/sites/default/files/2018-10/guide-methodologique-references-flux-de-matiere-juin2014.pdf> .
- Charron, M., Shearmur, R., Beauchemin, G., 2018, Données massives et sciences du territoire, *Revue canadienne des sciences régionales*, Vol. 1/3, n°41. URL : <http://www.cjrs-rcsr.org/V41/cjrsrcsr41-2cCharronShearmur.pdf> .

- Chartron, G., 2016, Stratégie, politique et reformulation de l'open access, *Revue française des sciences de l'information et de la communication*, Vol. 8. DOI : [10.4000/rfsic.1836](https://doi.org/10.4000/rfsic.1836)
- Chen, P., 1976, The entity-relationship model - toward a unified view of data, *ACM Transactions on Database Systems (TODS)*, Vol. 1, n°1, pp. 9-36. DOI : [10.1145/320434.320440](https://doi.org/10.1145/320434.320440)
- Chen, C., Härdle, W., Unwin, A., 2008, Handbook of Data Visualization. Springer-Verlag Berlin Heidelberg, 936 p. DOI : [10.1007/978-3-540-33037-0](https://doi.org/10.1007/978-3-540-33037-0)
- Chrysoulakis, N., Castro, E., Moors, E., 2015, Understanding urban metabolism: a tool for urban planning. Routledge, 240 p.
- Ciroth, A., Winter, S., GreenDelta Berlin, 2014, OpenLCA 1.4 overview and first steps, 31 p. URL : http://www.openlca.org/wp-content/uploads/2015/11/openLCA_1-4_overview_and_first_steps_v11.pdf .
- Cleveland, W., 2001, Data Science: an Action Plan for Expanding the Technical Areas of the Field of Statistics, *International Statistical Review*, Vol. 69, n°1, pp. 21-26. DOI : [10.1111/j.1751-5823.2001.tb00477.x](https://doi.org/10.1111/j.1751-5823.2001.tb00477.x)
- Cohen, B., 2004, Urban Growth in Developing Countries: A Review of Current Trends and a Caution Regarding Existing Forecasts, *World Development*, Vol. 32, n°1, pp. 23-51. DOI : [10.1016/j.worlddev.2003.04.008](https://doi.org/10.1016/j.worlddev.2003.04.008)
- Commission européenne, 2017, Communication de la Commission au Parlement Européen, au Conseil, au Comité économique et social européen et au comité des régions relative à la liste 2017 des matières premières critiques pour l'UE. URL : <https://ec.europa.eu/transparency/regdoc/rep/1/2017/FR/COM-2017-490-F1-FR-MAIN-PART-1.PDF> .
- Courtonne, J., 2016, Evaluation environnementale de territoires à travers l'analyse de filières : la comptabilité biophysique pour l'aide à la décision délibérative. Thèse de doctorat en Sciences de gestion. Université Grenoble Alpes. URL : <https://tel.archives-ouvertes.fr/tel-01421664> .
- Dao, Q., 2005, Le rôle des systèmes d'information géographique pour le développement urbain durable, *Enjeux du développement urbain durable*. Da Cunha, A., Knoepfel, P., Leresche, J.-P., Narath, S. Presses Polytechniques Universitaires Romandes, pp. 123-156. URL : <https://archive-ouverte.unige.ch/unige:74870> .
- Dauphiné, A., Provitolo, D., 2007, La résilience : un concept pour la gestion des risques, *Annales de géographie*, Vol. 654, n°2, pp. 115-125
- Davis, C., 2012, Making Sense of Open Data: From Raw Data to Actionable Insight. URL : <https://repository.tudelft.nl/islandora/object/uuid:88c3c6f9-d6a2-4a82-9353-884a3b77b6ed> .
- Davis, C., 2017, Using Linked Data to facilitate translations between product and industry classifications. ISIE-ISSST 2017 Conference: Science in Support of Sustainable and Resilient Communities. URL : <http://programme.exordo.com/isie2017/delegates/presentation/716/> .

- De Boisséson, J., Géhin, B., Jenvrin, F., Ribon, B., 2013, Vers une gestion durable des déchets à la Maison d'arrêt de Strasbourg - Rapport du chantier d'application ECO-Conseil
- De Perthuis, C., Trotignon, R., 2017, Marché européen des quotas de CO2 : Les enjeux du passage à la phase 3. URL : <https://hal.archives-ouvertes.fr/hal-01504990> .
- Debuisson, M., 2014, Les modes d'interaction pour une dynamique territoriale soutenable : un apport à l'écologie territoriale. Thèse de doctorat en Sciences de l'Homme et Société, Université de Technologie de Troyes. URL : <https://halshs.archives-ouvertes.fr/tel-01162416> .
- Delahaye, J., 2015, Les blockchains, clefs d'un nouveau monde, *Pour la Science*, Vol. 449, pp. 80-85. URL : <https://www.pourlascience.fr/sd/informatique/les-blockchains-clefs-daposun-nouveau-monde-8354.php> .
- Delzangles, H., Fleury, M., Monnier, L., 2018, Droit à l'information environnementale vs. secret industriel et commercial : une conciliation à l'épreuve en droit nucléaire, *Revue juridique de l'environnement*, Vol. 43, n°4, pp. 675-692. URL : <https://www.cairn.info/revue-revue-juridique-de-l-environnement-2018-4-page-675.htm>.
- Desrosières, A., 2008, L'Argument statistique I : Pour une sociologie historique de la quantification. Presses des MINES, 329 p. DOI : [10.4000/books.pressesmines.901](https://doi.org/10.4000/books.pressesmines.901)
- Dessaigne, N., 2005, Le modèle DOAN (DOcument ANnotation Model) Modélisation de l'information complexe appliquée à la plateforme Arisem Kaliwatch Server. Thèse de Doctorat en Informatique. Université de Nantes. URL : <https://tel.archives-ouvertes.fr/tel-00465962> .
- Diguet, C., Lopez, F., Lefèvre, L., 2019, L'impact spatial et énergétique des data centers sur les territoires, *Projet ENERNUM*. ADEME, Direction Villes et territoires durables, 141 p.
- DILA, 2011, Guide des bonnes pratiques d'hygiène de la distribution de produits alimentaires par les organismes caritatifs, *Direction de l'information légale et administrative. Les éditions des Journaux officiels*. URL : http://agriculture.gouv.fr/sites/minagri/files/documents/pdf/gph_20115943_0001_p000_cle0e8e3f.pdf .
- Dusek, T., 2005, The modifiable areal unit problem in regional economics, n°Ersa05p357. URL : <http://www-sre.wu.ac.at/ersa/ersaconfs/ersa05/papers/357.pdf> .
- Duvigneaud, P., Denaeyer-De Smet, S., 1975, L'écosystème Bruxelles, *L'Ecosystème urbain. Application à l'agglomération bruxelloise*. Colloque international, organisé par l'Agglomération de Bruxelles, 1974, pp. 45-47
- EcoInfo, 2014, L'énergie des métaux. URL : <https://ecoinfo.cnrs.fr/2014/09/03/2-lenergie-des-metaux/> .
- Elioth, 2014, Métabolisme Urbain de Paris. URL : <http://metabolisme.paris.fr/> .
- Eurostat, 2001, Economy-wide material flow accounts and derived indicators: a methodological guide, *Eurostat Theme 2, Economy and finance*. European Commission. URL : <https://ec.europa.eu/eurostat/documents/1798247/6191533/3-Economy-wide-material-flow-accounts...-A-methodological-guide-2001-edition.pdf> .

- Eurostat, 2013a, Economy-wide Material Flow Accounts (EW-MFA) - Compilation guide. URL : <http://ec.europa.eu/eurostat/documents/1798247/6191533/2013-EW-MFA-Guide-10Sep2013.pdf/54087dfb-1fb0-40f2-b1e4-64ed22ae3f4c> .
- Eurostat, 2013b, Système européen des comptes: SEC 2010. Publications Office. URL : <http://bookshop.europa.eu/uri?target=EUB:NOTICE:KS0213269:FR:HTML> .
- Faist Emmenegger, M., Frischknecht, R., Cornaglia, L., Rubli, S., 2003, Métabolisme des activités économiques du canton de Genève - Phase 1. URL : <https://www.genie.ch/library/h/metabolisme-des-activites-economiques-du-canton-de-geneve---phase-1.html> .
- Ferrão, P., Fernández, J., 2013, Sustainable urban metabolism. MIT Press, Cambridge, Mass., 244 p. DOI : [10.7551/mitpress/8617.001.0001](https://doi.org/10.7551/mitpress/8617.001.0001)
- Figuères, C., Guyomard, H., Rotillon, G., 2007, Une brève analyse économique orthodoxe du concept de développement durable, *Économie rurale. Agricultures, alimentations, territoires*, n°300, pp. 79-84. DOI : [10.4000/economierurale.2228](https://doi.org/10.4000/economierurale.2228)
- Fléty, Y., De Sède-Marceau, M., 2009, Vers une géo-ontologie pour les Systèmes Énergétiques Territoriaux (SET), *XVIe rencontres de Rochebrune sur les systèmes complexes naturels et artificiels : ontologie et dynamique des systèmes complexes*, pp. 1-12. URL : <https://hal.archives-ouvertes.fr/hal-00767229> .
- Flichy, P., 2013, Rendre visible l'information, *Réseaux*, n°178-179, pp. 55-89. URL : <http://www.cairn.info/revue-reseaux-2013-2-page-55.htm> .
- Flipo, F., Deltour, F., Dobré, M., 2016, Les technologies de l'information à l'épreuve du développement durable, *Natures Sciences Sociétés*, Vol. 24, n°1, pp. 36-47. DOI : [10.1051/nss/2016007](https://doi.org/10.1051/nss/2016007)
- Fonseca, F., Egenhofer, M., Davis, C., Câmara, G., 2002, Semantic granularity in ontology-driven geographic information systems, *Annals of mathematics and artificial intelligence*. Springer, Vol. 36, n°1-2, pp. 121-151. DOI : [10.1023/a:1015808104769](https://doi.org/10.1023/a:1015808104769)
- Francois-Lecompte, A., Gentric, M., Audigier, N., 2013, L'affichage environnemental, une nouvelle façon pour l'entreprise de rendre des comptes ?, *Revue française de gestion*, n°237, pp. 181-199. URL : http://www.cairn.info/resume.php?ID_ARTICLE=RFG_237_0181 .
- Gaeremynck, J., 2009, Le comité du secret statistique, *Courrier des statistiques*, Vol. 128, n°3, pp. 15-18. URL : <https://www.autorite-statistique-publique.fr/wp-content/uploads/2018/09/Le-comité-du-secret-statistique-Jean-Gaeremynck.pdf> .
- Garric, A., 2012, Le 7e continent de plastique : ces tourbillons de déchets dans les océans, *Le Monde*. URL : http://www.lemonde.fr/planete/article/2012/05/09/le-7e-continent-de-plastique-ces-tourbillons-de-dechets-dans-les-occeans_1696072_3244.html .
- Georgeault, L., 2015, Le potentiel d'écologie industrielle en France : approche territoriale et éléments de réalisation. Thèse de doctorat en Géographie-Aménagement. Université Paris 1. URL : <http://www.theses.fr/s91113> .

- Gérard, J., 2017, Incinérateur amianté : un diagnostic en avait trouvé dès 2009, *Rue89 Strasbourg*. URL : <https://www.rue89strasbourg.com/incinerateur-amianteurometropole-etait-elle-au-courant-116718> .
- Giljum, S., Hubacek, K., 2009, Conceptual Foundations and Applications of Physical Input-Output Tables, *Handbook of Input-Output Economics in Industrial Ecology, Eco-Efficiency in Industry and Science*. Suh, Sangwon. Springer Netherlands, n°23, pp. 61-75. DOI : [10.1007/978-1-4020-5737-3_4](https://doi.org/10.1007/978-1-4020-5737-3_4)
- Goldstein, B., Birkved, M., Quitzau, M., Hauschild, M., 2013, Quantification of urban metabolism through coupling with the life cycle assessment framework: concept development and case study, *Environmental Research Letters*, Vol. 8, n°3, pp. 035024. DOI : [10.1088/1748-9326/8/3/035024](https://doi.org/10.1088/1748-9326/8/3/035024)
- Gruber, T., 1993, A translation approach to portable ontology specifications, *Knowledge acquisition*. Elsevier, Vol. 5, n°2, pp. 199-220. DOI : [10.1006/knac.1993.1008](https://doi.org/10.1006/knac.1993.1008)
- Harpert, C., Gully, E., 2013, Ecologie industrielle et territoriale : quels outils d'aide à la décision ? De l'analyse des flux à l'approche intégrée, *Déchets, Sciences et Techniques*, n°63, pp. 45-55. DOI : [10.4267/dechets-sciences-techniques.2580](https://doi.org/10.4267/dechets-sciences-techniques.2580)
- Haynes, W., Lide, D., Bruno, T., 2017, CRC handbook of chemistry and physics: a ready-reference book of chemical and physical data. CRC Press, 2652 p.
- Hazas, M., Morley, J., Bates, O., Friday, A., 2016, Are There Limits to Growth in Data Traffic?: On Time Use, Data Generation and Speed, *Proceedings of the Second Workshop on Computing Within Limits, LIMITS '16*. ACM, pp. 14:1-14:5. DOI : [10.1145/2926676.2926690](https://doi.org/10.1145/2926676.2926690)
- Hendriks, C., Obernosterer, R., Müller, D., Kytzia, S., Baccini, P., Brunner, P., 2000, Material Flow Analysis: A tool to support environmental policy decision making. Case-studies on the city of Vienna and the Swiss lowlands, *Local Environment*, Vol. 5, n°3, pp. 311-328. DOI : [10.1080/13549830050134257](https://doi.org/10.1080/13549830050134257)
- Herbelin, A., 2018, Modèle:Ecologie territoriale et trajectoires de transitions : le cas du Rhône-Médian. Thèse de doctorat en Aménagement de l'espace, Urbanisme. Université Grenoble Alpes. URL : <https://tel.archives-ouvertes.fr/tel-02276808> .
- Herrenschmidt, C., 2007, Les trois écritures. Langue, nombre, code, *Bibliothèque des Sciences humaines*. Gallimard, 528 p.
- Hey, T., Tansley, S., Tolle, K., 2009, The Fourth Paradigm: Data-Intensive Scientific Discovery, *The Fourth Paradigm: Data-Intensive Scientific Discovery*, 284 p.
- Hoekstra, A., Chapagain, A., 2007, Water footprints of nations: Water use by people as a function of their consumption pattern, *Water Resources Management*, Vol. 21, n°1, pp. 35-48. DOI : [10.1007/s11269-006-9039-x](https://doi.org/10.1007/s11269-006-9039-x)

- Hoornweg, D., Campillo, G., Saldivar-Sali, A., Sugar, L., Linders, D., 2012, Mainstreaming Urban Metabolism: Advances and Challenges in City Participation, *Sixth Urban Research and Knowledge Symposium 2012*. URL : http://www.dennislinders.com/portfolio/Hoornweg_et_al_URKS6_Urban_Metabolism.pdf .
- Humbert, S., 2009, Geographically Differentiated Life-cycle Impact Assessment of Human Health. UC Berkeley Electronic Theses and Dissertations. URL : <https://escholarship.org/uc/item/1xv927gy> .
- Ibekwe-Sanjuan, F., 2014, Big Data, Big machines, Big Science : vers une société sans sujet et sans causalité ?, *XIXème Congrès de la Sfsic. Penser les techniques et les technologies : Apports des Sciences de l'Information et de la Communication et perspectives de recherches.*, pp. 1-10. URL : <https://hal.archives-ouvertes.fr/hal-01066202> .
- Idda, C., 2014, Caractéristiques biologiques spatialisées et influence des stratégies individuelles dans la gestion des ressources halieutiques : une approche par les jeux différentiels. Thèse de doctorat en Sciences économiques, Université de Corte. URL : <http://www.theses.fr/2014CORT0004> .
- Inmon, W., 2005, Building the Data Warehouse. Wiley, 576 p. URL : [http://homes.dcc.ufba.br/~mauricio052/Material_Artigo/Building_The_Data_Warehouse_\(2005\)_Fourth_Edition-Inmon-Wiley.pdf](http://homes.dcc.ufba.br/~mauricio052/Material_Artigo/Building_The_Data_Warehouse_(2005)_Fourth_Edition-Inmon-Wiley.pdf) .
- Ioppolo, G., Heijungs, R., Cucurachi, S., Salomone, R., Kleijn, R., 2014, Urban Metabolism: Many Open Questions for Future Answers, *Pathways to Environmental Sustainability, Pathways to Environmental Sustainability*. Salomone, Roberta and Saija, Giuseppe. Springer International Publishing, pp. 23-32. DOI : [10.1007/978-3-319-03826-1_3](https://doi.org/10.1007/978-3-319-03826-1_3)
- IPCC, 2014, Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, *IPCC 5th Assessment Synthesis Report*. Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.), Geneva, Switzerland. Intergovernmental Panel on Climate Change, 151 p. URL : http://ar5-syr.ipcc.ch/ipcc/ipcc/resources/pdf/IPCC_SynthesisReport.pdf .
- Ismayilov, A., Kontokostas, D., Auer, S., Lehmann, J., Hellmann, S., 2018, Wikidata through the Eyes of DBpedia, *Semantic Web*. IOS Press, Vol. 9, n°4, pp. 493-503. DOI : [10.3233/sw-170277](https://doi.org/10.3233/sw-170277)
- ISO, 2006, ISO 14040:2006 - Environmental management - Life cycle assessment - Principles and framework. Organisation internationale de normalisation, 23 p. URL : http://www.iso.org/iso/catalogue_detail?csnumber=37456 .
- Janin, M., 2000, Démarche d'éco-conception en entreprise. Un enjeu : construire la cohérence entre outils et processus. Thèse de doctorat en Sciences de l'ingénieur. Arts et Métiers ParisTech. URL : <https://pastel.archives-ouvertes.fr/tel-00005632> .
- Kahn, R., 2015, Le développement durable et les territoires : conceptions, impacts et perspectives - Parties 1 et 2, *La Lettre du financier territorial*, n°302 & 303
- Kennedy, C., Cuddihy, J., Engel-Yan, J., 2007, The Changing Metabolism of Cities, *Journal of Industrial Ecology*, Vol. 11, n°2, pp. 43-59. DOI : [10.1162/jie.2007.1107](https://doi.org/10.1162/jie.2007.1107)

- Kennedy, C., Pincetl, S., Bunje, P., 2011, The study of urban metabolism and its applications to urban planning and design, *Environmental Pollution*, Vol. 159, n°8–9, pp. 1965-1973. DOI : [10.1016/j.envpol.2010.10.022](https://doi.org/10.1016/j.envpol.2010.10.022)
- Kennedy, C., Hoornweg, D., 2012, Mainstreaming Urban Metabolism, *Journal of Industrial Ecology*, Vol. 16, n°6, pp. 780-782. DOI : [10.1111/j.1530-9290.2012.00548.x](https://doi.org/10.1111/j.1530-9290.2012.00548.x)
- Kennedy, C., Stewart, I., Ibrahim, N., Facchini, A., Mele, R., 2014, Developing a multi-layered indicator set for urban metabolism studies in megacities, *Ecological Indicators*, Vol. 47, pp. 7-15. DOI : [10.1016/j.ecolind.2014.07.039](https://doi.org/10.1016/j.ecolind.2014.07.039)
- Kitchin, R., Tate, N., 2000, *Conducting Research in Human Geography: Theory, Methodology and Practice*. Routledge, London, 344 p.
- Kitzes, J., Wackernagel, M., Loh, J., Peller, A., Goldfinger, S., Cheng, D., Tea, K., 2008, Shrink and share: humanity's present and future Ecological Footprint, *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, Vol. 363, n°1491, pp. 467-475. DOI : [10.1098/rstb.2007.2164](https://doi.org/10.1098/rstb.2007.2164)
- Kleijn, R., 2000, Adding It All Up The Sense and Non-Sense of Bulk-MFA, *Journal of Industrial Ecology*, Vol. 4, n°2, pp. 7-8. DOI : [10.1162/108819800569762](https://doi.org/10.1162/108819800569762)
- Kozderka, M., 2016, Parametric LCA approaches for efficient design. Thèse de doctorat en Sciences de l'ingénieur. Université de Strasbourg et Technical University of chemistry and technology (Prague). URL : <http://www.theses.fr/2016STRAD050> .
- Kuhn, T., 1959, Energy conservation as an example of simultaneous discovery, *Critical problems in the history of science*. The university of Wisconsin Press, pp. 321-356
- Lagain, A., 2016, La station d'épuration déborde, la Nesque encore polluée à Saint-Didier, *France Bleu Vaucluse*. URL : <https://www.francebleu.fr/infos/climat-environnement/la-station-d-epuration-deborde-la-nesque-encore-polluee-st-didier-1474602330> .
- Laganier, R., Villalba, B., Zuindeau, B., 2002, Le développement durable face au territoire : éléments pour une recherche pluridisciplinaire, *Développement durable et territoires*, n°1 (Dossier). DOI : [10.4000/developpementdurable.774](https://doi.org/10.4000/developpementdurable.774)
- Landner, L., Lindeström, L., 1998, Zinc in society and in the environment: an account of the facts on fluxes, amounts and effects of zinc in Sweden. Swedish Environmental Research Group. Kil, Sweden, 160 p.
- Laner, D., Rechberger, H., 2016, Material Flow Analysis, *Special Types of Life Cycle Assessment*. LCA Compendium – The Complete World of Life Cycle Assessment. Springer Netherlands, pp. 293-332. DOI : [10.1007/978-94-017-7610-3_7](https://doi.org/10.1007/978-94-017-7610-3_7)
- Laubard, B., Lissorgues, G., 2010, La logistique urbaine, fonction vitale pour la métropole parisienne. Constats, pistes d'actions, préconisations, *Rapport du CCIP*. Chambre de Commerce et d'Industrie de Paris, 65 p. URL : <https://www.cciparisidf.com/sites/default/files/etudes/pdf/documents/logistique-urbaine-lau1002.pdf> .
- Lemoine, P., 2018, La malédiction des données, *Esprit*. Editions Esprit, n°6, pp. 131-138. DOI : [10.3917/espri.1806.0131](https://doi.org/10.3917/espri.1806.0131)

- Le Noé, J., Billen, G., Lassaletta, L., Silvestre, M., Garnier, J., 2016, La place du transport de denrées agricoles dans le cycle biogéochimique de l'azote en France : un aspect de la spécialisation des territoires, *Cahiers Agricultures*, Vol. 25, n°1, pp. 15004. DOI : [10.1051/cagri/2016002](https://doi.org/10.1051/cagri/2016002)
- Loiseau, E., Junqua, G., Roux, P., Bellon-Maurel, V., 2012a, Environmental assessment of a territory: An overview of existing tools and methods, *Journal of Environmental Management*, Vol. 112, pp. 213 - 225. DOI : [10.1016/j.jenvman.2012.07.024](https://doi.org/10.1016/j.jenvman.2012.07.024)
- Loiseau, E., Junqua, G., Roux, P., Maurel, P., Bellon-Maurel, V., 2012b, Évaluation environnementale de territoires: apports, limites et adaptation du cadre méthodologique de l'Analyse du Cycle de Vie. Conférence Interdisciplinaire sur l'Écologie Industrielle et Territoriale, Oct 2012, Troyes, France. URL : <https://hal.archives-ouvertes.fr/hal-00811194/> .
- Mabi, C., 2015, La plate-forme "data.gouv.fr" ou l'open data à la française, *Informations sociales*, Vol. 5, n°191, pp. 52-59. DOI : [10.3917/inso.191.0052](https://doi.org/10.3917/inso.191.0052).
- Maillefert, M., Robert, I., 2014, Écologie industrielle, économie de la fonctionnalité, entreprises et territoires : vers de nouveaux modèles productifs et organisationnels ?, *Développement durable et territoires. Économie, géographie, politique, droit, sociologie*, Vol. 5, n°1. DOI : [10.4000/developpementdurable.10177](https://doi.org/10.4000/developpementdurable.10177)
- Mallowan, M., 2012, Intelligence et transculture de l'information, *Communication et organisation*, Vol. 42, pp. 27-48. DOI : [10.4000/communicationorganisation.3832](https://doi.org/10.4000/communicationorganisation.3832)
- Mardellat, P., 2010, Qualité de vie et consommation soutenable : une perspective pratique, *Développement durable et territoires*, Vol. 1, n°3. DOI : [10.4000/developpementdurable.8429](https://doi.org/10.4000/developpementdurable.8429)
- Mazaud, L., Suaud, C., Marquet, S., de Sède Marceau, M., Noucher, M., Hainaut, H., Bailly, B., 2017, Planification territoriale durable - Opportunités offertes par les données et les outils, *Revue Internationale de Géomatique*. Lavoisier, Vol. 27, n°1, pp. 11-36
- Meadows, D., Meadows, D., Randers, J., Behrens III, W., 1972, The limits to growth: a report for the Club of Rome's Project on the Predicament of Mankind. Club of Rome and Potomac Associates. Universe Books. URL : <https://www.clubofrome.org/report/the-limits-to-growth/> .
- MEDDTL, 2012, Chiffres clés du transport - Edition 2012, *Repères*. Ministère de l'Écologie, du Développement durable, des Transports et du Logement, 32 p. URL : http://temis.documentation.developpement-durable.gouv.fr/pj/17636/17636_2012.pdf .
- Mehlhorn, K., Sanders, P., 2008, Algorithms and Data Structures - The Basic Toolbox. Springer, 300 p. DOI : [10.1007/978-3-540-77978-0](https://doi.org/10.1007/978-3-540-77978-0)
- Metabolism of Cities, 2016, Data Visualizations in Urban Metabolism Research. URL : <https://metabolismofcities.org/datavisualization> .
- Minaud, G., 2014, Comptables et comptabilités dans l'Antiquité : conclusion, *Comptabilités*, Vol. 6. URL : <http://journals.openedition.org/comptabilites/1539> .

- Montero, L., 2013, L'analyse du métabolisme urbain : intérêt et limites pour l'évaluation de la durabilité de l'approvisionnement alimentaire d'un territoire ? Le cas de Montpellier agglomération. URL : https://web.supagro.inra.fr/pmb/opac_css/doc_num.php?explnum_id=1612 .
- Morris, A., 2016, L'analyse de flux de matières au Québec : Méthodes et enjeux d'opérationnalisation dans une perspective d'économie circulaire, 89 p. URL : http://savoirs.usherbrooke.ca/bitstream/handle/11143/8173/Morris_Audrey_MEnv_2016.pdf?sequence=1 .
- MTES, 2018, Bilan du dispositif de collecte et de diffusion des données locales d'énergie « article 179 ». Ministère de la Transition Écologique et Solidaire, 43 p. URL : <https://www.statistiques.developpement-durable.gouv.fr/sites/default/files/2018-12/bilan-donnees-locales-energie-article-179-decembre2018.pdf> .
- Muller, J., 2018, *The Tyranny of Metrics*. Princeton University Press, 248 p.
- Myers, R., Fishman, T., Reck, B., Graedel, T., 2019, Unified materials information system (UMIS): an integrated material stocks and flows data structure, *Journal of Industrial Ecology*. Wiley Online Library, Vol. 23, n°1, pp. 222-240. DOI : [10.1111/jiec.12730](https://doi.org/10.1111/jiec.12730)
- Nations Unies, Commission Européenne, Eurostat, Organisation des Nations Unies pour l'alimentation et l'agriculture, Fonds monétaire international, Organisation de coopération et de développement économiques, Banque mondiale, 2012, Cadre central du système de comptabilité économique et environnementale 2012. UNO, New York. UNO, 377 p. URL : https://unstats.un.org/unsd/envaccounting/seeaRev/CF_trans/SEEA_CF_Final_fr.pdf .
- Nikitin, M., 1992, La naissance de la comptabilité industrielle en France. Thèse de doctorat en Gestion et management. Université Paris Dauphine - Paris IX. URL : https://tel.archives-ouvertes.fr/tel-00649637/file/ThA_se_Nikitin.pdf .
- Noy, N., McGuinness, D., 2001, *Ontology Development 101: A Guide to Creating Your First Ontology*. URL : <http://www-ksl.stanford.edu/people/dlm/papers/ontology-tutorial-noy-mcguinness.pdf> .
- Odum, H., 2002, Emergy accounting, *Unveiling Wealth*. Bartelmus, Peter. Springer Netherlands, pp. 135-146. DOI : [10.1007/0-306-48221-5_13](https://doi.org/10.1007/0-306-48221-5_13)
- OECD, 2008, *Measuring material flows and resource productivity - The OECD Guide*, 164 p. URL : <https://www.oecd.org/environment/indicators-modelling-outlooks/MFA-Guide.pdf> .
- O'Neil, E., 2008, Object/Relational Mapping 2008: Hibernate and the Entity Data Model (Edm), *Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data, SIGMOD '08*. ACM, pp. 1351-1356. DOI : [10.1145/1376616.1376773](https://doi.org/10.1145/1376616.1376773)
- O'Neil, C., 2016, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, Crown, 272 p.

- Openshaw, S., Taylor, P., 1979, A million or so correlation coefficients: three experiments on the modifiable areal unit problem, *Statistical Applications in the Spatial Sciences*. Pion, pp. 127-144.
- Pauliuk, S., Heeren, N., Hasan, M., Müller, D., 2019, A general data model for socioeconomic metabolism and its implementation in an industrial ecology data commons prototype, *Journal of Industrial Ecology*, Vol. 23, n°5, pp. 1016-1027. DOI : [10.1111/jiec.12890](https://doi.org/10.1111/jiec.12890)
- Paquiénéguy, F., Dymytrava, V., 2018, Open data et métropoles, les enjeux d'une transformation à l'oeuvre, *Questions de communication*, Vol. 2, n°34, pp. 209-228. URL : <https://www.cairn.info/revue-questions-de-communication-2018-2-page-209.htm> .
- PAS, 2018, Transport multimodal - Navette Saverne. Port Autonome de Strasbourg. URL : <https://www.strasbourg.port.fr/wp-content/uploads/2018/11/Fiche-PAS-navette-saverne-vd1.pdf> .
- Pasquier, J., 2010, La comptabilité environnementale au niveau national. URL : <http://encyclopedie-dd.org/encyclopedie/neige-neige-economie-neige-neige/la-comptabilite-environnementale.html> .
- Pavel, I., 2017, La blockchain - Les défis de son implémentation, *Annales des Mines - Réalités industrielles*, n°3, pp. 20-24. DOI : [10.3917/rindu1.173.0020](https://doi.org/10.3917/rindu1.173.0020)
- Peyraud, J., Richard, G., Gascuel-Oudou, C., 2015, Boucler les grands cycles biogéochimiques, *Innovations Agronomiques*, Vol. 43, pp. 177-186. URL : <http://www6.inra.fr/ciag/content/download/5606/42546/file/Vol43-15-Peyraud.pdf> .
- Pierra, G., Dehainsala, H., Aneur, Y., Bellatreche, L., 2005, Bases de données à base ontologique. Principe et mise en oeuvre, *Ingénierie des Systèmes d'Information*, Vol. 10, n°2, pp. 91-115. URL : <http://lias-lab.isae-ensma.fr/publications/7198/2005-I-Pierra.pdf>.
- Pincetl, S., Graham, R., Murphy, S., Sivaraman, D., 2016, Analysis of High-Resolution Utility Data for Understanding Energy Use in Urban Systems: The Case of Los Angeles, California, *Journal of Industrial Ecology*, Vol. 20, n°1, pp. 166–178. DOI : [10.1111/jiec.12299](https://doi.org/10.1111/jiec.12299)
- Pitarch, Y., Favre, C., Laurent, A., Poncelet, P., 2010, Analyse flexible dans les entrepôts de données : quand les contextes s'en mêlent, *EDA*, Vol. 10, pp. 191-205. URL : <https://www.lirmm.fr/~poncelet/publications/papers/eda2010.pdf> .
- Polanco, X., 2008, Transformer l'information en connaissance avec stanalyst. Cadre conceptuel et modèle, *Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação*, Vol. 13, n°1, pp. 76-91. DOI : [10.5007/1518-2924.2008v13nesp1p76](https://doi.org/10.5007/1518-2924.2008v13nesp1p76)
- Press, G., 2013, A Very Short History Of Data Science, *Forbes*. URL : <https://www.forbes.com/sites/gilpress/2013/05/28/a-very-short-history-of-data-science> .
- Rahm, E., Do, H., 2000, Data cleaning: Problems and current approaches, *IEEE Data Engineering Bulletin*, Vol. 23, n°4, pp. 3-13. URL : <https://dbs.uni-leipzig.de/file/TBDE2000.pdf> .

- Région Alsace, Ademe, 2015, Consommations, besoins et richesses du territoire alsacien. Document complété par un entretien avec la mission régionale d'économie circulaire le 27/10/2016, voir Annexe 1
- Ribon, B., 2013, Réalisation du bilan carbone du patrimoine bâti et étude des consommations en eau et énergie des piscines publiques - Ville et Communauté Urbaine de Strasbourg. Mémoire / thèse professionnelle pour l'obtention du mastère spécialisé 'Éco-conseiller'
- Ribon, B., 2017, Territorial Metabolism through the Life Cycle of Products, *LCM 2017 Conference Cycle (Luxembourg)*. URL : <https://metabolisme-territorial.fr/wiki/index.php/LCM2017> .
- Rodhain, A., Rodhain, F., Fallery, B., Galy, J., 2017, TIC et/ou développement durable : le paradoxe écologique vécu par les utilisateurs, *Annales des Mines - Gérer et comprendre*, Vol. 2, n°128, pp. 48-61. DOI : [10.3917/geco1.128.0048](https://doi.org/10.3917/geco1.128.0048)
- Rosado, L., Niza, S., Ferrão, P., 2014, A Material Flow Accounting Case Study of the Lisbon Metropolitan Area using the Urban Metabolism Analyst Model, *Journal of Industrial Ecology*, Vol. 18, n°1, pp. 84-101. DOI : [10.1111/jiec.12083](https://doi.org/10.1111/jiec.12083)
- Rose, C., Parker, A., Jefferson, B., Cartmell, E., 2015, The Characterization of Feces and Urine: A Review of the Literature to Inform Advanced Treatment Technology, *Critical Reviews in Environmental Science and Technology*. Taylor & Francis, Vol. 45, n°17, pp. 1827-1879. DOI : [10.1080/10643389.2014.1000761](https://doi.org/10.1080/10643389.2014.1000761)
- Rouvreau, L., Michel, P., Monfort, D., Jayr, E., Morice, J., 2013, L'analyse systémique du métabolisme territorial, un outil pour favoriser le recours aux ressources secondaires dans le domaine du BTP. URL : <https://hal-brgm.archives-ouvertes.fr/hal-00788373/document>
- Saporta, G., 2011, Probabilités, analyse des données et statistique. Éd. Technip, Paris, 656 p.
- Schutt, R., O'Neil, C., 2013, Doing data science. O'Reilly Media, 406 p.
- Seppälä, J., Posch, M., Johansson, M., Hettelingh, J., 2006, Country-dependent Characterisation Factors for Acidification and Terrestrial Eutrophication Based on Accumulated Exceedance as an Impact Category Indicator (14 pp), *The International Journal of Life Cycle Assessment*, Vol. 11, n°6, pp. 403-416. DOI : [10.1065/lca2005.06.215](https://doi.org/10.1065/lca2005.06.215)
- Servigne, P., Stevens, R., 2015, Comment tout peut s'effondrer: petit manuel de collapsologie à l'usage des générations présentes. Editions du Seuil, Collection Anthropocène, 304 p.
- Shahrokni, H., Lazarevic, D., Brandt, N., 2015, Smart Urban Metabolism: Towards a Real-Time Understanding of the Energy and Material Flows of a City and Its Citizens, *Journal of Urban Technology*. Routledge, Vol. 22, n°1, pp. 65-86. DOI : [10.1080/10630732.2014.954899](https://doi.org/10.1080/10630732.2014.954899)
- Shao, M., Tang, X., Zhang, Y., Li, W., 2006, City clusters in China: air and surface water pollution, *Frontiers in Ecology and the Environment*, Vol. 4, n°7, pp. 353-361. DOI : [10.1890/1540-9295\(2006\)004\[0353:CCICAA\]2.0.CO;2](https://doi.org/10.1890/1540-9295(2006)004[0353:CCICAA]2.0.CO;2)

- Silvestre, M., Billen, G., Garnier, J., 2012, Évaluation de la provenance des marchandises consommées par un territoire : AmstraM, une application de webmapping basée sur les statistiques de transport et de production, *Écologie industrielle et territoriale: COLEIT*, pp. 361-370
- Speich, R., 2016, Lien ville-agriculture : L’empreinte écologique de l’Eurométropole, *Rapport de Stage Étude et Recherche. École des Ingénieurs de la Ville de Paris, Université de Strasbourg*
- Spyns, P., Meersman, R., Jarrar, M., 2002, Data Modelling Versus Ontology Engineering, *SIGMOD Rec.* ACM, Vol. 31, n°4, pp. 12-17. DOI : [10.1145/637411.637413](https://doi.org/10.1145/637411.637413)
- Steffen, W., Richardson, K., Rockström, J., Cornell, S., Fetzer, I., Bennett, E., Biggs, R., Carpenter, S., de Vries, W., de Wit, C., Folke, C., et autres, 2015, Planetary boundaries: Guiding human development on a changing planet, *Science*. American Association for the Advancement of Science, Vol. 347, n°6223. DOI : [10.1126/science.1259855](https://doi.org/10.1126/science.1259855)
- Strathern, M., 1997, 'Improving ratings': audit in the British University system, *European review*. Cambridge University Press, Vol. 5, n°3, pp. 305-321. URL : <https://www.cambridge.org/core/journals/european-review/article/improving-ratings-audit-in-the-british-university-system/FC2EE640C0C44E3DB87C29FB666E9AAB> .
- Sumaray, A., Makki, S., 2012, A Comparison of Data Serialization Formats for Optimal Efficiency on a Mobile Platform, *Proceedings of the 6th International Conference on Ubiquitous Information Management and Communication, ICUIMC '12*. ACM, pp. 48:1-48:6. DOI : [10.1145/2184751.2184810](https://doi.org/10.1145/2184751.2184810)
- Thessen, A., Patterson, D., 2011, Data issues in the life sciences, *ZooKeys*, Vol. 150, pp. 15-51. DOI : [10.3897/zookeys.150.1766](https://doi.org/10.3897/zookeys.150.1766)
- Thévenot, J., 2001, Évolution des perspectives de la conception des SI : analyse longitudinale des enjeux et des réponses méthodologiques, *Systèmes d'Information et Management*, Vol. 6, n°2, pp. 117-132. DOI : [10.9876/sim.v6i2.103](https://doi.org/10.9876/sim.v6i2.103)
- Thompson, C., 2013, Relying on Algorithms and Bots Can Be Really, Really Dangerous, *Wired*. URL : <https://www.wired.com/2013/03/clive-thompson-2104/> .
- Thomsen, C., Bach Pedersen, T., 2009, Pygrametl: A Powerful Programming Framework for Extract-transform-load Programmers, *Proceedings of the ACM Twelfth International Workshop on Data Warehousing and OLAP, DOLAP '09*. ACM, pp. 49-56. DOI : [10.1145/1651291.1651301](https://doi.org/10.1145/1651291.1651301)
- Toubin, M., Lhomme, S., Diab, Y., Serre, D., Laganier, R., 2012, La Résilience urbaine : un nouveau concept opérationnel vecteur de durabilité urbaine ?, *Développement durable et territoires. Économie, géographie, politique, droit, sociologie*, Vol. 3, n°1. DOI : [10.4000/developpementdurable.9208](https://doi.org/10.4000/developpementdurable.9208)

- UNEP, 2013, City-level decoupling: urban resource flows and the governance of infrastructure transitions : summary for policy makers. United Nations Environment Programme. À Report of the Working Group on Cities of the International Resource Panel. URL : <https://www.wrforum.org/uneppublicationspdf/city-level-decoupling-urban-resource-flows-and-the-governance-of-infrastructure-transitions/> .
- Valentin, A., Lancry, A., Lemarchand, C., 2010, La construction des échantillons dans la conception ergonomique de produits logiciels pour le grand public. Quel quantitatif pour les études qualitatives ?, *Le travail humain*. Presses Universitaires de France, Vol. 73, n°3, pp. 261-290. DOI : [10.3917/th.733.0261](https://doi.org/10.3917/th.733.0261)
- Van der Voet, E., van Oers, L., Nikolic, I., 2004, Dematerialization: Not Just a Matter of Weight, *Journal of Industrial Ecology*. MIT Press, Vol. 8, n°4, pp. 121-137. DOI : [10.1162/1088198043630432](https://doi.org/10.1162/1088198043630432)
- Veyret, Y., Ciattoni, A., 2011, Géo-environnement. Armand Colin, Paris, 253 p.
- VNF, 2019, Accélération du développement de la logistique urbaine durable à Strasbourg, *VNF et vous. La logistique, l'eau, les territoires*. URL : <https://vnfetvous.fr/2019/12/07/acceleration-du-developpement-de-la-logistique-urbaine-durable-a-strasbourg/> .
- Wackernagel, M., Beyers, B., 2019, Ecological Footprint - Managing Our Biocapacity Budget. New Society Publishers, 288 p.
- Waelbroeck, P., 2017, Les enjeux économiques de la blockchain, *Annales des Mines - Réalités industrielles*. Annales des Mines - Réalités industrielles, n°3, pp. 10-19. DOI : [10.3917/rindu1.173.0010](https://doi.org/10.3917/rindu1.173.0010)
- Wilkinson, M., Dumontier, M., Aalbersberg, I., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J., da Silva Santos, L., Bourne, P., autres, 2016, The FAIR Guiding Principles for scientific data management and stewardship, *Scientific data*. Nature Publishing Group, Vol. 3. DOI : [10.1038/sdata.2016.18](https://doi.org/10.1038/sdata.2016.18)
- Wolman, A., 1965, The Metabolism of Cities, *Scientific American*, Vol. 213, pp. 179-190. URL : <http://www.irows.ucr.edu/cd/courses/10/wolman.pdf> .
- Zhang, Y., 2013, Urban metabolism: A review of research methodologies, *Environmental Pollution*, Vol. 178, pp. 463 - 473. DOI : [10.1016/j.envpol.2013.03.052](https://doi.org/10.1016/j.envpol.2013.03.052)
- Zhang, Y., Yang, Z., Yu, X., 2015, Urban Metabolism: A Review of Current Knowledge and Directions for Future Study, *Environmental Science & Technology*, Vol. 49, n°19, pp. 11247-11263. DOI : [10.1021/acs.est.5b03060](https://doi.org/10.1021/acs.est.5b03060)
- Zuiderwijk, A., Janssen, M., Choenni, S., Meijer, R., Alibaks, R., 2012, Socio-technical Impediments of Open Data, *Electronic Journal of e-Government*, Vol. 10, n°2. URL : <http://www.ejeg.com/issue/download.html?idArticle=255> .

Liste des illustrations

Figure 1 : Le(s) cycle(s) de l'information	24
Figure 2 : Structures récursives : liste chaînée, arbre, graphe hiérarchisé, graphe	36
Figure 3 : Différentes dimensions des graphiques. Exemples de graphiques à deux, trois et quatre dimensions (c).	42
Figure 4 : « L'organisme territorial » et l'étude de son métabolisme	63
Figure 5 : Bilan de matière de la Région Alsace en 2010 à travers différents indicateurs,	73
Figure 6 : Illustration du Modifiable Temporal Unit Problem à travers la consommation d'un bâtiment fictif.	98
Figure 7 : Exemple de représentations d'indicateurs : diagramme en étoile et son équivalent discret	112
Figure 8 : Exemple de représentation de l'aire d'approvisionnement et d'exportation des fruits et légumes de Haute-Garonne (extrait de CGDD 2014)	112
Figure 9 : Huit dimensions pour l'analyse du métabolisme d'un territoire	114
Figure 10 : Extrait de la structure de la nomenclature NST2007	127
Figure 11 : Modèle logique de données implémenté dans le SINAMET	137
Figure 12 : Les étapes du traitement de données à travers le SINAMET, cas général et exemple	143
Figure 13 : Architecture logicielle du SINAMET	144
Figure 14 : Interface générale du prototype de SINAMET	150
Figure 15 : Affichage de la hiérarchie des produits de la nomenclature NST2007 et détail d'un produit spécifique	150
Figure 16 : Exemple du module de création d'un graphe régional du transport de marchandises.	151
Figure 17 : Conséquences graphiques d'une mauvaise convention de définition des périodes sur la consommation d'un bâtiment quelconque.....	162
Figure 18 : Profils de consommation (énergie finale) du patrimoine bâti	165
Figure 19 : Profil de consommation (énergie finale) du patrimoine bâti pour le gaz et de l'électricité.....	166
Figure 20 : Répartition des consommations du patrimoine bâti selon l'unité (en 2014).	167
Figure 21 : Évolution du coût moyen des différents types d'énergie (lissé sur 12 mois).	167

Figure 22 : Coût mensuel moyen de l'électricité.	168
Figure 23 : Coût moyen du chauffage urbain.....	168
Figure 24 : Consommation d'énergie (kWh EF) du patrimoine (en 2014).....	169
Figure 25 : Consommation cumulée des bâtiments les plus importants (en 2014).....	170
Figure 26 : Consommation électrique des « Vestiaire et tribune du stade de l'III » en MWh et euros et coût moyen de l'électricité en euro/MWh	171
Figure 27 : Consommation des bâtiments gérés par la collectivité en 2016 (en kWh EF/m ²) selon leur typologie	172
Figure 28 : Sensibilité climatique du centre administratif de l'Eurométropole de Strasbourg, pour chaque mois entre 2009 et 2017.....	173
Figure 29 : Sensibilité climatique des différents bâtiments (en 2016).....	173
Figure 30 : Trafic fluvial de marchandises (chargement + déchargement) des départements français.	183
Figure 31 : Tonnage des différentes marchandises transportées par voie navigable en France.	185
Figure 32 : Importations, exportations et trafic intérieur de marchandises dans le Bas-Rhin (par route et voie navigable), trafic de marchandises du Port Autonome de Strasbourg (voie navigable) en 2017.	186
Figure 33 : Évolution des quantités de marchandises transportées dans le Bas-Rhin par route et voie navigable, entre 2009 et 2018.....	187
Figure 34 : Nature et origine des marchandises importées dans le Bas-Rhin par voie navigable en 2017 (total : 2732 kt)	188
Figure 35 : Nature et destination des marchandises exportées depuis le Bas-Rhin par voie navigable en 2017 (total : 4981 kt).....	188
Figure 36 : Origine (destination) des marchandises importées (exportées) par route dans (depuis) le Bas-Rhin en 2017.	189
Figure 37 : Comparaison entre les statistiques de transport de marchandises par voie fluviale, dans le Bas-Rhin et dans les installations du PAS	190
Figure 38 : Trafic fluvial de marchandises du Port Autonome de Strasbourg (en kilotonne par mois).....	191
Figure 39 : Trafic de conteneurs dans les installations du PAS selon les différents modes de transport.....	192
Figure 40 : Trafic de conteneurs dans le Port Autonome de Strasbourg en 2017.....	193
Figure 41 : Fret ferroviaire à travers le PAS entre janvier et juin 2016.	193

Figure 42 : Ratios d'importation (en tonnes par habitant) de différents territoires administratifs en France. Marchandises transportées par camion et voie fluviale uniquement.....	204
Figure 43 : Masse des importations + flux internes de marchandises en fonction de la population pour les 96 départements français. Marchandises transportées par camion et voie fluviale uniquement.....	213
Figure 44 : Métabolisme moyen d'un corps humain.....	216
Figure 45 : Mesures de certains composés des eaux usées entrantes dans la station d'épuration Strasbourg Wantzenau.....	219
Figure 46 : Collecte des déchets ménagers (non triés) par le service de collecte de l'Eurométropole de Strasbourg	221
Figure 47 : Masse des ordures ménagères collectées en porte-à-porte en 2018 dans l'Eurométropole de Strasbourg	221
Figure 48 : Localisation des composteurs collectifs sur voie publique de l'Eurométropole de Strasbourg et zones à prioriser pour l'installation de nouveaux composteurs.	222
Figure 49 : Circuit de la matière alimentaire (hors boissons) dans une agglomération (valeurs en kt/an estimées pour l'Eurométropole de Strasbourg).	224

Liste des tableaux

Tableau 1 : Critères d'utilisabilité des données	28
Tableau 2 : Adéquation ressentie des structures de données pour différents usages.	43
Tableau 3 : Bilan de combustion complète d'un kilogramme de méthane (gaz naturel).....	67
Tableau 4 : Quelques exemples d'unités disponibles pour l'étude de différentes thématiques	87
Tableau 5 : Exemple d'application de l' « octomètre » du métabolisme territorial pour esquisser différentes problématiques autour de la thématique des déchets alimentaires.....	115
Tableau 6 : Adéquation ressentie de différentes techniques graphiques pour la représentation des différentes dimensions du métabolisme territorial.....	116
Tableau 7 : Périmètre de notre analyse sur les données de facturation d'énergie.....	158
Tableau 8 : Données de facturation du patrimoine de l'EMS	159
Tableau 9 : Données du Bilan énergie et eau patrimoine de l'EMS	160
Tableau 10 : Données sur le patrimoine bâti de l'Eurométropole de Strasbourg	160
Tableau 11 : Données climatiques (DJU18)	161
Tableau 12 : Périmètre de notre analyse sur les flux de marchandises dans les installations portuaires.....	179
Tableau 13 : Données du Système d'Information sur le Transport de Marchandises.....	180
Tableau 14 : Données sur le trafic fluvial du Port Autonome de Strasbourg.....	180
Tableau 15 : Données sur l'activité conteneur du Port Autonome de Strasbourg	181
Tableau 16 : Départements les plus significatifs au regard du transport de marchandises par voie navigable.	184
Tableau 17 : Principaux indicateurs de la comptabilité des flux de matières pour une économie entière (EWMFA)	202
Tableau 18 : Structure administrative (2015) et importations par camion et voie navigable (2010) pour la région Grand Est et ses subdivisions.	205
Tableau 19 : Données mesurées sur la station d'épuration de Strasbourg	218
Tableau 20 : Données de collecte des déchets en porte à porte (ménages et assimilés)	220
Tableau 21 : Propositions d'indicateurs d'accessibilité des composteurs collectifs	222

Annexes

Annexe 1	Entretien sur l'étude des flux de matières, d'eau et d'énergie réalisée à l'échelle de l'Alsace	255
Annexe 2	Synthèse partielle des sources de données pour la comptabilité des flux de matières et d'énergie.....	261
Annexe 3	Modèle de données	265
Annexe 4	Exemples de modules d'importation (ETL)	267
Annexe 5	Exemple d'un module de visualisation et exportation.....	269
Annexe 6	Cartographies autour des infrastructures portuaires	271
Annexe 7	Métabolisme d'un humain	277

Annexe 1 Entretien sur l'étude des flux de matières, d'eau et d'énergie réalisée à l'échelle de l'Alsace

Entretien réalisé le 27/10/2016 avec Virginie WOLFF (Chargée de mission économie circulaire – Région Grand Est) et Guillaume KAUFFMANN (Idée Alsace / Initiatives durables) qui ont contribué à la réalisation de l'étude. L'objectif de l'entretien était de :

- Mieux comprendre l'exercice d'une analyse de flux de matières et d'énergie (AMFE) à l'échelle d'une région,
- Identifier les parties prenantes de l'exercice et de la restitution,
- Mettre en exergue les difficultés rencontrées.

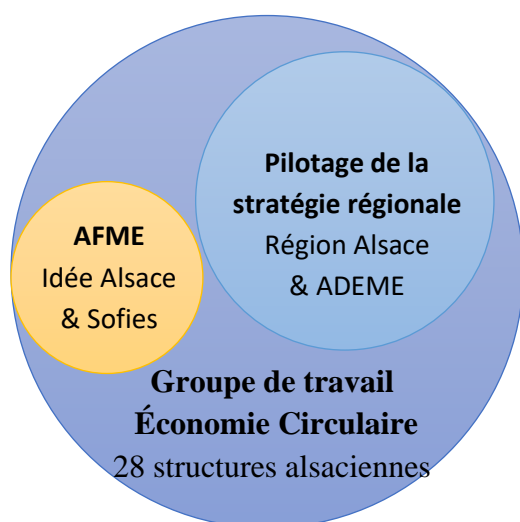
Cet entretien vient ainsi compléter le document de communication diffusé à la suite de l'étude : Région Alsace, Ademe, 2015, « Consommations, besoins et richesses du territoire alsacien ». Nous présentons ici une synthèse des échanges.

I. Stratégie régionale de soutien à l'économie circulaire

La région Alsace, avec le soutien de l'ADEME, s'est officiellement engagée en juillet 2014 à soutenir l'économie circulaire. La thématique a des implications vastes, bien plus larges que la simple question de la réutilisation des déchets.

Cette stratégie se décline en plusieurs objectifs, allant de la sensibilisation et communication à l'accompagnement technique et financier de projets. Il s'agit ainsi de :

- Accompagner les acteurs (collectivités, élus, entreprises, EPCI en charge des déchets ou de l'eau...) vers une prise de conscience des enjeux et une « acculturation » à l'économie circulaire,
- Présenter des retours d'expériences et des bonnes pratiques (valeur de l'exemple),
- Formaliser des priorités, des enjeux et des envies partagés par les acteurs du territoire grâce à l'implication d'un groupe de travail de 28 structures alsaciennes dans la démarche,
- Accompagner les porteurs de projets dans ce domaine techniquement et financièrement à travers un appel à projet lancé fin 2015 et début 2016 et qui a retenu 10 projets. D'autres sessions de l'appel à projet sont prévues en 2017.



Afin de pouvoir appuyer l'argumentaire autour de l'économie circulaire, la région et l'ADEME ont souhaité réaliser une analyse de flux de matières et d'énergie (AFME) à l'échelle régional. Un appel d'offre a été lancé pour cela, remporté par le consortium Sofies (apportant une expertise méthodologique et analytique et ayant déjà participé à une telle étude sur le canton de Genève) et Idée Alsace (ayant une bonne connaissance du territoire et des acteurs alsaciens).

Quelques structures du groupe de travail Économie Circulaire :

Grandes collectivités (départements, communes...), représentations de l'État (DIRRECTE, DREAL, DRAAF), pôles de compétitivité, CRITT, fédérations professionnelles (UNICEM, FRBTP), chambres économiques (CCI, CAA, JCEF, CRESS), Port Autonome de Strasbourg ASPA, Alsace Active

II. Méthodologie de l'AFME

Il convient de préciser que l'exercice entrepris n'avait pas pour vocation à obtenir une comptabilité fine et précise, mais surtout d'obtenir, en un temps limité, un ordre de grandeur des enjeux matériels territoriaux. L'objectif était ainsi avant tout d'aller vers un même langage entre les différents acteurs du territoire, de partager une vision sur les ressources, de proposer une grille de lecture des enjeux et d'amener progressivement l'économie circulaire dans le quotidien des décideurs.

Ainsi, la quantification de ces flux visait à appuyer l'argumentaire autour de l'économie circulaire (les chiffres sont souvent parlants, même s'ils ne sont pas la finalité), et de prioriser l'attention sur certaines ressources spécifiques.

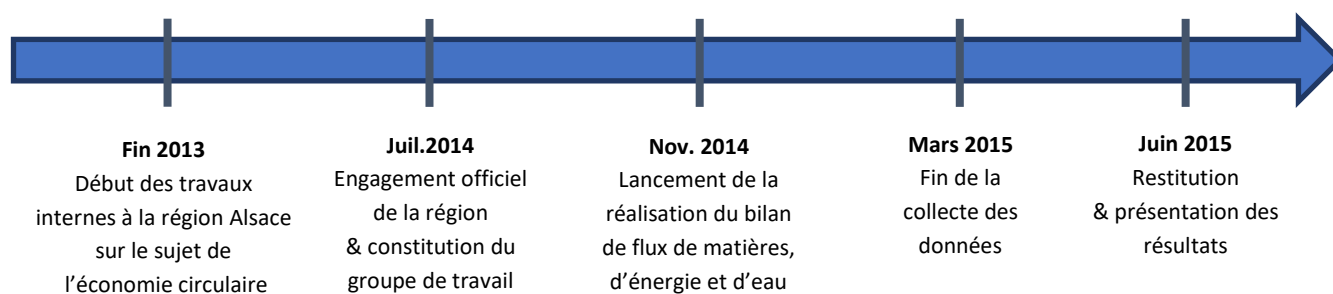
La méthode utilisée par Idée Alsace et Sofies repose en très grande partie sur le « Guide Méthodologique : Comptabilité des flux de matières dans les régions et les départements » édité par le Commissariat Général au Développement Durable en juin 2014. Cette méthode n'est pas exempte de problématiques qui sont explicitées dans le guide : - faire entrer des produits dans la nomenclature des matières (produits composites, manufacturés, terres d'excavation...) ; - accès difficile à certaines sources de données (difficiles à identifier, payantes – SitraM –, ou

ancienne – FRET ferroviaire – ...); - suivi de la chaîne logistique (un produit importé de Chine, transitant par le port Marseille, apparaît finalement comme étant en provenance de Marseille).

L'étude a été complétée par des questionnaires auprès d'acteurs locaux (fédération de métiers, acteurs institutionnels...) et par des données bibliographiques sur l'usage des ressources (par exemple : le suivi du recyclage n'est pas directement traité dans le guide). Les études concernant les flux d'énergie et les flux d'eau ne sont pas l'objet du guide méthodologique. Elles ont été réalisées pour le territoire alsacien grâce au soutien de l'ASPA (qui dispose d'une connaissance fine de la consommation d'énergie en Alsace) et du SDEA.

Idée Alsace et Sofies ont compilé manuellement les données dans un fichier Excel, seules quelques formules ont permis d'équilibrer automatiquement certains flux et de synthétiser les résultats. Un rapport final de l'étude a également été rédigé. Ces documents récapitulant l'ensemble de la démarche n'ont pas été diffusés au public par souci d'intelligibilité, mais ne contiennent pas d'informations sensibles : beaucoup des données sources sont publiques. La région et l'ADEME ont préféré présenter aux acteurs territoriaux une version synthétique de 6 pages que nous ne retrouvons plus en ligne suite aux changements d'adresse URL de la nouvelle région.

L'étude aura nécessité 105 jours de travail (jours × homme) par Idée Alsace et Sofies répartis entre novembre 2014 et juin 2015. Le suivi aura été mensuel par la région Alsace et l'Ademe. Le groupe de travail a été associé sur la thématique spécifique de l'AFME lors de 3 réunions. Il en ressort que tous les acteurs n'ont pas la même facilité (et le même intérêt) à appréhender le concept du bilan de flux de matières qui reste une notion parfois abstraite. La question de la fiabilité de la démarche a été souvent évoquée voire remise en cause sur certains chiffres, notamment ceux issus de coefficients techniques génériques qui ne tiendraient pas toujours compte des réalités technico-économiques (par exemple, l'utilisation des résidus dans la filière bois). Dans tous les cas, la démarche aura au moins permis d'apporter des éléments de dialogue transversaux.



III. Résultats

Cette analyse des flux de matières, d'énergie et d'eau, complétée par les travaux du groupe de travail a permis d'aboutir à une image de la matérialité du territoire alsacien (dépendance, ressources locales...). Cependant, l'image obtenue est plutôt celle d'une boîte noire, et la circulation des flux au sein du territoire reste encore mal connue, par exemple en ce qui concerne le recyclage intérieur. Cela a néanmoins permis de formuler une trentaine de leviers d'actions visant à limiter la dépendance du territoire et de ses entreprises, et à créer de nouvelles activités économiques. Cela permettra également de situer l'Alsace par rapport à d'autres régions.

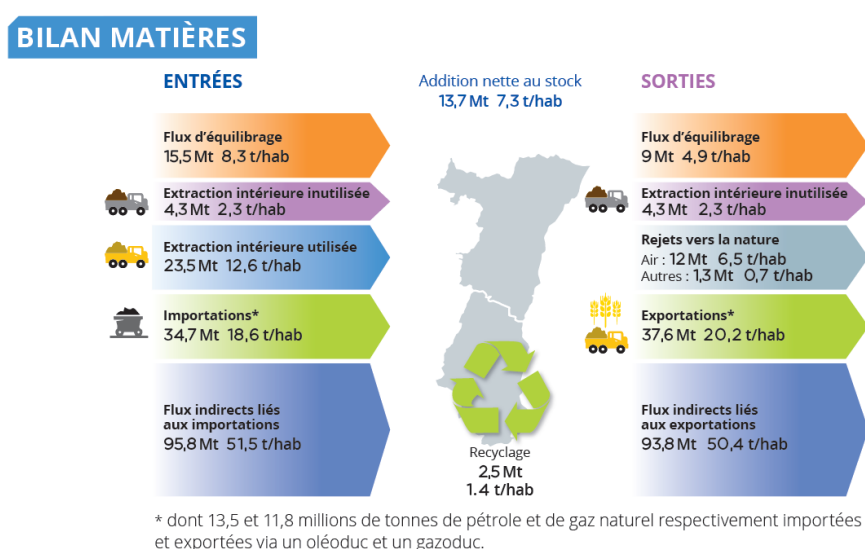


Figure extraite du document de synthèse « Consommations, besoins et richesses du territoire alsacien » édité par la région Alsace et l'ADEME

Trois axes prioritaires ont été définis avec le groupe de travail : le BTP, avec des problématiques d'accès, de promotion et de stockage des matériaux, en particulier recyclés ; la biomasse ; un axe divers (textile / plastiques / métaux). Le secteur du BTP a montré un vif intérêt dans la démarche d'économie circulaire, sachant qu'il est l'un des principaux mobilisateurs de la matière sur le territoire. Les axes doivent encore être approfondis, précisés et validés. Il s'agit de décliner les tâches entre acteurs privés et publics. En effet, ces derniers n'ont pas toujours la légitimité pour agir sur certains aspects, il est alors plus cohérent que les acteurs économiques privés portent la démarche en ce qui les concerne. Des animations de sensibilisation auprès de clubs d'entreprises, de zones d'activités... seront mises en place pour accompagner le dialogue entre les différents acteurs.

On notera que les restructurations régionales inhérentes à la récente réforme territoriale ont mis en pause depuis le début de l'année 2016 le processus de développement de l'économie circulaire (et pas seulement celui-là). La restructuration touche néanmoins à sa fin, et la dynamique devrait reprendre avec une meilleure coordination entre les trois régions Alsace, Lorraine et Champagne-Ardenne. Un chargé de mission dédié à l'économie circulaire a été nommé pour chacune des trois anciennes régions.

Avant la réforme, les régions Champagne-Ardenne et Lorraine avaient également engagé un travail de réflexion sur l'économie circulaire. Les démarches entreprises divergent cependant, ainsi que les acteurs impliqués :

- En Lorraine, c'est la DREAL, et non l'ADEME qui s'est associé à la région autour de cette thématique. La DREAL a d'ailleurs lancé un appel d'offres en juin 2015 pour la réalisation d'une comptabilité des flux de matières en Lorraine, mais les résultats ne semblent pas avoir encore été communiqués.
- En Champagne-Ardenne, la stratégie économie circulaire s'est plutôt orientée vers la constitution et le travail de plusieurs groupes thématiques.

Annexe 2 Synthèse partielle des sources de données pour la comptabilité des flux de matières et d'énergie

Cette annexe est un extrait du document numérique qui contient d'autres informations sur les sources de données :

https://docs.google.com/spreadsheets/d/1QEFdXToeoLe_wVEyChV9B49KuM9o9iZy505Rqy11mYs/edit?usp=sharing

Organisme / Service	Nom de la base de données	Catégorie	Type de données	Périmètre	Niveau de détail	Accès sur Internet
Eurostat				Union Européenne	Pays	Oui
Ministère de l'Agriculture / Service de la statistique et de la prospective (SSP)	Agreste		Cultures & Résidus de récoltes et pâturages,	France	Département	Oui
			Récolte de bois (Grumes, Bois d'industrie, Bois énergie)	France		Oui
			Surface jardins familiaux	France		Oui
		Rejets dissipatifs F.4.2	Surfaces agricoles	France	Département	
		Rejets dissipatifs F.4.1	Population animale (cheptel)	France		
Insee			Population des territoires	France	Régions, Départements, Communes	Oui
			CLAP - Nombre de salariés dans l'industrie	France	Département	
		Importation	Population dans les logements chauffés au gaz	France		

Ceren (Observatoire statistique national de la consommation d'énergie)			Bois Énergie (voir aussi SoES)			
FranceAgriMer Chambre syndicale des algues et des végétaux marin			Extraction de la biomasse Aquatique			
Fédérations départementales des chasseurs / Directions départementales des territoires/Bureaux de chasse			Extraction biomasse animale lors de la chasse			
DREAL (Direction Régionale de l'Environnement, de l'Aménagement et du Logement) Service prévention des risques / mines et carrières			Minerai métallique, Combustible fossile (base S3IC : Minerai non métallique			
UNICEM (Union nationale des industries de carrières et matériaux de construction)			Extraction de minerai non métallique			
Ministère de la Transition écologique et solidaire (MTES)	EIDER		Production gaz, charbon, pétrole Granulat (Tableaux détaillés – rubrique Sols)			
			Consommation de produits pétroliers			
			Déchets			
		Rejets dans l'eau F.3	Rubrique Eau : Rejets d'origine industrielle			
	Enquête annuelle sur le marché du gaz naturel				Région	
			Données locales d'énergie (consommation électricité, gaz, produits pétroliers)	France	Communes / IRIS	Oui
MTES / Service de la donnée et des études statistiques (SDES) / Sous-direction de l'observation statistique des transports / Bureau de la statistique des transports et des véhicules	SitraM	Importations / Exportations	Importations / Exportations via le transport de marchandises (route, rail, voie navigable)			Partiel

Cerema (Centre d'études et d'expertise sur les risques, l'environnement, la mobilité et l'aménagement)	Enquête Dragage	Extraction intérieure inutilisée	Boues de dragage dans les ports de façade maritime			Oui
		Produits dissipatifs F.4.7	Utilisation de sel (et autres) étendus sur les routes			
Cerema / Direction technique eau, mer et fleuves (ex-Cetmef)		Extraction intérieure inutilisée	Boues de dragage voies navigables (VNF)			-
Arrêté préfectoral	Plans de Gestion Pluriannuels des Opérations de Dragage	Extraction intérieure inutilisée				
ASPA (Association pour la surveillance de la pollution atmosphérique en Alsace)	Invent'Air		Émissions atmosphériques, Consommation d'énergie	Alsace		
Union des aéroports français	Statistiques de trafic		Traffic aéroportuaire pour estimer la consommation de carburacteur			
ATMO Agence agréée pour la surveillance de la qualité de l'air			Émissions atmosphériques, Consommation d'énergie	Régional		
CITEPA (Centre Interprofessionnel Technique d'Etudes de la Pollution Atmosphérique)		Émissions F.1	Émissions dans l'air			
ADEME	SINOE	Émissions F.2	Déchets municipaux	France	Régions et Département	
	Enquête ITOMA	Émissions solides F.2	Déchets entrants dans les installations de stockage de déchet non dangereux			Non
		Produits dissipatifs F.4	Déchets traités par les plateformes de compostage et quantité de compost produit			Non
ORDECO (Observatoire Régional des Déchets et de l'Economie Circulaire en Occitanie), ex-ORDIMIP				Occitanie		

Chambre de métiers et de l'artisanat	Outil Egida		Déchets des artisans			
	Plan régional de prévention et de gestion des déchets dangereux		Déchets dangereux			
	GEREP		Déchets dangereux collectés, émissions polluantes			
	Plan de prévention et de gestion des déchets issus de chantier de BTP		Déchets du BTP			
			Déchets agricoles			
SATESE (Service d'assistance technique aux exploitants de station d'épuration)		Rejets F.2, F.3				
UNIFA (Union Nationale des Industries de la Fertilisation)			Engrais livré	France	Région	
Chambre départementale d'agriculture			Épandage boues st. épuration			
SRAL (Services de l'alimentation) de la Draaf		Produits dissipatifs F.4.5	Vente de produits sanitaires			
SRISE (Service Régional d'Information statistiques et Economique)		Produits dissipatifs F.4.5	Pratiques culturelles en grandes cultures et viticultures			
UIPP (Union des Industries de la protection des plantes)		Produits dissipatifs F.4.5	Substances actives vendues en France			
DIR (Direction Interdépartementales des Routes)		Produits dissipatifs F.4.7	Utilisation de sel (et autres) étendus sur les routes			
		Pertes dissipatives F.5	Mémento de statistiques de transport			Oui
Agences de l'eau		Rejets F.2, F.3	Déchets d'assainissement (Rejets liquides et solides)			
Agence de l'eau Rhin-Meuse	SIERM (Système d'Information sur l'Eau Rhin-Meuse)		Eau	Bassin Versant Rhin-Meuse		Oui

Annexe 3 Modèle de données

Modèle de données non disponible dans la version publique de la thèse

Annexe 4 Exemples de modules d'importation (ETL)

I. Chargement de territoires

Ce module permet le chargement des données des départements (nom en français, code Insee et région d'appartenance) dans la base du SINAMET à partir d'un fichier récupéré sur le site de l'Insee.

Code non disponible dans la version publique de la thèse

II. Chargement de flux

Ce module permet le chargement des données sur le transport de marchandises entre départements, grâce aux fichiers extraits de la base SitraM et transmis par les services du Ministère de la Transition Ecologique et Solidaire.

Code non disponible dans la version publique de la thèse

Annexe 5 Exemple d'un module de visualisation et exportation

Nous présentons ici le code source exhaustif d'un module de visualisation et d'exportation des données ainsi que les résultats obtenus structurés sous trois formes. Ce module vise à calculer les tonnages de marchandises transportées entre les départements d'une région. La région est renseignée à l'aide d'une fenêtre de paramétrage (voir Figure 16 p. 151). Pour la bonne exécution du module, il est nécessaire d'avoir chargé au préalable les données de la base SitraM et un modèle territorial définissant les régions françaises et les départements correspondants (voir Annexe 4).

I. Code source

Code non disponible dans la version publique de la thèse.

II. Résultats (output)

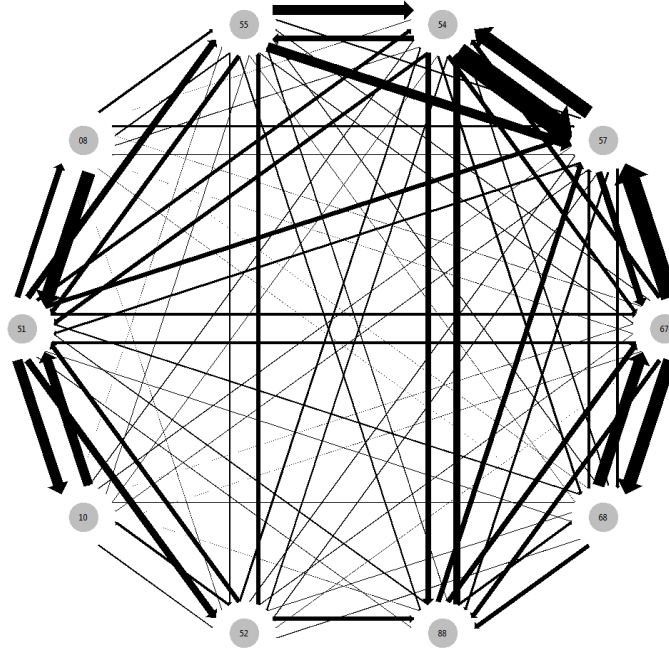
1) Affichage sous forme sérialisée textuelle

« `print(data[r])` » affiche les résultats dans la console sous forme sérialisée textuelle. La syntaxe est : ['dép. expéditeur' : { 'dép. destinataires' : qtés en kt }]. Ce format reste lisible, même s'il demande un certain temps de lecture.

```
{'54': {'54': 0.0, '55': 828.869, '57': 3075.020, '88': 817.675, '67': 582.512, '68': 373.217, '08': 153.69, '10': 103.731, '51': 521.165, '52': 193.986}, '55': {'54': 1381.51, '55': 0.0, '57': 1339.783, '88': 197.619, '67': 185.385, '68': 73.41, '08': 212.073, '10': 52.026, '51': 635.786, '52': 284.048}, '57': {'54': 2003.744, '55': 218.585, '57': 0.0, '88': 339.665, '67': 1014.594, '68': 312.765, '08': 172.918, '10': 103.318, '51': 335.159, '52': 140.459}, '88': {'54': 1047.532, '55': 162.89, '57': 716.509, '88': 0.0, '67': 724.997, '68': 314.674, '08': 0.0, '10': 24.154, '51': 213.338, '52': 433.613}, '67': {'54': 669.100, '55': 149.756, '57': 2511.346, '88': 660.090, '67': 0.0, '68': 1882.74, '08': 37.474, '10': 76.716, '51': 444.485, '52': 33.981}, '68': {'54': 170.654, '55': 90.727, '57': 464.087, '88': 584.524, '67': 1663.370, '68': 0.0, '08': 23.298, '10': 49.532, '51': 151.684, '52': 128.685}, '08': {'54': 55.127, '55': 343.895, '57': 212.858, '88': 5.887, '67': 76.837, '68': 22.346, '08': 0.0, '10': 7.675, '51': 1911.979, '52': 7.746}, '10': {'54': 137.406, '55': 44.156, '57': 107.455, '88': 16.996, '67': 63.724, '68': 33.588, '08': 29.496, '10': 0.0, '51': 1371.267, '52': 209.654}, '51': {'54': 510.085, '55': 893.175, '57': 630.506, '88': 125.140, '67': 468.154, '68': 182.798, '08': 833.374, '10': 1461.334, '51': 0.0, '52': 994.345}, '52': {'54': 273.424, '55': 527.352, '57': 159.168, '88': 572.513, '67': 81.370, '68': 91.497, '08': 3.858, '10': 355.384, '51': 647.251, '52': 0.0}}
```

2) Affichage sous forme graphique

« `drawgraph(data[r])` » affiche le résultat, sous la forme graphique ci-dessous (les numéros de départements ne sont pas bien visibles dans les bulles).



3) Export sous forme d'un tableau 2D

« `write_excel_dictmatrix(data[r])` » exporte les résultats dans un fichier tableur .xlsx sous la forme du tableau 2D ci-dessous.

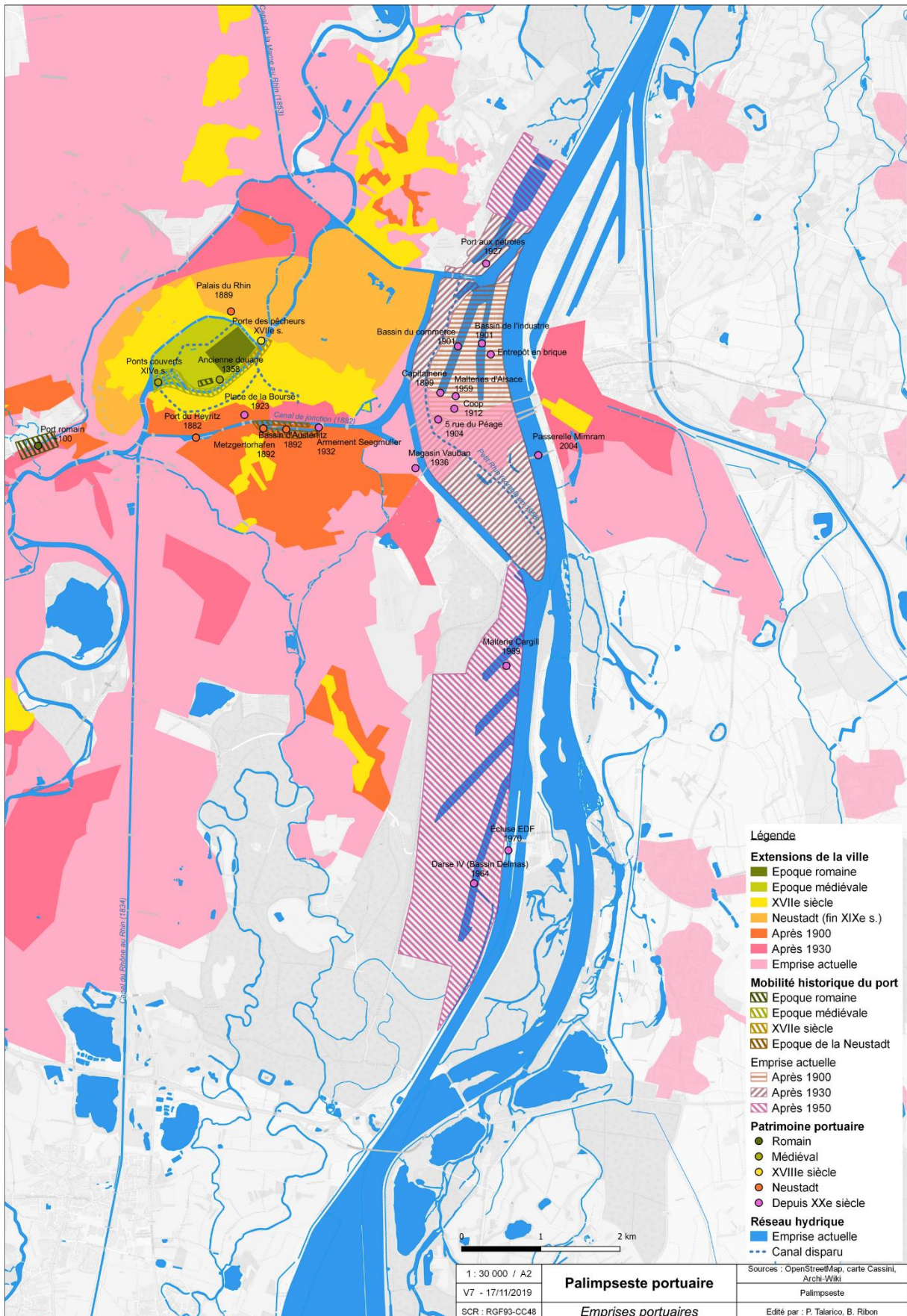
D \ O	54	55	57	88	67	68	08	10	51	52
54	0	829	3075	818	583	373	154	104	521	194
55	1382	0	1340	198	185	73	212	52	636	284
57	2004	219	0	340	1015	313	173	103	335	140
88	1048	163	717	0	725	315	0	24	213	434
67	669	150	2511	660	0	1883	37	77	444	34
68	171	91	464	585	1663	0	23	50	152	129
08	55	344	213	6	77	22	0	8	1912	8
10	137	44	107	17	64	34	29	0	1371	210
51	510	893	631	125	468	183	833	1461	0	994
52	273	527	159	573	81	91	4	355	647	0

Annexe 6 Cartographies autour des infrastructures portuaires

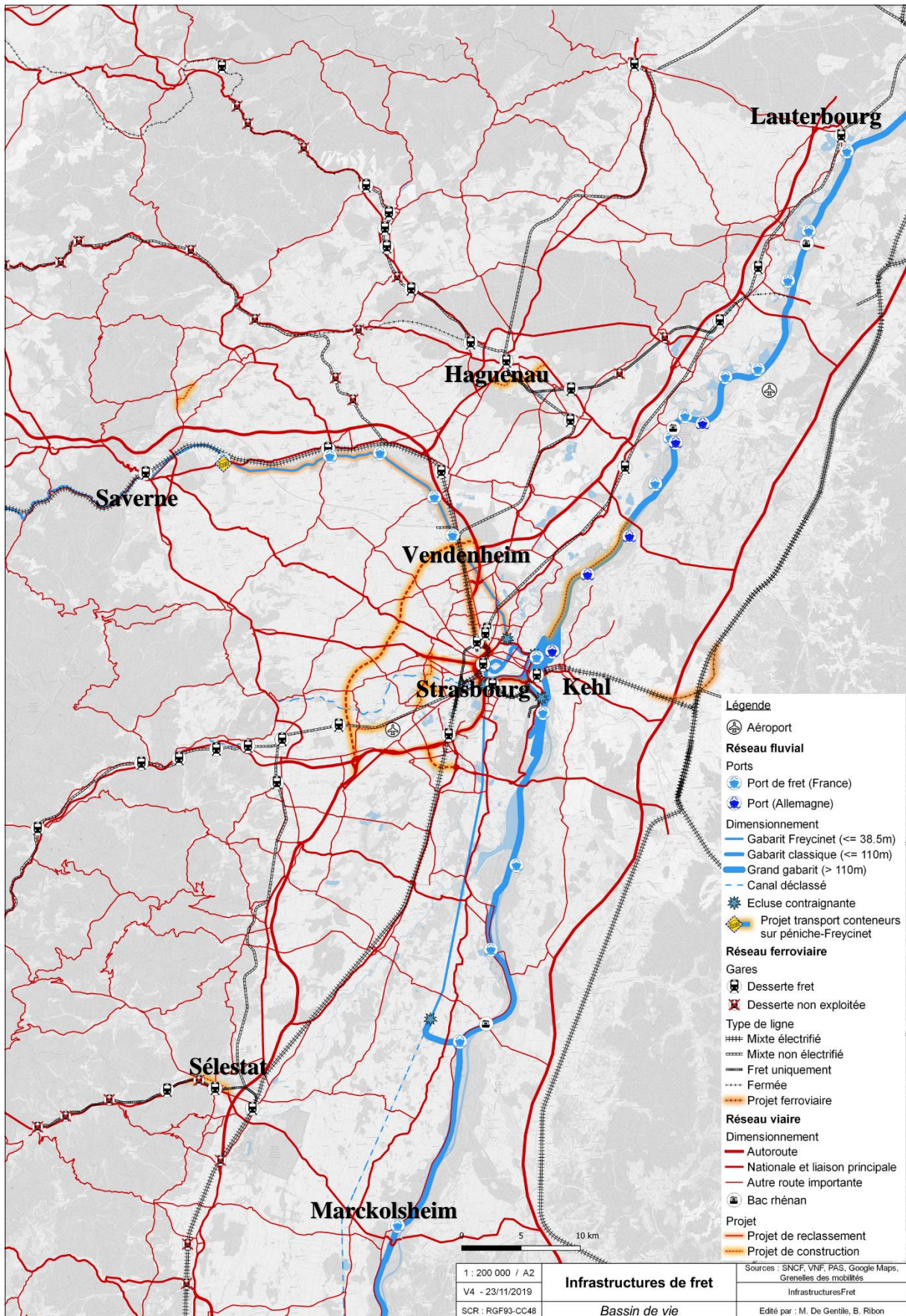
Les cartes créées durant le programme POPSU ont été réalisées à trois échelles : « Emprises portuaires », « Eurométropole » et « Bassin de vie ». Elles sont centrées sur Strasbourg et ses espaces portuaires.

Dans cette annexe, seules 4 des 28 cartes réalisées pendant le projet sont présentées, ce sont celles qui appuient nos propos sur la question des flux de marchandises à travers les installations portuaires du territoire (Chapitre 6).

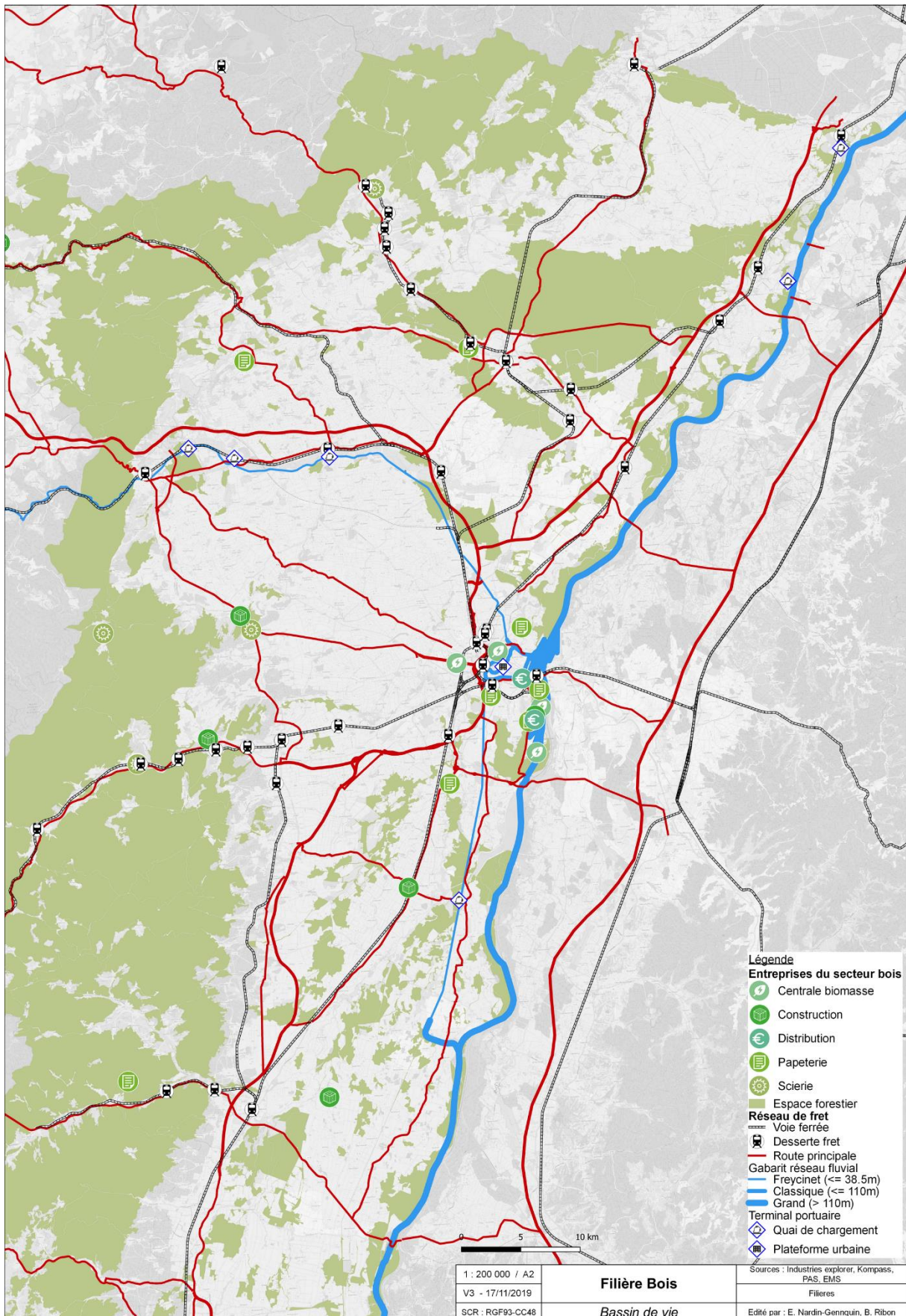
I. Palimpseste portuaire



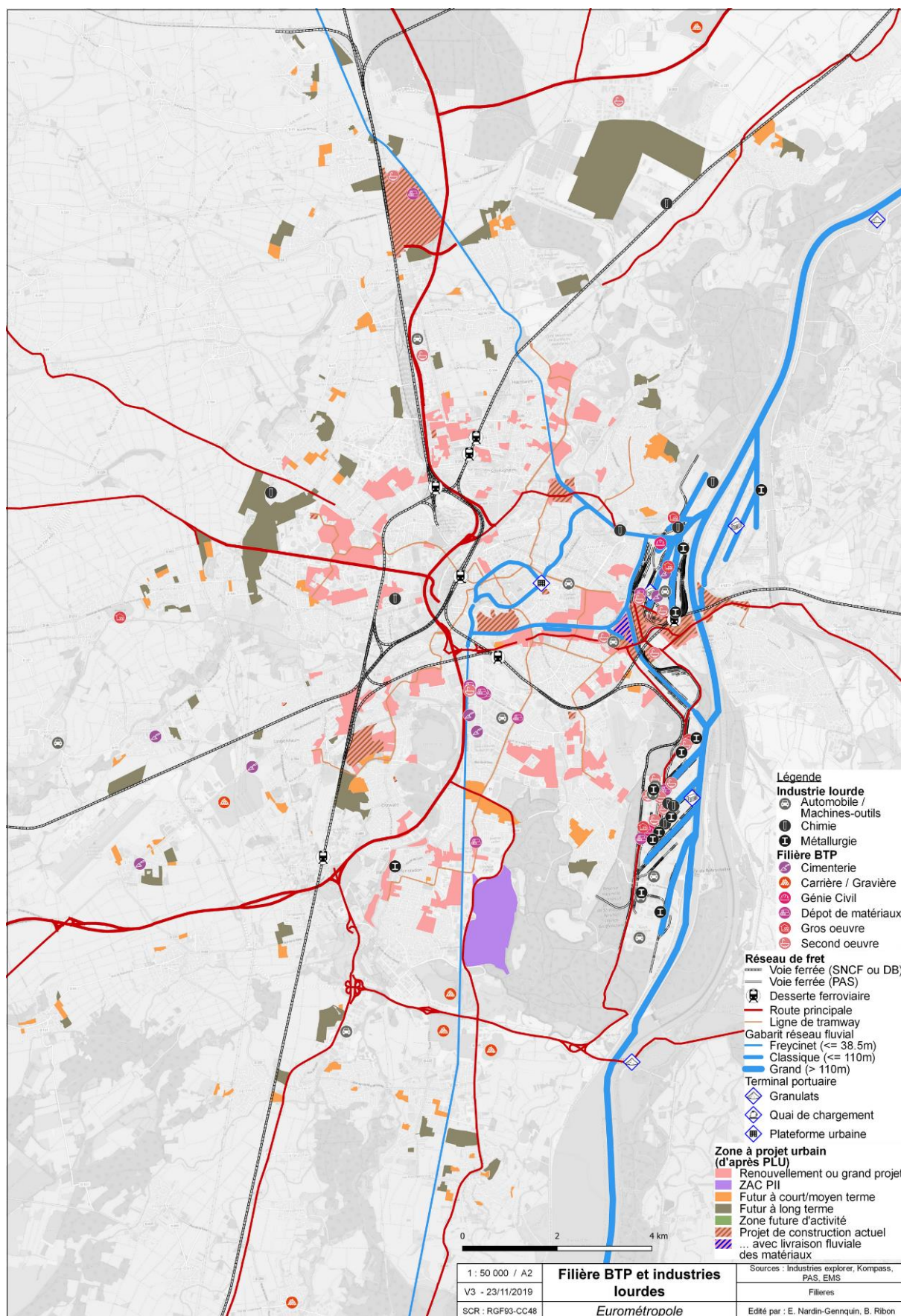
II. Infrastructures de fret



III. Filière bois



IV. Filière BTP et industrie lourde



Annexe 7 Métabolisme d'un humain

Nous présentons ici la démarche ayant mené à la réalisation du bilan entrées/sorties présenté Figure 44 p. 216. Les sources spécifiques de cette annexe sont présentées à la suite.

I. Respiration

Le CGDD (2014) présente des estimations de la masse de matière mobilisée lors de la respiration : 0,68 kg d'O₂ consommé en entrée, et 0,82 kg de CO₂ et 0,96 kg de vapeur d'eau rejetés en sortie. Ces chiffres sont compatibles avec ceux trouvés dans certaines références (18 000 à 20 000 mmol, soit 0,79 à 0,88 kg de CO₂ d'après Agasti 2011), bien que l'on puisse trouver des estimations un peu supérieures (0,83 kg d'O₂ consommé et 1 kg de CO₂ rejeté d'après MacElroy et al. 1992). Le carbone composant le CO₂ expiré (0,22 kg, 27 % de la masse du CO₂) provient des aliments solides consommés. 0,60 kg d'oxygène vont être mobilisés pour former le CO₂, il reste ainsi 0,08 kg d'oxygène dont nous faisons l'hypothèse qu'il participe à la formation d'eau métabolique.

II. Alimentation

Les habitudes de consommations alimentaires sont extrêmement diverses selon les cultures et les catégories socio-professionnelles. Les travaux de l'Anses (2017) nous permettent toutefois d'avoir une idée des habitudes de consommation moyenne pour une personne habitant en France. Ainsi, un adulte ingère quotidiennement 2,94 kg d'aliments : 1,17 kg d'aliments solides et 1,77 kg de boissons. Nous trouvons toutefois des valeurs supérieures chez MacElroy et al. (1992) : 1,84 kg d'aliments solides (composés à 64 % d'eau), et dans l'étude de l'Adeus (2017) basée sur la proposition d'un nutritionniste : 1,88 kg de besoins journaliers (dont 0,50 kg de produits laitiers, donc avec le lait classé dans les boissons par l'Anses).

III. Selles et urines

Les selles et urines permettent d'évacuer les déchets produits par le corps humain. Bien qu'essentiellement composés d'eau, environ 0,10 kg de matières solides sont évacués de cette manière : 0,04 kg par les selles et 0,06 kg par les urines selon Rose et al. (2015). Ces chiffres correspondent à une personne moyenne dans un pays à haut revenu. Par exemple, la quantité de fibres consommée, plus importante dans les pays à bas revenus, influe à la hausse les

quantités de selles produites (Rose et al. 2015). On retrouve également 0,10 kg évacués dans l'étude de MacElroy et al. (1992), mais uniquement par les fèces.

IV. Bilan hydrique

L'eau est un des principaux composants du corps humain, et nécessite un renouvellement permanent (EFSA 2010). L'eau mobilisée dans le corps humain provient des boissons consommées : 1,50 kg (Bauer et al. 2010) ; 1,77 kg (Anses 2017), 1,85 kg (MacElroy et al. 1992), des aliments solides : 0,7 kg (Bauer et al. 2010) ; 1,17 kg (MacElroy et al. 1992), de la production d'eau métabolique résultant des processus chimiques à l'intérieur du corps humain : 0,3 kg (Bauer et al. 2010) : 0,25 à 0,35 kg (EFSA 2010) ; 0,10 kg (MacElroy et al. 1992). Cette eau est ensuite rejetée via les selles : 0,11 kg (Rose et al. 2015) ; 0,10 kg (Bauer et al. 2010) ; 0,20 kg (MacElroy et al. 1992 ; EFSA 2010), via les urines : 1,40 kg (Rose et al. 2015) ; 1,50 kg (Bauer et al. 2010 ; MacElroy et al. 1992), et à travers la peau ou par la respiration (pertes imperceptibles) : 0,90 kg (Bauer et al. 2010) ; 0,96 kg (CGDD 2014 qui ne mentionne que la respiration) ; 1,82 kg (MacElroy et al. 1992). La quantité d'urine est généralement égale à la quantité de boisson ingérée.

V. Synthèse

Comme nous pouvons le voir, les valeurs diffèrent selon les travaux et il n'est pas évident de les rendre compatibles. Nous procédons à plusieurs choix arbitraires. Tout d'abord, nos travaux s'inscrivant dans la continuité de ceux du CGDD (2014), nous utilisons les mêmes données disponibles pour la respiration (0,68 kg d'O₂, 0,82 kg de CO₂ et 0,96 kg de vapeur d'eau). Nous considérons ensuite 0,10 kg de matière solide rejetée par les selles et les urines. La matière sèche non rejetée sous forme de carbone (dans le CO₂) ou par les urines et selles, ainsi que l'oxygène non rejeté sous forme de carbone vont servir la production d'eau métabolique. Du point de vue hydrique, nous fixons la quantité de boissons et la quantité d'urine égales à 1,50 kg, la quantité d'eau métabolique produite à 0,30 kg et la quantité d'eau dans les fèces à 0,15 kg. Ces hypothèses imposent le reste des valeurs possibles si nous voulons que la conservation de la masse soit vérifiée, à la fois pour la Matière Sèche (MS), mais aussi pour l'eau, ce qui n'est pas le cas dans les travaux de MacElroy et al. (1992). La quantité de MS dans les aliments solides consommés doit être ainsi de 0,54 kg et la part d'eau de 0,81 kg (soit 60 % d'eau). Cela représente donc 1,35 kg d'aliments solides consommés quotidiennement, ce qui se situe entre les valeurs présentées par l'Anses (2017) et par l'ADEUS (2017). Étant donné la

divergence des valeurs trouvée dans la littérature, mais aussi la cohérence globale de notre analyse dans ce nuage de valeurs, nous estimons raisonnable de retenir cette valeur moyenne de 1,35 kg d'aliments consommés.

VI. Bibliographie de l'annexe

- ADEUS, 2017, Le système alimentaire local bas-rhinois - Première analyse, *Les notes de l'ADEUS*. Agence de Développement et d'Urbanisme de l'Agglomération Strasbourgeoise, 8 p.. [En ligne] URL : http://www.adeus.org/productions/les-notes-de-ladeus-ndeg253-economie/files/note-253_systeme_alimentaire_local_67-web.pdf. Consulté le 19 avril 2018.
- Agasti, T., 2011, Textbook of Anesthesia for Postgraduates - 1st Edition. Jaypee Brothers Medical Pub, 1186 p.
- Anses, 2017, Étude individuelle nationale des consommations alimentaires 3 (INCA 3). Avis de l'Anses et Rapport d'expertise collective., 535 p.. [En ligne] URL : <https://www.anses.fr/fr/system/files/NUT2014SA0234Ra.pdf>. Consulté le 19 avril 2018.
- Bauer, W., Badoud, R., Lölliger, J., Eturnaud, A., 2010, Science et technologie des aliments : principes de chimie des constituants et de technologie des procédés. Lausanne : Presses polytechniques et universitaires romandes, 720 p.
- CGDD, 2014, Comptabilité des flux de matières dans les régions et les départements - Guide méthodologique, *Références du Service de l'observation et des statistiques (SOeS)*. Commissariat Général au Développement Durable. [En ligne] URL : <https://www.statistiques.developpement-durable.gouv.fr/sites/default/files/2018-10/guide-methodologique-references-flux-de-matiere-juin2014.pdf>. Consulté le 19 avril 2018.
- EFSA, 2010, Scientific Opinion on Dietary Reference Values for water, *EFSA Journal*, 8, pp. 1459. DOI : [10.2903/j.efsa.2010.1459](https://doi.org/10.2903/j.efsa.2010.1459)
- MacElroy, R., Kliss, M., Straight, C., 1992, Life support systems for Mars transit, *Advances in Space Research*, 5, pp. 159--166. DOI : [10.1016/0273-1177\(92\)90022-p](https://doi.org/10.1016/0273-1177(92)90022-p)
- Rose, C., Parker, A., Jefferson, B., Cartmell, E., 2015, The Characterization of Feces and Urine: A Review of the Literature to Inform Advanced Treatment Technology, *Critical Reviews in Environmental Science and Technology*, 45, pp. 1827-1879. Taylor & Francis. DOI : [10.1080/10643389.2014.1000761](https://doi.org/10.1080/10643389.2014.1000761)

Le paradigme du métabolisme territorial : Une approche par le traitement des données

Résumé

Le métabolisme territorial offre un paradigme pour étudier les flux de matières et d'énergie sur un territoire. Il vise à mieux qualifier et quantifier les ressources mobilisées et rejetées dans l'environnement. Néanmoins, l'étude du métabolisme reste complexe par la quantité de données à mobiliser et à traiter. Dans cette thèse, nous abordons directement cette problématique. Pour commencer, nous formalisons les notions et approches à mobiliser autour du traitement des données et du métabolisme. Nous concevons ensuite un Système d'Information pour l'Analyse du Métabolisme des Territoires (SINAMET). Enfin, nous mettons en application ces outils sur quatre cas d'études : la consommation d'énergie du patrimoine de l'Eurométropole de Strasbourg, les marchandises transportées par voie navigable dans le port de Strasbourg, la sensibilité à l'échelle des indicateurs d'importation et d'exportation, et les flux de matières alimentaire à l'échelle de l'Eurométropole.

Comptabilité ; Flux ; Matière ; Énergie ; Ecologie ; Territoire ; Système d'Information

Résumé en anglais

Territorial Metabolism Paradigm: A Data Processing Approach

Territorial metabolism is a paradigm that helps study material and energy flows on a territory level. It aims to better quantify and qualify resources that are used and rejected in the environment. Metabolism studies remain complex due to the amount of data that needs to be gathered and processed. In this thesis, we tackle this problem in a straightforward way. We begin with formalising basic concepts and approaches in the domains of data processing and metabolism. We then conceptualise an Information System for Territorial Metabolism Analysis. Finally, we apply these tools on four case studies as follow: energy consumption of the Eurométropole de Strasbourg's property assets, waterway-freight in Strasbourg's harbour, scale-sensitivity of import and export indicators, and foodstuff flows at the Eurométropole's scale.

Accounting ; Flow ; Material ; Energy ; Ecology ; Territory ; Information System