

Bulletin du Statec n° 8-2010

Sommaire

Système statistique intégré de l'énergie – sources et méthodes

1. L'utilité des statistiques de l'énergie	415
2. Les centralisations internationales	415
3. Les complexités de la centralisation des données de l'énergie	417
4. La conception d'un système d'information statistique intégré pour l'énergie	418
5. Les différentes phases du projet luxembourgeois sur l'énergie	419
6. Les éléments-clés du système d'information statistique intégré luxembourgeois de l'énergie	422
7. Les conclusions et perspectives	446
Annexe 1 : Liste non exhaustive des textes législatifs européens du domaine de l'énergie	448
Annexe 2 : Structure de la base de données de l'énergie	450

Table des matières

Système statistique intégré de l'énergie – sources et méthodes*

1. L'utilité des statistiques de l'énergie	415
2. Les centralisations internationales	415
3. Les complexités de la centralisation des données de l'énergie	417
4. La conception d'un système d'information statistique intégré pour l'énergie	418
5. Les différentes phases du projet luxembourgeois sur l'énergie	419
6. Les éléments-clés du système d'information statistique intégré luxembourgeois de l'énergie	422
6.1 La centralisation des données	422
6.1.1 L'inventaire des sources et besoins	422
6.1.2 La confrontation des sources et besoins	423
6.1.3 La conception du nouveau schéma organisationnel	424
6.1.4 La mise en place du système de centralisation de l'information	427
6.1.5 La création de la base de données	428
6.2 La production des statistiques	429
6.2.1 La production de bilans énergétiques (en volume et valeur) et le développement de méthodes d'estimation	429
6.2.2 La production des questionnaires internationaux	431
6.2.3 La production de tables spécifiques	434
6.3 La vérification de la cohérence	435
6.3.1 L'analyse de la qualité des données	435
6.3.2 La cohérence des séries chronologiques	436
6.4 La diffusion des statistiques	443
6.4.1 Les principes de publication et regroupement thématique	443
6.4.2 Les supports de publication des tableaux	444
6.4.3 Le développement d'indicateurs	444
7. Les conclusions et perspectives	446
 Les encadrés	
Encadré 1: Les institutions internationales du domaine de l'énergie	416
Encadré 2: L'évaluation des processus de collecte, de traitement et de communication	423
Encadré 3: La méthodologie pour l'évaluation des transmissions de données	433
Encadré 4: Les critères d'évaluation de la qualité des données	436
Encadré 5: La méthodologie d'une analyse de la tendance et saisonnalité	442
Encadré 6: Le bilan énergétique restreint du Luxembourg en 2009	445
Encadré 7: Quelques indicateurs mis en valeur	446

Les figures

Figure 1: Procédures de collecte et de traitement des données et analyses de qualité s'y référant	417
Figure 2: Capture d'écran d'une partie de la matrice de correspondance en sa table principale (matrice des sources et besoins)	423
Figure 3: Différents types de schéma organisationnel des données	425
Figure 4: Schéma organisationnel des données établi au Luxembourg	426
Figure 5: Outil pour lister les transmissions de données à réaliser	432
Figure 6: Evolution mensuelle des livraisons intérieures en gasoil carburant en tonnes	438
Figure 7: Evolution mensuelle des importations d'essence 95oct en tonnes	439
Figure 8: Evolution mensuelle de la production brute thermique en GWh	440

Les annexes

Annexe 1 : Liste non exhaustive des textes législatifs européens du domaine de l'énergie	448
Annexe 2 : Structure de la base de données de l'énergie	450

Les abréviations

AIE :	Agence internationale de l'énergie
ARIMA :	autoregressive integrated moving average
Eurostat :	Institut statistiques de l'Union Européenne
FED :	fichier d'échange des données
GES :	gaz à effet de serre
GJ :	gigajoule
GWh:	gigawatt-heure
ILR :	Institut luxembourgeois de la régulation
INS :	Institut national statistique
OCRA :	Office de la commercialisation et du ravitaillement
oct:	octane
SIG :	système d'information géographique
SIS :	système d'information statistique
STATEC :	Institut national de la statistique et des études économiques
tep :	tonne équivalent pétrole
TJ :	terajoule
TRAMO:	time series regression with ARIMA noise, missing values and outliers

Le résumé

A l'occasion de l'entrée en vigueur du règlement européen 1099/2008/CE, le Luxembourg a entamé une révision profonde de son système de centralisation des statistiques de l'énergie.

Une nouvelle organisation de la collecte, du traitement, de la vérification et de la compilation des données a été mise en place. Cette nouvelle organisation a été mise en œuvre en veillant, d'une part, à l'intégration des missions des administrations publiques travaillant dans le domaine de l'énergie, et d'autre part, aux nouvelles missions de transmission de données du Service central de la statistique et des études économiques (STATEC) suite au nouveau règlement européen.

Ce travail a permis d'évaluer l'existant en termes d'inventaires de données énergétiques et de leurs schémas de transmission, de construire un nouveau schéma organisationnel pour la collecte et la diffusion de données relatives à l'énergie, de développer les outils nécessaires à sa mise en place, d'améliorer la cohérence des statistiques énergétiques luxembourgeoises et d'ébaucher une meilleure communication sur ces statistiques.

1. L'utilité des statistiques de l'énergie

A l'aube du 21^{ème} siècle, l'humanité fait face à l'un des plus grands challenges de son histoire : comment maintenir voire accroître sa qualité de vie pour une population grandissante où de plus en plus de peuples aspirent à de hauts critères de bien-être.

Du point de vue des matières énergétiques, la grande majorité des modes de fonctionnement des sociétés modernes actuelles sont basées sur l'exploitation et la transformation de ressources primaires finies. Ces modes de fonctionnement sont arrivés à un point d'inflexion où la ressource n'est plus suffisante pour répondre à la demande. Conséquence visible de cette situation au sein des systèmes économiques : les prix de l'énergie sont devenus beaucoup plus volatiles et atteignent des records jamais égalés tout en ayant une tendance généralement croissante.

Du point de vue de l'environnement, la consommation croissante de produits fossiles à des fins énergétiques entraîne un accroissement des émissions de gaz dans l'atmosphère. Ces gaz, en trop grande quantité, sont responsables de l'accroissement de « l'effet de serre » qui a notamment pour conséquence un dérèglement climatique planétaire et *in fine* des désastres écologiques et économiques touchant aussi bien les pays du nord que du sud.

A la conférence de Copenhague, en décembre 2009, tous les pays présents ont marqué leurs accords pour limiter l'émission de ces gaz à effet de serre et réduire leur empreinte énergétique. Chaque gouvernement s'est engagé à un triple objectif, à savoir : limiter sa consommation en énergies fossiles, améliorer l'efficacité énergétique des processus et développer les énergies dites renouvelables.

Il est dès lors plus que jamais indispensable de disposer de statistiques énergétiques cohérentes, fiables et reproductibles afin que les gouvernements puissent mener des politiques efficaces, que les gestionnaires d'entreprises puissent décider d'investissements durables et que les ménages puissent orienter leurs achats en connaissance de cause.

La problématique de l'énergie est à ce point importante et complexe, qu'au sein d'un même pays de nombreuses institutions, administrations et organisations compilent leurs propres statistiques énergétiques répondant spécifiquement à leurs besoins. Ce mode de fonctionnement dans le processus d'établissement de la statistique pose de nombreux problèmes et notamment celui de la diffusion d'informations incohérentes et incomplètes.

Pour pallier à ces difficultés, l'instauration d'un système d'information statistique (SIS) centralisé auprès d'un seul et unique organisme est la solution opérationnelle la plus probante. Habituellement, ce sont les instituts nationaux de la statistique (INS) qui se chargent de cette centralisation. L'expérience acquise dans le domaine des méthodes statistiques permet aux INS de garantir la fiabilité et la représentativité des statistiques produites. De plus, les INS sont généralement les interlocuteurs privilégiés des organismes internationaux et ce, tout particulièrement, dans les pays membres de l'union européenne.

2. Les centralisations internationales

Plusieurs institutions supranationales se sont engagées dans la voie de la collecte et de la comparaison de statistiques énergétiques nationales en vue d'œuvrer à la diffusion d'informations cohérentes au niveau international. Le Luxembourg participe activement à cette démarche en rapportant ses statistiques plus particulièrement à deux d'entre elles : l'Agence Internationale de l'Energie (AIE) et Eurostat.

Afin de garantir la cohérence des statistiques nationales collectées, l'AIE et Eurostat ont établi une série de questionnaires harmonisés ainsi qu'un manuel d'utilisation à destination des agents responsables de la compilation de ces statistiques nationales¹.

En ce qui concerne les Etats membres de la Communauté européenne, la transmission de ces questionnaires harmonisés a été entérinée dans le règlement européen (1099/2008/CE) du Parlement européen et du Conseil du 22 octobre 2008 concernant les statistiques de l'énergie et prenant effet au 1^{er} janvier 2009.

¹ Energy statistics manual, IEA Publications, France – septembre 2004

Encadré 1: Les institutions internationales du domaine de l'énergie

L'Agence Internationale de l'Energie (AIE) est un organe autonome institué en novembre 1974 dans le cadre de l'Organisation de Coopération et de Développement Economiques (OCDE). Elle a notamment pour mission première de tenir à jour et d'améliorer les systèmes permettant de faire face à des perturbations des approvisionnements pétroliers ainsi que de fournir des données sur différents aspects des marchés internationaux de l'énergie. Les transmissions de données à l'AIE se font sur base volontaire par les pays membres.

Eurostat est l'office statistique de la Communauté européenne. Sa mission est de fournir des informations statistiques de haute qualité à l'Union européenne. Pour ce faire, elle s'appuie sur le réseau des Instituts

Nationaux de Statistiques (INS) qui délivrent des statistiques comparables et harmonisées.

Deux cas de figure sont possibles :

- Dans le cas d'un règlement, les données doivent être transmises dans leurs entières et ce dès l'entrée en vigueur du règlement européen (aucune transposition nationale n'est nécessaire).
- Dans le cas d'une directive, le pays doit tout d'abord transposer la directive européenne en droit national et peut, à cette occasion, modifier le texte selon les spécificités du pays. Les données sont transmises dans des rapports dont le format est défini par le pays.

Ce règlement décrit les produits énergétiques, les agrégats et leurs unités à rapporter dans ses annexes B, C et D : respectivement annuellement, mensuellement (3 mois après le mois de référence) et mensuellement à court terme (1 mois après le mois de référence). Les vecteurs énergétiques concernés sont les combustibles solides, le gaz naturel, l'électricité et la chaleur, les produits pétroliers et enfin les énergies renouvelables. Quant aux agrégats demandés, ils couvrent aussi bien les domaines de l'approvisionnement, que ceux de la transformation et de la consommation de ces vecteurs énergétiques.

Les statistiques ainsi collectées permettent d'établir des tableaux comparatifs entre pays. Les deux principaux types de publication sont la compilation des bilans énergétiques nationaux et le développement d'indicateurs de développement durable.

D'autres textes législatifs et réglementaires concernent également, en totalité ou en partie, les statistiques de l'énergie. Une liste non exhaustive de ces législations est reprise à l'annexe 1.

Les statistiques énergétiques collectées par le biais des questionnaires harmonisés AIE / Eurostat ne réclament que des variables portant sur les volumes. Pour ce qui est des statistiques portant sur les prix des énergies, les variables collectées par ces deux institutions sont différentes et ne font pas l'objet d'un questionnaire commun.

L'AIE collecte, de son côté, les prix agrégés des principaux produits pétroliers ainsi que ceux du gaz et de l'électricité à un rythme trimestriel. Eurostat, quant à lui, collecte semestriellement les prix du gaz naturel et de l'électricité à destination des consommateurs finals. La transmission à Eurostat des prix aux consommateurs finals industriels est rendu obligatoire par voie de directive européenne (2008/92/CE) tandis que la transmission des prix aux consommateurs finals résidentiels se fait sur base volontaire.

Au sein d'un marché de l'énergie libéralisé, il y a un intérêt marqué pour que les prix pratiqués (et leurs structures) par les distributeurs et fournisseurs deviennent publiques. Cela dit, le respect du principe de concurrence impose de garder confidentielles les données individuelles. Ces deux exigences deviennent antagonistes dans le cas de catégories de consommation regroupant un nombre restreint de clients. Les instituts nationaux de la statistique jouent ici un rôle prépondérant. En effet, la grande majorité des instituts nationaux de la statistique fonctionnent sur base d'un respect de la confidentialité des données transmises (principe dit « du secret statistique »). Il apparaît dès lors que les instituts nationaux de la statistique sont les organismes les mieux à même, au sein d'un pays, pour collecter, traiter et compiler ce type de statistiques.

3. Les complexités de la centralisation des données de l'énergie

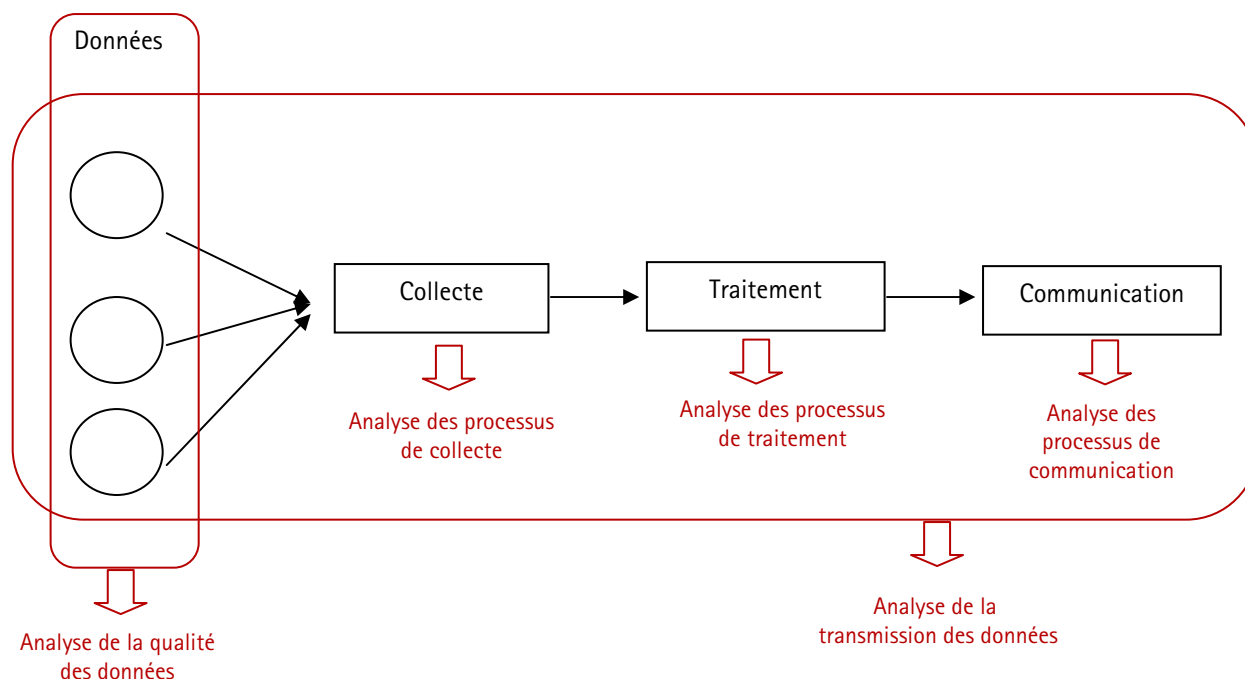
Au cours des vingt dernières années, la demande en statistiques énergétiques a crû de façon importante. Pouvoirs politiques, administrations publiques, journalistes, associations environnementales, universités, entreprises,... ont souhaité avoir une vision plus claire du marché de l'énergie afin d'orienter leurs propres travaux.

L'enjeu majeur pour le statisticien est de produire des statistiques cohérentes et de qualité. Cette cohérence des statistiques et la qualité des données collectées sont parfois mises en défaut suite à des problèmes sur la procédure de collecte et de traitement des données. Afin de minimiser ces erreurs, le statisticien réalise des évaluations des différents processus qui composent la procédure statistique.

Dans le domaine de l'énergie, les difficultés suivantes méritent d'être soulignées :

- Suite à la libéralisation totale du marché national, le statisticien doit adapter ses procédures de collecte et de traitement pour préserver la couverture des données sur la distribution énergétique. Anciennement, seule la compagnie nationale devait être contactée pour obtenir l'ensemble des données nécessaires. Aujourd'hui, la multitude des intervenants implique de multiples collectes ainsi que le recoupement des déclarations et l'ajustement des données sur base de l'importance et du rôle de chacun des intervenants.
- Suite au développement des échanges du marché international de l'énergie, le statisticien voit se réduire les données disponibles auprès des opérateurs finaux. Ainsi certaines informations sur le pays d'origine ou de la destination finale d'un produit ne sont plus automatiquement connues des acteurs nationaux actifs dans le pays.

Figure 1: Procédures de collecte et de traitement des données et analyses de qualité s'y référant



- Les récents objectifs de réduction de la consommation énergétique et de l'augmentation de l'efficacité énergétique impliquent que le statisticien ne doit plus seulement s'intéresser à la production, l'importation, l'exportation, les stocks et la consommation, mais également à la manière de consommer l'énergie. Il doit alors combiner les statistiques de l'énergie avec d'autres statistiques (comme le nombre de téléviseurs par ménage,...) qui ne sont pas forcément collectées dans le même contexte que les données de l'énergie. Se pose alors la question de la cohérence des données collectées.
- La progressive décentralisation de la production d'énergie (principalement active dans la cogénération et les énergies renouvelables) implique qu'une partie de l'utilisation de l'énergie sort du réseau public (facilement échantillonné) et entre dans le domaine de l'autoproduction et/ou du marché non régulé (comme le bois de chauffage) qui est par définition difficilement appréhendable par les procédures statistiques actuelles.
- La libre circulation des biens et personnes entre les pays implique également pour le statisticien de devoir estimer des statistiques sur des unités qui n'apparaissent pas dans la population potentiellement enquêtée. Pour le Luxembourg, l'exemple le plus significatif est la consommation de carburant par les non-résidents.

Ces quelques exemples de spécificité du domaine de l'énergie montrent le défi lancé au statisticien pour établir des statistiques cohérentes et de qualité. Outre le développement de méthodes d'échantillonnage, de traitement et d'estimation spécifiques, le statisticien doit se munir d'un système d'information à la hauteur de ce challenge. L'utilité de la centralisation des données énergétiques provenant de sources de données différentes est donc explicitement établie.

La complexité croissante du domaine de l'énergie et des techniques mises en place par le statisticien pour suivre cette tendance implique également de faire évoluer son rôle. Celui-ci ne doit plus être considéré comme un simple producteur de statistiques, mais doit devenir un interlocuteur à part entière dans le débat sur l'énergie. Un interlocuteur qui sera amené à expliquer les statistiques et leurs tendances.

4. La conception d'un système d'information statistique intégré pour l'énergie

L'outil le plus probant dans le dessein d'une mise à la disposition de données fiables nécessaires à la conception, au suivi et à l'évaluation de politiques économiques et/ou énergétiques est très certainement le système d'information statistique intégré.

Il est vrai que chaque secteur et chaque administration essaie, de façon séparée, d'améliorer son système d'information afin de garantir la qualité de leur travail respectif. Mais ce type de schéma organisationnel linéaire où chaque système d'information est mis côte à côte ne présente généralement pas les critères requis pour permettre des analyses multi-secteurs ou multi-produits. Tout l'intérêt d'un système d'information intégré est de faire converger, sur le plan national, l'ensemble des efforts sectoriels fournis pour atteindre un niveau élevé de cohérence statistique.

Les objectifs supplétifs d'un système d'information statistique intégré en comparaison de systèmes d'informations linéaires sectoriels sont :

- d'assurer la maîtrise de la production d'informations harmonisées, fiables et utiles ;
- d'améliorer la circulation des informations entre les différents secteurs ;
- et de rendre disponibles et accessibles à tous, les informations nécessaires pour la prise de décision.

En d'autres termes, le développement d'un système d'information statistique intégré doit permettre l'accroissement de l'utilisation des statistiques de l'énergie.

La définition d'un système d'information statistique intégré

De façon générale, un système d'information se définit comme un ensemble de moyens humains, de moyens matériels et de méthodes se rapportant au traitement de l'information de base et devant conduire à la fourniture d'informations synthétisées ou agrégées dans un but décisionnel.

Le système d'information n'est pas un concept purement théorique. Il doit répondre à des objectifs précis qui sont la compilation et la diffusion de statistiques nécessaires à la prise de décision.

Le système d'information intégré, en sus des objectifs de compilation et de diffusion, incorpore des objectifs supplémentaires de centralisation et de vérification de la qualité et de la cohérence des informations collectées et compilées.

Le système d'information statistique intégré peut donc être défini comme l'ensemble :

- *Des moyens humains* : personnes qui manipulent, émettent ou transmettent de l'information ;
- *Des moyens matériels* : ordinateurs, supports de l'information (papier, numérique) ;
- *Des méthodes* : logiciels d'ordinateurs, statistiques, analyses quantitatives et qualitatives des données, modèles mathématiques, etc.

Cet ensemble se rapporte aux phases de collecte, de traitement et de diffusion des différentes formes d'informations utiles au suivi et à l'évaluation de la mise en œuvre de décisions.

La phase de collecte inclut la recherche et l'obtention d'information *via* des moyens variés (enquêtes, déclarations obligatoires,...) dans le respect des règles de techniques statistiques de sondage et de collecte.

La phase de traitement inclut l'enregistrement et le stockage d'information sur un support adéquat (soit momentanément, soit pour archivage), sa consultation (consultation de fichiers, d'une base de données, recherche documentaire,...), sa présentation (modification de sa forme), son agrégation (modification de son contenu) et sa vérification (description de sa qualité et de sa cohérence).

La phase de diffusion inclut la communication de l'information sous une forme adaptée vers un ou un ensemble d'individus, en direct ou à distance.

La création d'un système d'information statistique intégré

L'approche méthodologique préconisée pour créer un système d'information statistique intégré comporte différentes phases :

1. L'analyse de l'existant
2. La conception du système
3. La mise en œuvre du système

Dans le cadre de la mise en place, au Luxembourg, d'un système d'information statistique intégré appliqué au domaine de l'énergie, il a fallu un peu plus de deux années pour accomplir ces trois phases.

Il va s'en dire qu'un système d'information statistique centralisé n'est pleinement opérationnel que si les acteurs de ce système jouent parfaitement leurs rôles. Dans l'exemple du Luxembourg, il est important de souligner l'entière participation des administrations telles que l'Institut luxembourgeois de régulation, l'Office de la commercialisation et du ravitaillement, la Direction de l'énergie, l'Administration de l'environnement, le Ministère du développement durable et le STATEC qui ont compris l'importance de la fourniture de données énergétiques de qualité et de l'exploitation conjointe de celles-ci.

5. Les différentes phases du projet luxembourgeois sur l'énergie

Les tâches devant mener à la constitution d'un système d'information statistique intégré appliqué au domaine de l'énergie au Luxembourg ont été articulées en trois phases :

- La phase 1 concernait la collecte des informations sur les sources de données et l'analyse des besoins en statistiques énergétiques. Cette phase 1 incluait également une analyse de la qualité des données collectées.
- La phase 2 concernait la confrontation des sources et des besoins, ainsi que l'élaboration de propositions pour l'organisation de la collecte et de la transmission des données du domaine de l'énergie, y compris la collecte de données manquantes ou incomplètes.

- La phase 3 concernait la mise en application du schéma organisationnel défini à la phase 2 et la compilation des statistiques à destination des instances nationales et internationales.

En préalable à ces trois phases, une étude du cadre législatif du domaine des statistiques de l'énergie a été réalisée. Cette étude portait d'une part sur le nouveau règlement européen en lui-même et d'autre part sur les législations actuellement en vigueur au Luxembourg.

Les analyses accomplies lors de la phase 1

L'analyse des législations du domaine de l'énergie

Sur base des documents législatifs officiels et des documents préparatoires à ces législations, un inventaire des obligations de différentes institutions luxembourgeoises en matière de transmission de statistiques énergétiques a été réalisé.

L'analyse de la collecte et de la transmission des données

Deux analyses ont été faites : l'une portant sur la phase de collecte, l'autre portant sur la phase de transmission des données. La même méthodologie a été appliquée pour ces deux analyses, pour lesquelles la collecte des informations nécessaires était généralement réalisée conjointement.

Les procédures de collecte et de transmission ont finalement été retranscrites sous forme de schémas organisationnels, présentant les grandes étapes et/ou intervenants de la procédure, ainsi que certains éléments détaillés de la procédure.

Ces schémas organisationnels ont été produits par institution mais également par vecteur énergétique. Un schéma organisationnel de l'ensemble des procédures existantes de collecte et de transmission des données (en volume et en valeur) a également été produit.

Les caractéristiques des données collectées et des statistiques rapportées ont été retranscrites dans deux dictionnaires de variables distincts.

L'analyse de la qualité des données collectées

La qualité des données a été caractérisée suivant certains critères de qualité proposés par Eurostat (précision, ponctualité, comparabilité, cohérence).

Les analyses accomplies lors de la phase 2

La confrontation des sources aux besoins

Tout d'abord un tableau de correspondance entre les enquêtes-sources de données sur les produits énergétiques et les besoins (administratifs et statistiques) en statistiques énergétiques a été produit. La matrice de correspondance ainsi constituée est un tableau à double entrée avec caractérisation des intersections. Cette matrice permet une bonne vision d'ensemble des statistiques de l'énergie au Luxembourg et est assez précise pour cerner les transmissions de données actuelles entre les sources et les besoins mais également les transmissions envisageables et intéressantes. Cette matrice a également servi à l'élaboration de propositions de recoupement des données.

Un autre tableau de correspondance, spécifique aux agrégats demandés par le nouveau règlement, a été produit à l'aide des deux dictionnaires de variables préalablement établis. Cette matrice à simple entrée permet de caractériser la transmission de données actuellement réalisée dans le domaine de l'énergie :

- identification des agrégats actuellement rapportés ;
- identification des lacunes de transmission ou des imprécisions de la transmission ;
- identification des fournisseurs de données et des services centralisateurs ;
- spécification des unités et pouvoirs calorifiques exigés.

Cette matrice a servi à l'élaboration de propositions de nouvelles enquêtes pour la transmission de données suivant le nouveau règlement.

L'analyse des schémas organisationnels

Les schémas organisationnels ont été analysés et mis en parallèle des souhaits des différentes administrations luxembourgeoises abordant la thématique de l'énergie, dans l'optique d'une réorganisation de la transmission des questionnaires internationaux. Un nouveau schéma organisationnel a ainsi été établi.

L'optimisation des procédures existantes

L'analyse des procédures de collecte des données et de la qualité des données collectées a conduit à la proposition de solutions permettant d'optimiser ces procédures de collecte mais également le stockage et la vérification de la qualité des données.

Les recommandations sur les procédures existantes portaient sur diverses rubriques comme le cycle de production ou les techniques de collecte, de transmission et de stockage.

La conception de nouvelles procédures

Les agrégats identifiés comme pertinents au Luxembourg mais pour lesquels les données collectées étaient absentes ou incomplètes ont fait l'objet d'une réflexion approfondie visant à déboucher sur des propositions pour compléter les enquêtes actuelles, pour mettre en place de nouvelles enquêtes ou pour développer des méthodes d'estimation pertinentes.

Dans ce dernier cas, une description explicite de chacune des méthodes d'estimation envisagées a été produite.

Les analyses accomplies lors de la phase 3La mise en œuvre du schéma organisationnel

Les procédures de collecte des informations auprès des institutions participantes ont été activées. Les transmissions des informations destinées à la centralisation sont réalisées *via* des fichiers d'échange de données (FED) spécifiquement conçus. Ces FED contiennent des modules de vérification de l'encodage et de demande de commentaires lors de variations mensuelles ou annuelles importantes. Ces données sont ensuite chargées dans la base de données dédiée aux statistiques de l'énergie.

La vérification de la cohérence des séries chronologiques

L'ensemble des données historiques luxembourgeoises du domaine de l'énergie ont été collectées et formatées pour entrer dans une analyse de séries temporelles de type ARIMA. La méthode TRAMO a été prioritairement utilisée pour détecter les valeurs aberrantes et suspectes sur ces séries chronologiques.

La réalisation des transmissions de données nationales et internationales

Différents codes de programmation ont été développés pour permettre la compilation automatique des agrégats réclamés par les transmissions mensuels auprès des instances internationales comme l'AIE et Eurostat, mais également auprès des unités de l'institut national de la statistique et des études économiques du Luxembourg (STATEC).

La constitution de bilan énergétique équilibré

Le bilan énergétique est un outil primordial pour appréhender de manière holistique le domaine de l'énergie au sein d'un pays. Le bilan énergétique (habituellement composé de trois parties : approvisionnement, transformation et consommation) a pour objectif de mettre en relation les flux d'énergie opérés sur un territoire défini. L'approche innovante utilisée dans le bilan énergétique luxembourgeois a été de structurer la partie « consommation » du bilan selon une approche sectorielle de type comptabilité nationale, c'est-à-dire que les postes sont classifiés par branche d'activité et non par utilisation comme cela est généralement réalisé dans les bilans énergétiques classiques. Cette approche a abouti à une structure standard permettant la comparaison aisée des statistiques énergétiques en volume et en valeur.

Comme première étape à cet objectif, le bilan énergétique a nécessité le développement d'une méthode de répartition des consommations par branche ainsi que plusieurs méthodes d'estimation de production et de consommation.

La définition de tableaux et indicateurs pertinents

L'utilisation des statistiques énergétiques implique la diffusion de tableaux répondant aux principales attentes des utilisateurs potentiels. Sous réserve du principe de confidentialité des données, la majorité des statistiques disponibles au sein de la base de données ont été formatées pour une communication aisée sur le Portail statistique (outil central de la diffusion des statistiques établies par le STATEC). Les séries de tableaux n'ont pas été classifiées selon la structure d'un bilan, mais selon les thématiques actuelles de questionnement de la société luxembourgeoise (conjoncture énergétique, sécurité et dépendance énergétique, efficacité énergétique, développement des énergies renouvelables, marché et prix de l'énergie, impacts sur l'environnement). Cette structure par thématique permet plus facilement la création d'indicateurs spécifiques à ces thématiques.

6. Les éléments-clés du système d'information statistique intégré luxembourgeois de l'énergie

6.1 La centralisation des données

6.1.1 L'inventaire des sources et besoins

Le développement d'un système d'information statistique intégré à l'échelon d'un pays est une tâche complexe car elle demande l'intégration des tenants et aboutissants de chacune des collectes de données.

La collecte de données auprès d'entreprises par les administrations, à l'exception de celle réalisée par le STATEC, a pour vocation première de répondre à des besoins administratifs. L'utilisation de ces données à des fins statistiques est possible sous réserve de concordance des définitions des variables collectées.

Mais à l'inverse, les données individualisables collectées à des fins statistiques ne peuvent être utilisées à des fins administratives. En effet la législation relative au secret statistique interdit au STATEC de divulguer des données individualisables.

Il est dès lors important que la procédure de collecte tente au maximum d'intégrer les administrations comme étape intermédiaire. Dans le cas d'une configuration idéale, l'institut national de la statistique ne joue que le rôle de centralisateur des données.

De ce fait, l'inventaire des sources et besoins a porté une attention particulière aux données collectées par les administrations publiques en tant que futurs partenaires privilégiés ainsi qu'à la définition exacte des variables collectées.

L'organisation historique de la collecte et du traitement des données énergétiques au Luxembourg est fortement conditionnée à la mise en application de textes législatifs internationaux.

La Direction de l'énergie, administration centrale dans ce domaine, est chargée de l'élaboration des orientations futures en matière de politiques énergétiques. Elle est soutenue par les travaux de l'Office commercial du ravitaillement (OCRA) qui est en charge, d'une part, de l'observation de l'évolution des prix des produits pétroliers et des combustibles solides, et d'autre part, de l'enregistrement des importations, des exportations et de la mise à la consommation finale. De plus, les analyses de la Direction de l'énergie sont complétées par celles du Service de l'énergie de l'Institut luxembourgeois de régulation (ILR) qui assure la fonction de régulateur du marché de l'électricité et du marché du gaz naturel. Ces trois organismes sont sous la responsabilité du Ministère de l'économie et du commerce extérieur.

L'Administration de l'environnement collecte également des données du domaine de l'énergie en vue de calculs des émissions de gaz à effet de serre. Cette administration est sous la direction du Ministère du développement durable.

L'institut national de la statistique et des études économiques (STATEC) n'intervenait pas dans le schéma organisationnel historique.

L'analyse des inventaires des sources a permis de mettre en avant un élément directeur de l'organisation historique à savoir que l'approche « top-down » était privilégiée lors des collectes de la filière du Ministère de l'économie et du commerce extérieur alors que l'approche « bottom-up » caractérisait la collecte de la filière du Ministère de l'environnement. En d'autres termes, d'un côté les thématiques de l'approvisionnement et de la transformation étaient travaillées alors que d'un autre côté la thématique de la consommation était abordée.

Encadré 2: L'évaluation des processus de collecte, de traitement et de communication

La collecte de données n'a de sens que si, une fois les données enregistrées, classées et organisées, on leur donne une signification en les insérant dans un cadre d'interprétation. La conversion de données en information nécessite la réalisation d'un ensemble de processus de transmission et d'opérations sur ces données (cf. figure 1).

Trois grandes étapes sont identifiées : la collecte, le traitement et la communication. Dans le cadre de ce projet sur les statistiques de l'énergie au Luxembourg, différentes analyses de processus ont été envisagées (en rouge dans la figure 1) :

- Les processus de collecte analysés sur base de critères : de codage des variables, de saisie des données et de vérification des données ;
- Les processus de transmission évalués sur base de critères : de formalisme de la transmission, de format du module de transmission, de

stockage des données transmises et de centralisation de ces données ;

- Les processus de traitement étudiés sur base de critères : de complexité de l'opération, de fréquence de réalisation et d'automatisation de l'opération ;
- Les processus de communication appréciés sur base de critères : de multiplicité des outils de diffusion, d'interactivité des outils de communication et de possibilité de mises à jour.

Chaque source de données inventoriée a fait l'objet de ce type d'évaluation. Chaque analyse était accompagnée de recommandations visant à optimiser certains éléments de ces processus.

6.1.2 La confrontation des sources et besoins

Les inventaires des « sources » et des « besoins » de chacune des administrations travaillant dans le domaine de l'énergie, ont été représentés sous la forme d'un schéma de transmission de données. Cette méthode est habituellement appliquée dans des optiques d'optimisation des systèmes. Mais cette approche méthodologique est apparue trop

conceptuelle et entachée d'incertitudes quant à la réalisation effective des missions de chacun.

Disposant de l'ensemble des variables collectées par les différentes administrations (dictionnaire des « sources ») et de l'ensemble des statistiques à rapporter (dictionnaire des « besoins »), il est possible de créer une matrice de correspondance entre ces deux dictionnaires de variables.

Figure 2: Capture d'écran d'une partie de la matrice de correspondance en sa table principale (matrice des sources et besoins)

			S1-UC3 QO	S2-UC3 QO	S1-UC4 QO	S2-UC4 QO	S1-UD1 QO	S1-OCRA QO	S2-OCRA QO	S3-OCRA QO	S4-OCRA QO	S1-DirEner GA
B1-ILR (30/30)	✓	BS-BA										
B2-ILR (14/14)	✓	BA										
B3-ILR (12/12)	✓	BA										
B1-UA1 (30/30)	✓	BS						var indir	var indir			
B2-UA1 (4/4)	✓	BS										
B1-UB1 (6/6)	✓	BA			var indir	var indir	var indir					
B2-UB1 (6/6)	✓	BS										
B3-UB1 (4/5)	✗	BS									var cont	
B1-UB4 (3/3)	✓	BS*										
B2-UB4 (2/2)	✓	BA										
B1-UC1	✓	BA	var indir		var indir		var indir					
B2-UC1=NR	✓	BA										
B1-UC2 (14/14)	✓	BS	var indir		var indir		var indir					
B1-UC3 (1/1)	✓	BA	var dir (1)				var cont					
B2-UC3 (4/4)	✓	BA		var dir (4)								
B1-UC4 (2/2)	✓	BA			var dir (1)	var dir (1)						
B2-UC4 (1/1)	✓	BS			var indir	var indir						
B3-UC4 (1/1)	✓	BA				var dir (1)						
B4-UC4 (1/1)	✓	BS				var indir						
B1-UD1 (38/38)	✓	BA					var dir (38)					
B2-UD1 (38/38)	✓	BS					var indir					
B1-OCRA-NR (10/10)	✓	BS						var dir (10)				
B2-OCRA-NR (4/12)	✗	BS-BA						var dir (4)				
B3-OCRA (7/8)	✗	BS						var dir (7)				
B4-OCRA (2/2)	✓	BA							var dir (2)	var cont		
B5-OCRA-NR (15/16)	✗	BS-BA								var dir (15)		
B6-OCRA-NR (4/4)	✓	BS								var dir (4)		
B7-OCRA-NR (34/38)	✗	BS								var dir (34)		
B8-OCRA (3/3)	✓	BS								var dir (3)		

L'analyse de cette matrice de correspondance a fait apparaître différents cas de figure, à savoir :

- *Une source qui répond directement à un besoin.* Ce cas correspond généralement à des enquêtes réalisées spécifiquement pour répondre à un besoin administratif.
- *Plusieurs sources combinées qui répondent à un besoin.* Ce cas de figure correspond aux besoins statistiques qui combinent différents enquêtes administratives et statistiques.
- *Une source qui répond partiellement à un besoin,* des variables sont donc manquantes. C'est le cas où une source est utilisée à d'autres fins que le besoin initial et où seule une partie des variables sont disponibles. Ce cas implique une transmission incomplète.
- *Une source qui répond partiellement à un besoin que remplit complètement une autre source.* Ce cas de figure apparaît lorsque deux enquêtes différentes sont réalisées à des fins différentes mais pour lesquelles des variables communes sont demandés aux répondants. Dans le présent cas, des opérations de contrôle de la validité des données sont possibles.
- *Un besoin auquel aucune source ne répond.* Ce cas de figure apparaît généralement lors de la création de nouveaux besoins ou lorsque aucun système de collecte ne permet d'appréhender les variables réclamées par le besoin.

Le schéma organisationnel vise à tirer parti de ces différents cas de figure et de permettre *via* la transmission complète ou partielle des variables de sources de données de répondre à un maximum de besoins. Dans certains cas, le recours à de nouvelles sources est nécessaire.

6.1.3 La conception du nouveau schéma organisationnel

L'un des intérêts d'un système d'information statistique intégré est de convertir un schéma organisationnel linéaire (comme l'était l'organisation historique) en un schéma plus centralisé où l'information collectée est unique (cf. figure 3 et 4). Cet intérêt répond parfaitement au souhait de la simplification administrative mais sa mise en application n'est pas aisée. En effet, un système

linéaire unique permettant de traiter l'ensemble des besoins des différentes administrations implique une mise en commun totale des données au sein d'un *datawarehouse*. Cette solution, bien que recommandée pour des questions de cohérence et de vérification des données et des statistiques produites, n'est que très rarement applicable tant les besoins des uns et des autres sont éloignés et relèvent parfois d'objectifs incompatibles aux yeux des fournisseurs de données. Une autre solution aurait été d'introduire des procédures d'échange de partie ou de totalité des bases de données mais cette solution n'apporte qu'une amélioration fictive du système, car elle sous-entend des tâches réalisées en doublon au sein de chaque système linéaire voir d'un surplus de travail lors des vérifications de la cohérence des données nationales agrégées.

La solution préconisée fut d'introduire au sein du schéma organisationnel historique, un nouveau partenaire, qui ne soit pas lié à des missions administratives, et qui prendrait en charge la centralisation d'une partie des données collectées par les administrations en vue d'apporter une plus-value quant à des critères de vérification de la cohérence et de la qualité des données ainsi qu'une plus-value sur la compilation de statistiques intégrant les approches des différents systèmes linéaires.

Ces plus-values ont été quantifiées préalablement lors de la mise en place du nouveau schéma organisationnel grâce à la réalisation d'une analyse approfondie des dictionnaires des variables. En effet, la matrice de correspondance est un outil permettant au travers de la sélection (ou non) de l'intersection d'une donnée collectée et de sa statistique rapportée d'évaluer les intérêts de différentes possibilités d'organisation du système d'information statistique intégré.

Les critères de sélection pour le schéma organisationnel du Luxembourg ont été répartis en thématiques :

1. Comparaison du nombre de besoins dans chacun des schémas organisationnels et du nombre de rapports complets qui sont transmis à l'institution demanderesse.
2. Comparaison du nombre de variables demandées dans chacun des schémas organisationnels et du nombre de variables réellement transmises.

3. Présentation de la répartition des sources des données. Deux catégories ont été retenues : variables obtenues par enquête et variables obtenues par estimation.
4. Présentation de la répartition de la standardisation des collectes de données. Deux catégories ont été retenues : variables obtenues par enquête standardisée et variables obtenues sur demande.

Figure 3: Différents types de schéma organisationnel des données

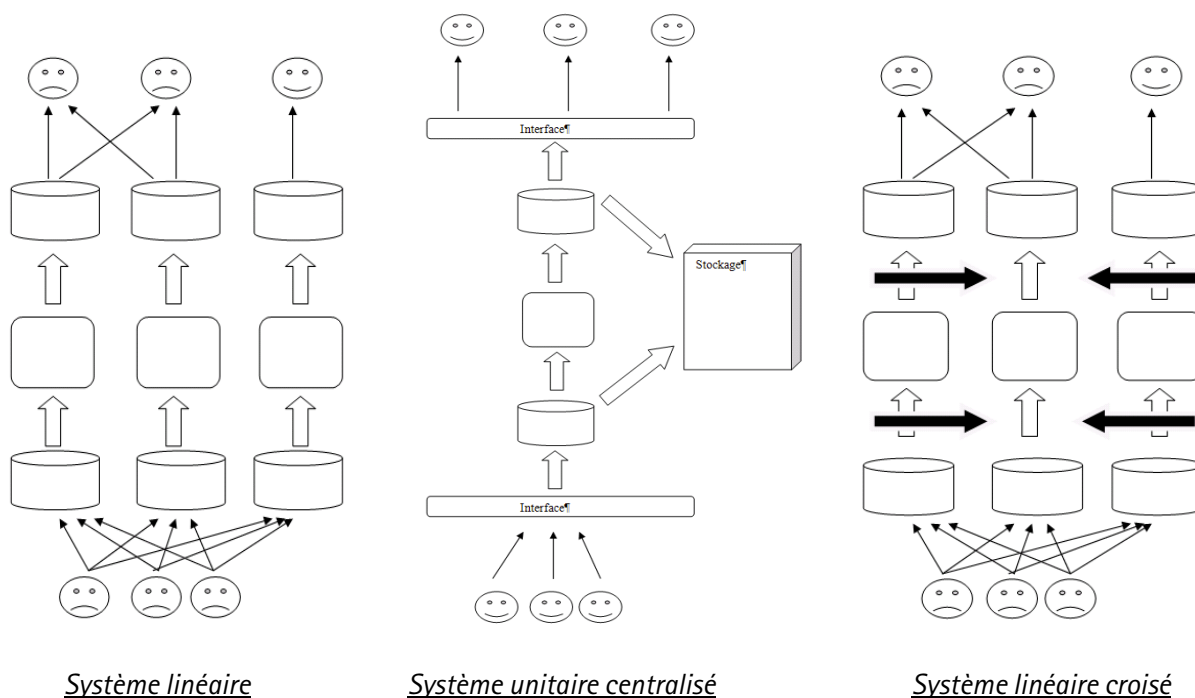
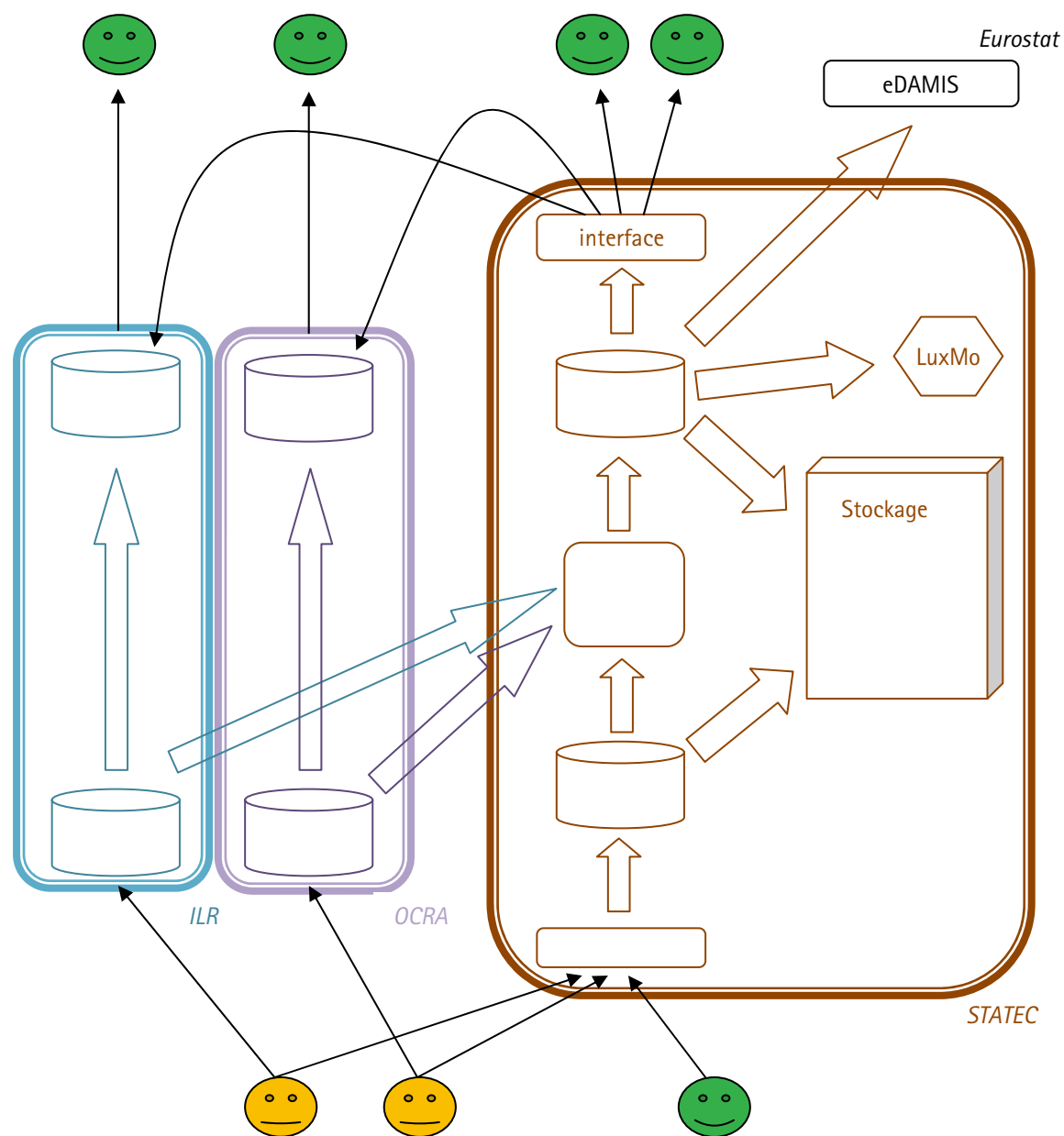


Figure 4: Schéma organisationnel des données établi au Luxembourg



Système partiellement centralisé développé au Luxembourg

6.1.4 La mise en place du système de centralisation de l'information

La mise en place d'un nouveau schéma organisationnel des données et statistiques énergétiques doit nécessairement être précédée d'une réflexion sur l'efficacité et la durabilité d'un système de centralisation de l'information.

A l'instar des systèmes d'information géographique (SIG), les systèmes d'information statistique sont des outils d'aide à la décision qui regroupent plusieurs séries de données ou de statistiques sur base d'un dénominateur commun. Dans le cas des SIG, l'élément comparatif des séries est la localisation spatiale des données. Un système de centralisation de l'information du domaine de l'énergie, quant à lui, compilera des séries de données permettant de répondre aux questions que se posent les utilisateurs en lien aux problématiques énergétiques. En d'autres termes, le dénominateur commun est le produit énergétique.

Les missions d'un système de centralisation de l'information sont diverses et multiples. Trois d'entre elles sont primordiales :

- Faire état des connaissances, c'est-à-dire faciliter le suivi des séries dans le temps, dans l'espace ou selon tout autre élément pertinent ainsi que la qualité statistique de ces séries.
- Permettre l'évaluation, c'est-à-dire faciliter le calcul d'indicateurs qui visent à obtenir une vision rapide et claire des séries.
- Améliorer la communication, c'est-à-dire faciliter les échanges d'information entre les différents acteurs du domaine de l'énergie.

Un système de centralisation de l'information correctement construit et opérationnel est également indispensable à toute analyse prospective.

L'efficacité et la durabilité du système de centralisation ne sont effectives que s'il répond aux utilisations directes et indirectes des acteurs du domaine de l'énergie.

- L'utilisation directe de l'information est immédiate au sein du système de centralisation et se traduit par la compilation de statistiques pour les transmissions internationales, la production de tableaux de bord, de notes statistiques ou conjoncturelles, etc.

- L'utilisation indirecte ne prend de sens que par le biais de l'instauration d'un outil d'interprétation ou instrument comptable : bilan énergétique, comptes sectoriels, efficacité énergétique (énergie et émissions de gaz à effet de serre (GES)), etc.

Avant de définir le cahier des charges d'un système de centralisation de l'information du domaine de l'énergie, il est nécessaire de caractériser les utilisations directes et indirectes du système.

L'utilisation directe du système de centralisation de l'information énergétique correspond typiquement à la transmission de statistiques de l'énergie réclamées par les règlements et directives européens. A titre d'exemple, citons le règlement 1099/2008/CE portant sur les statistiques de l'énergie et la directive 2008/92/CE portant sur la transparence des prix du gaz et de l'électricité aux consommateurs finals. Cette utilisation directe nécessite bien évidemment la connaissance des besoins et des ressources en énergie. Elle exige également la description de l'organisation du secteur énergétique, c'est-à-dire des flux entre besoins et ressources. Cette description implique la connaissance des acteurs en présence, des moyens de production disponibles et des paramètres économiques (coûts et prix). Enfin, elle ne sera complète que si le fonctionnement du secteur énergétique est également caractérisé en termes de fiabilité (par exemple la défaillance de fourniture), de concurrence (entre opérateurs), d'efficacité (par exemple le taux d'utilisation des équipements) et de contrainte (par exemple émissions de gaz à effets de serre).

L'utilisation indirecte du système de centralisation de l'information est intimement liée aux outils d'analyse et d'évaluation mis à disposition. Citons les trois principaux outils implantés dans la majorité des systèmes d'information énergétique :

- *Le bilan énergétique* : représente une description exhaustive et synthétique de l'ensemble des flux énergétiques. Il est par excellence l'instrument de base de la compréhension de la situation énergétique d'un pays. Instrument comptable faisant appel à une même unité de mesure des flux énergétiques (par exemple la tonne équivalent pétrole ou le térajoule), le bilan permet les agrégations entre formes d'énergie différentes et autorise les comparaisons entre filières et les analyses de structure interne du système énergétique (structure des flux).

- *Les comptes sectoriels*: permettent une approche plus fine d'analyse des besoins énergétiques alors que le bilan énergétique informe sur les grands postes de consommation d'un pays. Les séries de données nécessaires à cette analyse proviennent généralement d'enquêtes spécifiques auprès d'un secteur particulier ou d'estimations sur base d'enquêtes dérivées (par exemple, la consommation d'électricité pour la réfrigération et la cuisson pourra être estimée en couplant le plan tarifaire de chacun des ménages et la composition en appareils ménagers de ceux-ci).
- *Les indicateurs*: relient entre elles deux informations pour dériver une information complémentaire (indicateurs de structure d'approvisionnement, de parts de marché des énergies, de sécurité énergétique, d'efficacité énergétique, etc.).

Enfin, tout système de centralisation de l'information doit s'appuyer sur une base de données informatique unique et correctement structurée, permettant le stockage et le traitement des requêtes des utilisateurs.

6.1.5 La création de la base de données

La base de données qui soutient le système d'information intégré doit être conçue de manière à faciliter son exploitation. Plusieurs conditions sont initialement fixées :

- Elle doit permettre de garder des traces des opérations menées sans pour autant générer de multiples tables redondantes.
- Son exploitation doit être aisée et indépendante du logiciel d'exploitation utilisé.
- Sa structure doit être facilement compréhensible pour l'utilisateur occasionnel mais également permettre des requêtes complexes pour l'utilisateur avancé.
- Elle doit être conviviale dans son système de chargement au vu des fréquentes mises à jour.
- Elle doit s'intégrer dans le schéma de transmission de données établi entre les partenaires actuels.

Les options choisies lors de la création base de données luxembourgeoise de l'énergie ont permis de répondre à toutes ces conditions.

En effet, les fichiers d'échange de données (FED) entre les acteurs du schéma organisationnel ont été maintenus dans leurs formats actuels à savoir des tableurs contenant des macros VBA de vérification des données. De cette manière la vérification initiale des données à l'encodage est maintenue et l'échange d'informations factuelles sur les données et leurs tendances reste possible.

Les fichiers FED sont ensuite convertis dans un format spécifique à la base de données. Le principe d'une valeur quantitative par ligne est utilisé, et chaque valeur est caractérisée de manière unique. Cette caractérisation se fait *via* la définition du mois, de l'année, d'un type de données (définitive, provisoire, etc.), d'un code par produit, d'un code pour l'unité, d'un code par poste du bilan (importation, stock, etc.), d'un code pour le pays d'origine, d'un code pour le pays d'importation/exportation, d'un identifiant pour la source de données et d'un champ pour les commentaires. Ce format et la standardisation des codes permettent de stocker l'ensemble des données dans une même table.

Les fichiers FED correctement formatés sont transférés (*via* une interface graphique joignable sous n'importe quel logiciel d'exploration d'Internet) sur un serveur spécifiquement dédié et ensuite chargés dans la base de données SQL.

La structure de la base de données (cf. annexe 2) est constituée deux tables de valeurs et de huit tables de dimension, qui permettent de vérifier et de convertir les codes des caractéristiques en intitulés compréhensibles.

Afin de garder le suivi des opérations, des champs comme la date et l'utilisateur qui crée la ligne ainsi que la date et le dernier utilisateur qui modifie la ligne sont ajoutés à chaque ligne. De plus, un champ de type booléen décrit la validité de la valeur. A chaque modification de la valeur de la ligne, c'est en réalité une nouvelle ligne qui est créée et seule la dernière modification est jugée positivement valide. Par ce système, le suivi des ajustements des valeurs est rendu possible.

La base de données est exploitable *via* tous les logiciels de traitement statistique (comme SAS, SPSS, STATA, etc.) mais une interface permettant la création automatique de tables prédéfinies a également été

créée pour l'utilisateur occasionnel. Cette interface permet de sauvegarder les projets établis par l'utilisateur pour une application ultérieure.

Une sauvegarde journalière de la base de données et du « cube » servant à l'interface de création de rapports a été planifiée. Le « cube » n'est rien d'autre qu'un agencement multidimensionnel de la base de données selon les caractéristiques de chacune des valeurs.

Toutes les données du domaine de l'énergie sont stockées dans ces deux tables de valeurs : l'une contenant les données mensuelles, l'autre les données annuelles. La caractérisation unique de chaque valeur permet de stocker des données en volume et en valeur dans la même table. De même, les données en volume peuvent être entre autres des quantités (importations, stocks, etc.), des capacités techniques (puissance de centrale, longueur de réseau électrique, etc.), des facteurs de conversion (pouvoir calorifique inférieur et supérieur, etc.).

De plus, la caractérisation unique n'est nullement limitative puisqu'il a été prévu d'attribuer un type à chacune des valeurs. Actuellement, quatre types sont possibles : définitive, provisoire, estimée ou prévisionnelle. Un cinquième type « historique ajustée » est à l'étude pour les opérations de normalisation sur les données, comme par exemple la normalisation au degré-jour.

Les sommes et les indicateurs calculés ne sont pas stockés dans les tables de valeurs mais générés, à chaque lancement, dans les requêtes sauvegardés. De cette manière, tous les résultats générés bénéficient en toute situation des données les plus actualisées.

Une base de données selon les règles de l'art permet de découpler le potentiel du système de centralisation de l'information. Elle permet une exploitation rapide, efficace et reproductible des données énergétiques.

De plus, l'utilisation de métadonnées sur les valeurs (notamment les dates de création et dates de modification) permet l'historisation des opérations de chargement et de validation des valeurs.

6.2 La production des statistiques

6.2.1 La production de bilans énergétiques (en volume et valeur) et le développement de méthodes d'estimation

La flambée des prix du pétrole au cours des trois premiers trimestres de l'année 2008 et les problèmes d'acheminement du gaz russe en début d'année 2009 ne sont que des exemples de l'importance du domaine énergétique pour notre société et surtout de notre dépendance envers ces produits énergétiques.

Afin d'anticiper et de gérer les incidences économiques et sociales liées aux problématiques de l'énergie, la production de statistiques énergétiques fiables est indispensable.

Un système de centralisation de l'information du domaine de l'énergie devrait permettre d'apporter des arguments chiffrés à un nombre de problématiques comme :

- Notre dépendance à l'égard des importations d'énergie ;
- La part du commerce de produits énergétiques non surveillés (bois de chauffage, ...) ;
- L'impact des sources d'énergies renouvelables sur le marché ;
- La répartition de la consommation entre secteurs et au sein de chaque secteur ;
- Les champs d'amélioration de l'efficacité énergétique ;
- La demande future en énergie ;
- L'importance de l'autoproduction industrielle et résidentielle ;
- Les effets directs et indirects des changements de prix des différentes énergies ;
- L'impact de notre consommation de combustibles fossiles sur l'émission de gaz à effet de serre.

Cette liste est loin d'être exhaustive, mais elle suffit pour attirer l'attention sur toute une série de défis statistiques qu'il faut résoudre lors de l'élaboration d'un système intégré de statistiques de l'énergie.

L'outil le plus adapté pour présenter l'ensemble des données énergétiques en volume est le 'bilan énergétique'.

Le bilan énergétique : intérêts et difficultés

Un bilan énergétique global fait apparaître dans un cadre comptable cohérent les stocks et les flux de toutes les formes d'énergie depuis leur origine jusqu'à leur utilisation finale. Un tel bilan fournit une base d'analyse et de synthèse rigoureuse qui assure précisément le respect du principe de cohérence. Si elle ne remplace pas des statistiques détaillées sur certaines thématiques comme la répartition des consommations électriques au sein des ménages, la compilation d'un bilan énergétique global constitue un test clé de la cohérence interne et externe, tant à l'intérieur de bilans distincts des produits énergétiques qu'entre ceux-ci. Un tel bilan énergétique constitue en même temps un cadre indispensable pour l'élaboration d'analyses plus poussées.

On a l'habitude de produire des bilans énergétiques portant sur une seule source d'énergie (par exemple, le gaz naturel) ou sur un groupe de produits énergétiques étroitement liés (par exemple, les produits pétroliers). De tels bilans indiquent correctement les origines (production, importation, prélèvement de stock) et les utilisations (exportation, augmentation des stocks, consommation), mais ils font généralement abstraction des processus de transformation. En effet, les processus de conversion d'énergie (par exemple, de gaz en électricité) sont considérés comme une consommation parmi d'autres.

L'intérêt d'un bilan énergétique global est d'intégrer la perception de ces conversions d'énergie au sein de la comptabilité des approvisionnements et des utilisations de toutes les sources d'énergie employées dans le pays.

a) Le choix de l'unité commune

Cette comptabilité commune de tous les produits énergétiques soulève par ailleurs la question des conversions d'unités de mesures et de la comparabilité des bilans de plusieurs années. En effet, le pouvoir calorifique de certains produits énergétiques varie selon le pays producteur et selon l'année de production. L'unité choisie pour établir le bilan énergétique global se doit de tenir compte de cette variabilité. L'unité habituellement choisie est une unité de mesure calorifique ; le gigajoule (GJ). Dans le cadre de l'établissement du bilan énergétique du Luxembourg, nous avons pris le parti de constituer le bilan en unités physiques spécifiques aux produits et postes du bilan et ensuite de convertir l'ensemble des valeurs en unité universelle (GJ) ainsi qu'en unité historique (tep). Cette conversion implique l'utilisation

des pouvoirs calorifiques spécifiques à chaque produit et origine (quand cela est possible). Selon les exigences des règlements et directives européennes, les pouvoirs calorifiques inférieurs ou supérieurs sont utilisés.

b) La correction de degré-jour

La comparabilité des bilans sur plusieurs années est également impactée par les conditions climatiques. Un hiver très froid augmentera significativement les dépenses énergétiques de chauffage au sein des ménages, commerces et services publics. Afin d'évaluer l'impact des politiques énergétiques, il est nécessaire de tenir compte des conditions climatiques par l'application d'un coefficient de correction lié aux degrés-jour par rapport à une valeur dite normale.

c) Le principe coût/avantage

Dans ce bilan énergétique global, toutes les sources d'énergie devraient apparaître : les combustibles fossiles mais également les énergies renouvelables par la production d'électricité et/ou de chaleur.

Cela dit, s'il est facile d'insérer dans un bilan global des lignes et des colonnes qui, si elles étaient remplies, fourniraient de précieux renseignements à certains spécialistes, il est important de tenir compte du coût engendré par la collecte et le traitement de l'information (coût financier pour le commanditaire de l'enquête et coût temporel pour le fournisseur de données). Il faut donc trouver un point d'équilibre (au sens coût/avantages) entre le caractère plus ou moins complet, plus ou moins précis du bilan énergétique et le coût de l'opération de production de ce bilan ainsi que de celui des erreurs entraînées par un système de données imparfait sur la prise de décision. Le bilan énergétique nouvellement établi pour le Luxembourg présente trente produits énergétiques regroupés en six vecteurs énergétiques au travers de soixante-quatre postes répartis en trois modules : approvisionnement, transformation, consommation.

d) La qualité des données

Outre le choix des postes et produits à inclure dans le bilan énergétique global, la qualité des statistiques produites relève d'une grande importance.

En principe, l'approvisionnement et la transformation des produits énergétiques sont bien surveillés (exception faite des biomasses), ce qui permet la production de statistiques relativement exactes

(encore faut-il que la définition de la variable du bilan corresponde à la définition utilisée pour la surveillance).

Par contre, la consommation finale est bien moins connue. Le recours à des enquêtes spécifiques est habituel dans ce domaine. Mais ces enquêtes ne couvrent généralement pas l'ensemble des secteurs de la consommation et leurs marges d'erreur rendent quasi impossible l'obtention d'un bilan équilibré.

La démarche choisie pour le bilan énergétique du Luxembourg est d'une part d'utiliser pleinement les résultats des enquêtes pour la compilation des postes du module consommation et d'autre part, pour les postes dont on ne dispose pas d'information, de répartir le solde des livraisons nettes à la consommation intérieure selon une clé de répartition inspirée de la comptabilité nationale de l'année. Cette méthode de répartition implique un équilibrage automatique du bilan.

e) L'usage non énergétique

Certaines sources d'énergie primaire (par exemple, le charbon ou le gaz naturel) et certaines sources d'énergie dérivée (par exemple, le naphte) peuvent être utilisées soit comme combustibles, soit comme matières premières par l'industrie pétrochimique. Certains produits pétroliers (par exemple, le bitume, les lubrifiants et les cires) ne sont presque jamais utilisés comme combustibles.

Le bilan énergétique veille à scinder l'utilisation énergétique et celle non-énergétique des produits comptabilisés.

f) Le marché non surveillé

Une partie de l'approvisionnement ou de la transformation de produits énergétiques n'est pas suivie explicitement. L'exemple le plus connu est la production et la vente de bois de chauffage entre particuliers en région rurale. Bien que non surveillé, le marché du bois de chauffage est loin d'être négligeable. Des méthodes indirectes d'estimation, sur base des résultats de l'enquête sur les budgets des ménages, ont été conçues pour évaluer ce marché.

g) L'autoproduction d'énergie

Un autre cas est l'autoproduction de chaleur, généralement par les particuliers, au moyen de panneaux solaires thermiques ou de pompes à chaleur.

Dans la mesure où cette énergie se substitue à celle qui était précédemment achetée sur le marché, une analyse rapide du bilan arriverait à la conclusion que la demande en énergie des particuliers diminue. Or ce n'est pas le cas ! Le bilan énergétique intègre l'autoproduction privée d'énergie au même titre que la production dite « publique ».

6.2.2 La production des questionnaires internationaux

L'une des principales tâches de type « production » dévolue au STATEC dans le domaine des statistiques de l'énergie est la collecte, le traitement et la transmission des données mensuelles et annuelles à Eurostat selon le règlement européen 1099/2008/CE.

Les données mensuelles nécessaires à ces transmissions proviennent, depuis la mise en place du nouveau schéma organisationnel, exclusivement de l'ILR, de l'OCRA et du STATEC, qui sont désignés comme les organismes collecteurs des variables de production et importations/exportations des vecteurs énergétiques, respectivement : gaz naturel, électricité, produits pétroliers et combustibles solides.

En principe, les fichiers FED contenant l'ensemble des variables utiles sont envoyés au 20^{ème} jour de chaque mois. Les fichiers sont alors sauvegardés (avec création d'un métafichier) et compilés dans des fichiers de synthèse dans la base de données de l'énergie.

Aux données mensuelles viennent s'ajouter les données collectées annuellement par l'Administration de l'environnement et le Service d'économie rural (SER).

Les données mensuelles et annuelles ainsi collectées permettent la compilation des questionnaires requis par le règlement européen, à savoir:

- *Les transmissions mensuelles à court terme en m+1* : concernent électricité, gaz naturel et produits pétroliers).
- *Les transmissions mensuelles en m+3* : concernent électricité, gaz naturel, produits pétroliers et combustibles solides.
- *Les transmissions annuelles en a+4 et en a+11* : concernent électricité et chaleur, gaz naturel, produits pétroliers, combustibles solides et énergies renouvelables.

A ces transmissions internationales viennent s'ajouter des transmissions régulières à différentes unités du STATEC (comptes nationaux, balance des paiements, statistiques à court terme).

Au vu de la multiplicité des transmissions, un outil de suivi a été développé. (cf. figure 5).

Afin de garantir la continuité de la compilation des données, un manuel méthodologique spécifique a été créé. Il décrit les opérations de traitement des données énergétiques collectées dans l'objectif de compiler les statistiques nécessaires aux questionnaires requis par le règlement européen.

Pour chaque variable au sein de chaque questionnaire sont présentés les sources de données et les éventuels traitements statistiques réalisés. Occasionnellement un paragraphe spécifique est rédigé sur des probléma-

tiques ponctuellement rencontrées lors du remplissage des questionnaires.

Toute phase de transmission nécessite une évaluation régulière de la qualité de la transmission des statistiques produites. Habituellement cette analyse de la qualité se base sur la caractérisation spécifique de critères comme : la précision, l'actualité et la ponctualité, la comparabilité, la cohérence des statistiques produites, la pertinence, l'accessibilité et la clarté.

Dans le cadre d'une évaluation des transmissions mensuelles et mensuels à court terme, les quatre premiers critères ont été retenus (voir encadré 3). Le critère de pertinence n'a pas été choisi de par le fait que les transmissions étant obligatoires, ce point ne peut être débattu.

Figure 5: Outil pour lister les transmissions de données à réaliser

Les reportages portant sur les statistiques de l'énergie en volume

Quels sont les reportages à réaliser ce mois-ci ?

Imprimer

Questionnaire	Nom du fichier	Echéance	Compilation	Rapporteur actuel	Adresse d'envoi à Eurostat	DATASET ID	Adresse d'envoi à AIE
Questionnaire annuel Eurostat/AIE selon Règlement 1099/2008/CE							
Q annuel combustibles solides	Coalques.xls	a+11 m		STATEC - Mr Thunus	via eDAMIS	ENERGY_SOLID_A	Webmail AIE ou Email coalag@iea.org
Q annuel gaz naturel	Gasques.xls	a+11 m		STATEC - Mr Thunus	via eDAMIS	ENERGY_NTGAS_A	Webmail AIE ou Email gasag@iea.org
Q annuel électricité et chaleur	Eleques.xls	a+11 m		STATEC - Mr Thunus	via eDAMIS	ENERGY_ELECT_A	Webmail AIE ou Email eleag@iea.org
Q annuel produits pétroliers	Oilques.xls	a+11 m		STATEC - Mr Thunus	via eDAMIS	ENERGY_PETRO_A	Webmail AIE ou Email oilag@iea.org
Q annuel énergies renouvelables	Renques.xls	a+11 m		STATEC - Mr Thunus	via eDAMIS	ENERGY_RENEW_A	Webmail AIE ou Email renevaq@iea.org
Questionnaire mensuel Eurostat/AIE selon Règlement 1099/2008/CE							
Q mensuel combustibles solides	Monthly coal questionnaire.xls	m+3	code "Coal" - projet SAS "code_rapportage.epg"	ST.		ENERGY_SOLID_M	
Q mensuel gaz naturel	MOS.xls	m+3 / m+2	code "MOS_gas" - projet SAS "code_rapportage.epg"	ST.		ENERGY_MOSPG_M	Email AIE: mos@iea.org
Q mensuel électricité	Elecmnthquest2001.xls	m+3	code "Elecmnth" - projet SAS "code_rapportage.epg"	ST.		ENERGY_ELEC3_M	
Q mensuel produits pétroliers	MOS.xls	m+3 / m+2	code "MOS_oil" - projet SAS "code_rapportage.epg"	ST.		ENERGY_MOSPG_M	Email AIE: mos@iea.org
Questionnaire mensuel à court terme							
Q gaz naturel	SEG_gas.doc	m+1	code "SEG_gas" - projet SAS "code_rapportage.epg"	ST.		ENERGY_NTGAS_M	
Q électricité	SEG_ele.doc	m+1	code "SEG_ele" - projet SAS "code_rapportage.epg"	ST.		ENERGY_ELEC1_M	
Q produits pétroliers	MAXI_JODI.xls	m+25j	code "JODI" - projet SAS "code_rapportage.epg"	ST.		ENERGY_PETRO_M	Email AIE: mos@iea.org
Q gaz naturel	Gas_JODI_OCDE.xls	m+25j	code "GAS_JODI" - projet SAS "code_rapportage.epg"	ST.			Email AIE: mos@iea.org
Rapportage à Eurostat selon la Directive 2004/8/CE (cogénération)							
Tables EU-1 et EU-2	compris dans: Eleques.xls	a+11 m		ST.			
Rapportage à DG ENER selon le Règlement 405/2003/CE							
Q steam coal et coking coal	Coal prices.doc	s+2 m		ST.			
Q approv et besoin en combustibles solides	imports.xls	a+2 m		ST.			
Mini-questionnaire annuel AIE							
Combustibles solides	minquestcoal2006.xls	a+4 m		STATEC - Mr Thunus			Webmail AIE
Gaz naturel	minquestgas2006.xls	a+4 m		STATEC - Mr Thunus			Webmail AIE

Microsoft Excel

LES QUESTIONNAIRES MENSUELS

Rapportage m+1 du mois de référence : octobre

SEG_gas.doc

SEG_ele.doc

MAXI_JODI.xls (Jour 25)

Gas_JODI_OCDE.xls (Jour 25)

Rapportage m+2 du mois de référence : septembre

MOS.xls à AIE (Jour 25)

LIQUIDES statec.xls

séries_pour_UC4.xls

Rapportage m+3 du mois de référence : août

Monthly coal questionnaire.xls

MOS.xls à Eurostat

Elecmnthquest2001.xls

LES QUESTIONNAIRES ANNUELS

Coalques.xls

Gasques.xls

Eleques.xls

Oilques.xls

Renques.xls

OK

Encadré 3: La méthodologie pour l'évaluation des transmissions de données

A) le critère de précision est défini, dans cette analyse, comme une comparaison des statistiques estimées aux statistiques définitives. En effet, les questionnaires mensuels à court terme ($m+1$) requièrent la transmission de variables statistiques alors que les données nécessaires à leurs calculs ne sont pas encore disponibles. Habituellement ces variables estimées sont qualifiées de provisoires. Les valeurs définitives de ces variables seront ensuite calculées suite à la mise à disposition des données. Les nouvelles valeurs sont rapportées via les questionnaires mensuels ($m+3$).

Au cours de cette analyse, le critère de précision se traduit par le calcul de l'erreur entre les valeurs estimées et les valeurs définitives.

B) le critère d'actualité et de ponctualité se définit ici comme la mise à disposition de statistiques récentes et dans les délais impartis par les textes législatifs. Dans le cadre du nouveau règlement européen sur les statistiques de l'énergie, l'actualité des statistiques et les délais de transmission sont imposés. L'étude de ce critère permet d'évaluer le respect de ces contraintes.

Le critère d'actualité et de ponctualité sera analysé par la catégorisation des délais des transmissions effectives.

C) le critère de comparabilité est généralement défini comme la possibilité de comparer les statistiques produites aussi bien dans le temps, qu'entre zones géographiques différentes ou qu'entre sous-domaines. Le manque de comparabilité provient habituellement de différences dans la méthode de collecte et de traitement des données.

Dans le cadre de cette analyse, le critère de comparabilité se concentrera sur l'évaluation de la cohérence des séries chronologiques entre les six premiers mois de 2008 et les six premiers mois de 2009. L'objectif de l'étude de ce critère se limite à évaluer l'équivalence des méthodes et procédures de traitement des données énergétiques appliquées en 2008 par la Direction de l'Energie et l'OCRA et en 2009 par le Statec.

Le critère de comparabilité se traduira par la recherche de biais systématiques entre les deux séries.

D) le critère de cohérence se définit comme la possibilité de combiner plusieurs sources de données. Dans le cadre de la présente analyse, le champ d'étude de ce critère a été limité au suivi de la cohérence entre les statistiques provisoires et définitives.

Au cours de cette analyse, le critère de précision se traduit par le calcul de l'erreur entre les valeurs provisoires et les valeurs définitives.

De même, les questionnaires à rapporter à Eurostat étant standardisés, le critère de l'accessibilité et de la clarté des statistiques produites n'est pas un objet d'intérêt dans cette évaluation.

L'objectif de ce type d'évaluation est de juger la mise en place du nouveau schéma organisationnel de collecte et de transmission des données énergétiques au Luxembourg dans le cadre de sa mission de transmission de données suite au règlement européen. Cette analyse a mis en exergue de petits dysfonctionnements pour l'un des vecteurs énergétiques. Ces dysfonctionnements ont été corrigés en vue d'optimiser les processus de diffusion des statistiques énergétiques à Eurostat.

6.2.3 La production de tables spécifiques

Selon le nouveau schéma organisationnel de la collecte, du traitement et de la diffusion des statistiques de l'énergie, il est possible d'identifier trois catégories de partenaires :

1. L'ILR, l'OCRA, l'Administration de l'environnement et le STATEC sont des partenaires spécifiquement concernés par la collecte des données énergétiques.
2. Le STATEC est le partenaire central du schéma, il centralise les données, compile les statistiques énergétiques, vérifie leurs cohérences et diffuse ces statistiques.
3. Le Ministère du développement durable et la Direction de l'énergie sont des partenaires spécifiquement concernés par l'utilisation finale et l'analyse des statistiques de l'énergie.

La mission du STATEC, outre le stockage et la vérification de la qualité des données collectées, réside dans la production de statistiques énergétiques utiles aux utilisateurs finaux. Ces utilisateurs finaux sont généralement caractérisés selon leur utilisation des statistiques :

- Utilisation à des fins d'informations : le grand public ;
- Utilisation à des fins décisionnelles : les entreprises ;
- Utilisation à des fins politiques ou de gestion : les ministères et administrations.

Pour les deux premiers groupes, l'intérêt réside généralement dans l'évolution des tendances de certains chiffres et/ou indicateurs.

L'optique du troisième groupe est de remplir ses objectifs, à savoir la définition de politiques (environnementales et énergétiques) et le suivi des plans de gestion et/ou prévisionnels. Ce troisième groupe présente donc une série de besoins en statistiques, généralement quantitatives, pour accomplir ses propres missions.

Concernant les statistiques énergétiques, les besoins spécifiques du Ministère du développement durable et de la Direction de l'énergie ont été inventoriés. Ensuite, l'ensemble des variables du domaine de

l'énergie nécessaires à la confection des transmissions de données de ces besoins ont été identifiées et listées dans le dictionnaire des variables du nouveau schéma organisationnel.

Il est important de noter que les tableaux produits doivent rester cohérents avec les autres statistiques diffusées auprès d'autres utilisateurs. Pour ce faire, on travaille selon une approche hiérarchique. On produit une matrice dont les lignes correspondent aux titres et sous-titres du bilan et dont les colonnes correspondent aux produits énergétiques et vecteurs énergétiques. Cette manière de présenter le bilan permet de voyager d'un niveau à l'autre de la hiérarchie selon les besoins.

Cette approche implique la définition d'un tableau synthétique mais holistique reprenant les principales statistiques énergétiques du Luxembourg. Ce tableau figure au sommet de notre hiérarchie. Par la suite, des tableaux plus précis et parfois spécifiques à certains secteurs sont produits et viennent s'imbriquer sous le tableau synthétique

Dans la pratique, les tableaux de type « bilan énergétique » conviennent parfaitement à cette approche.

Au premier niveau de notre hiérarchie (le sommet), on trouve un bilan énergétique ne reprenant que les grands titres (production/importation/exportation/stock/livraison brute/livraison nette) et tous les produits confondus.

Au deuxième niveau, on trouve un bilan ne reprenant que les grands titres mais scindés par vecteurs énergétiques (gaz naturel, électricité, produits pétroliers, combustibles solides, énergies renouvelables).

Au troisième niveau, on trouve des tableaux spécifiques à chaque grand titre (par exemple, les importations par pays, les consommations finales par secteur,...) et scindés par produits énergétiques (essence, diesel,...).

Au quatrième niveau, on trouve les tableaux les plus spécialisés : consommation finale de produits énergétiques au sein de chaque secteur (par exemple pour le transport : ferroviaire, routier, aérien, navigation).

Toute la difficulté réside dans l'adéquation du niveau de détail aux besoins du Ministère du développement durable et de la Direction de l'énergie.

De plus, les délais de transmission, qui jalonnent les missions de ces deux partenaires privilégiés, impliquent la diffusion en deux étapes :

- Une première diffusion – pour la fin avril – des statistiques sur l'approvisionnement du Luxembourg en produits énergétiques ;
- Une deuxième diffusion – pour la fin septembre – des statistiques complètes (approvisionnement et consommation).

Il est important de noter qu'au vu des méthodes d'estimation utilisées pour la partie « consommation » du bilan énergétique, une version consolidée du bilan sera produite 12 mois après la première diffusion du bilan.

6.3 La vérification de la cohérence

6.3.1 L'analyse de la qualité des données

L'évaluation de la qualité des données collectées est un élément fondamental du processus de production de statistiques car il permet d'estimer la validité des données vis-à-vis de la population observée.

Dès à présent, la distinction est à faire entre un exercice d'évaluation de la qualité des statistiques et un exercice d'évaluation de la qualité des données.

- Le premier vise à analyser l'adéquation entre les besoins des utilisateurs et les statistiques produites par le statisticien, y compris l'incertitude sur les statistiques fournies. Habituellement cette analyse se base sur la caractérisation spécifique de critères comme : la pertinence, la précision, l'actualité et la ponctualité, l'accessibilité et la clarté, la

comparabilité et la cohérence des statistiques produites.

- Le deuxième exercice vise à concentrer son analyse sur les données collectées qui permettront de calculer les statistiques. Les critères retenus pour cette analyse sont la précision, la comparabilité et la cohérence des données collectées.

Une évaluation de la qualité des données collectées a été réalisée sur l'ensemble des sources permettant de répondre aux besoins du règlement européen 1099/2008/CE portant sur les statistiques de l'énergie. Les critères de précision, comparabilité et cohérence ont été qualifiés par la recherche d'erreurs ou de biais.

Le STATEC a également collecté, pour le Luxembourg, le maximum des données énergétiques existantes pour les années antérieures à 2009.

L'objectif majeur de cette démarche était de reconstituer des séries historiques continues pour le plus grand nombre possible de variables et de produits énergétiques. Toutes les données collectées ont été intégrées à la base de données de l'énergie du STATEC.

En l'absence de centralisation des données de l'énergie, il apparaît qu'une bonne partie des données historiques (avant 2000) ne sont plus disponibles au sein des administrations luxembourgeoises.

Cependant, afin de produire des analyses tendanciellles dans le domaine de l'énergie, il est important de reconstituer des séries chronologiques remontant à des dates les plus reculées possibles.

Sachant que le Luxembourg communiquait, sous *Gentlemen Agreement*, certaines statistiques de l'énergie à Eurostat et à l'AIE, un exercice de récupération et de validation de ces données a été réalisé.

Encadré 4: Les critères d'évaluation de la qualité des données

A/ Le critère de précision peut se définir comme la proximité entre la valeur retenue et la vraie valeur, mais inconnue, pour l'ensemble de la population. L'analyse de la précision se résume au calcul du biais et de la variabilité de l'échantillon. Elle est structurée par l'évaluation des erreurs liées et non-liées à l'échantillonnage.

Les erreurs examinées dans le cadre de la collecte de données énergétiques ont été :

- L'erreur d'échantillonnage ;
- L'erreur de couverture ;
- L'erreur de mesure ;
- L'erreur de traitement ;
- L'erreur de non-réponse ;
- L'erreur d'hypothèse du modèle.

B/Le critère de comparabilité se définit comme la possibilité de comparaison des données collectées aussi bien dans le temps, qu'entre zones géographiques différentes, qu'entre des sous-domaines. Le manque de comparabilité provient généralement d'une différence dans la définition de

l'échantillon (population de référence, caractéristiques, classification,...) ou d'une différence dans la méthode d'analyse. Seules les deux premières comparabilités ont été étudiées :

- La comparabilité dans le temps ;
- La comparabilité dans l'espace.

C/ Le critère de cohérence se définit comme la possibilité de combiner plusieurs sources de données. Différents types de cohérence sont examinés :

- La cohérence entre les statistiques provisoires et définitives ;
- La cohérence entre les statistiques à court terme et à long terme ;
- La cohérence entre les mesures différentes d'un même domaine.

Au vu de la spécificité du domaine de l'énergie au Luxembourg, certaines de ces erreurs (comparabilité et cohérence) sont caduques. Pour exemple, l'erreur de couverture est considérée comme nulle si l'échantillonnage couvre toute la population.

6.3.2 La cohérence des séries chronologiques

L'évaluation de la cohérence de séries chronologiques peut porter sur différents éléments. Deux de ces éléments ont actuellement été travaillés :

- La recherche de valeurs suspectes ou aberrantes ;
- L'étude des tendances et saisonnalités des séries.

La recherche de valeurs suspectes ou aberrantes

Trois approches différentes ont été appliquées pour évaluer les séries chronologiques et si possible détecter les valeurs suspectes et aberrantes :

- L'approche graphique, utilisée par exemple par l'AIE pour vérifier la cohérence des statistiques énergétiques envoyées via les questionnaires mensuels.

- L'approche statistique *a priori*, utilisée actuellement dans les fichiers FED (valeurs aberrantes et double interquartile).

- L'approche statistique *a posteriori*, utilisée pour l'analyse de longues séries chronologiques.

Une fois identifiées, les valeurs suspectes doivent être examinées par le statisticien pour évaluer leur plausibilité et déterminer le traitement à appliquer à ces valeurs suspectes. Certaines valeurs ont été identifiées comme aberrantes, mais sont tout à fait exactes comme par exemple lors d'un arrêt de production d'une centrale pour problème technique. Si la valeur est correcte, il faut évidemment la garder. Dans le cas inverse, un travail d'imputation au 1^{er} ou 99^{ème} centile sera réalisé.

En effet, en cas de valeurs aberrantes avérées, deux techniques s'offrent au choix du statisticien : soit l'élimination, soit l'imputation. L'élimination est souvent utilisée pour les séries chronologiques constituées d'un grand nombre de données. L'imputation est généralement utilisée pour compléter

les valeurs manquantes, mais elle peut aussi être utilisée pour corriger les valeurs aberrantes. L'expérience a montré qu'un remplacement par la moyenne est rarement probant, l'imputation au 1^{er} ou 99^{ème} centile donne généralement de meilleurs résultats.

a) La méthode appliquée par l'AIE – L'approche graphique

L'AIE utilise cette méthode dans le cadre de l'évaluation de la cohérence des données transmises dans des délais mensuels à court terme ($m+1$) et celles transmises dans des délais mensuels classiques ($m+3$).

Le principe de cette approche consiste à comparer les données une à une sur un graphique ayant pour abscisse les mois consécutifs et en ordonnée les quantités. Sur base de la courbe des valeurs définitives (les valeurs transmises en $m+3$), une bande de couleur est dessinée pour représenter l'amplitude de la différence acceptée entre les deux jeux de données qui sont comparés. La largeur de cette bande est fixée arbitrairement. L'AIE travaille habituellement avec un écart admis de 3 % en positif et en négatif, soit une amplitude de 6% centrée sur la valeur définitive.

Cette méthode peut également se décliner sous forme de tableau en calculant le pourcentage de différence entre la donnée provisoire et la donnée définitive ou d'une représentation graphique du tableau des pourcentages avec en abscisse les mois consécutifs et en ordonnée le pourcentage d'écart entre la valeur provisoire et définitive. Une bande grisée centrée sur la valeur nulle symbolise l'amplitude habituellement admise (3% négatif et positif).

L'avantage indéniable de cette approche est d'obtenir une vision rapide de la cohérence des deux séries de données et d'identifier très rapidement les mois où la différence est problématique. Cette méthode peut également permettre de détecter des valeurs aberrantes.

Cette méthode présente un inconvénient de taille à savoir qu'elle n'est efficace que si les valeurs comparées sont grandes. Dans le cas de valeurs proches de l'unité, cette analyse devient vite non utilisable. Par exemple, pour un pays comme le Luxembourg, les variations de stocks en produits pétroliers sont généralement faibles voir nulles. La proximité de ces valeurs à l'unité donne, dans le cas d'une unité de différence entre les données, des pourcentages de l'ordre des 100% voire plus.

Cet inconvénient est également lié à l'inadéquation des unités utilisées dans les transmissions internationales en comparaison des quantités utilisées au Luxembourg. Pour exemple, les unités des questionnaires mensuels sont exprimées en 1000 tonnes alors que les livraisons intérieures de la classe « autres kérosènes » ne dépassent pas les 466 tonnes ou encore que les livraisons intérieures de la classe « lpg » fluctuent mensuellement de 540 à 1381 tonnes.

Un autre inconvénient de la méthode est qu'elle ne peut être appliquée que sur deux séries totalement comparables. Cette méthode ne permet donc pas la recherche de valeurs aberrantes sur une seule série. De plus, une valeur aberrante présente dans les deux séries n'est pas détectée. Elle ne permet pas non plus d'évaluer la variabilité des données au sein de la série chronologique.

b) La méthode appliquée dans les fichiers FED – L'approche statistique *a priori*

Dans le cadre des fichiers d'échange de données énergétiques entre les organismes collecteurs et le STATEC, une méthode statistique a été mise au point pour identifier, à l'encodage (c'est-à-dire *a priori*), les valeurs aberrantes ou suspectes.

Cette méthode consiste à vérifier la cohérence de la donnée encodée à deux niveaux.

Le premier niveau de vérification permet de s'assurer que la donnée est contenue dans un intervalle prédéfini. Cet intervalle, décrit par ses limites inférieures et supérieures, est déterminé sur base du calcul de l'amplitude des données récoltées les trois dernières années et du principe que les valeurs aberrantes correspondent à moins d'un pourcent d'une population (principe édicté dans la norme ISO05725). Les limites de l'intervalle sont donc fixées en soustrayant 0,5% de l'amplitude aux valeurs extrêmes (minimum et maximum) de la population récoltée sur les trois dernières années.

Le deuxième niveau de vérification permet de s'assurer que la variation entre deux données consécutives n'est pas suspecte. Cette variation est jugée suspecte si elle est supérieure au double de l'écart-type de la population récoltée sur les trois dernières années. Chaque donnée encodée doit donc être contenue dans un intervalle dont les limites correspondent à la somme (limite supérieure) et à la soustraction (limite inférieure) du double écart-type à la donnée encodée le mois précédent.

L'avantage de cette méthode à deux niveaux est d'identifier *a priori* les valeurs aberrantes et suspectes et demander une vérification directe de la donnée lors de l'encodage. Dans le cadre de son application actuelle, elle permet donc de vérifier, à l'encodage, les erreurs de frappe et de demander un commentaire à l'encodeur sur des valeurs qui semblent suspectes.

Cette méthode est adaptée à l'analyse d'une seule série chronologique et s'ajuste à l'unité des données récoltées. De plus, elle est relativement simple à mettre en place. Les statistiques calculées sur la série chronologique sont des plus classiques. Cela dit, elle a l'inconvénient de se baser sur l'hypothèse de la normalité de la population des données et néglige l'impact de la présence de valeurs aberrantes dans le calcul de l'amplitude et de l'écart-type.

Pour pallier à ce dernier inconvénient, une méthode plus robuste aux valeurs aberrantes a été choisie. La méthode du double écart-type a été remplacée par une approche utilisant les quartiles.

L'approche des quartiles se base sur le calcul des quartiles 0.25 et 0.75 et de l'interquartile ($iqr = q3 - q1$). La règle suivante est établie pour ce calcul :

*Si valeur est supérieur à $q3 + (1.5 * iqr)$ ou si valeur est inférieure à $q1 - (1.5 * iqr)$, alors la valeur est considérée comme aberrante.*

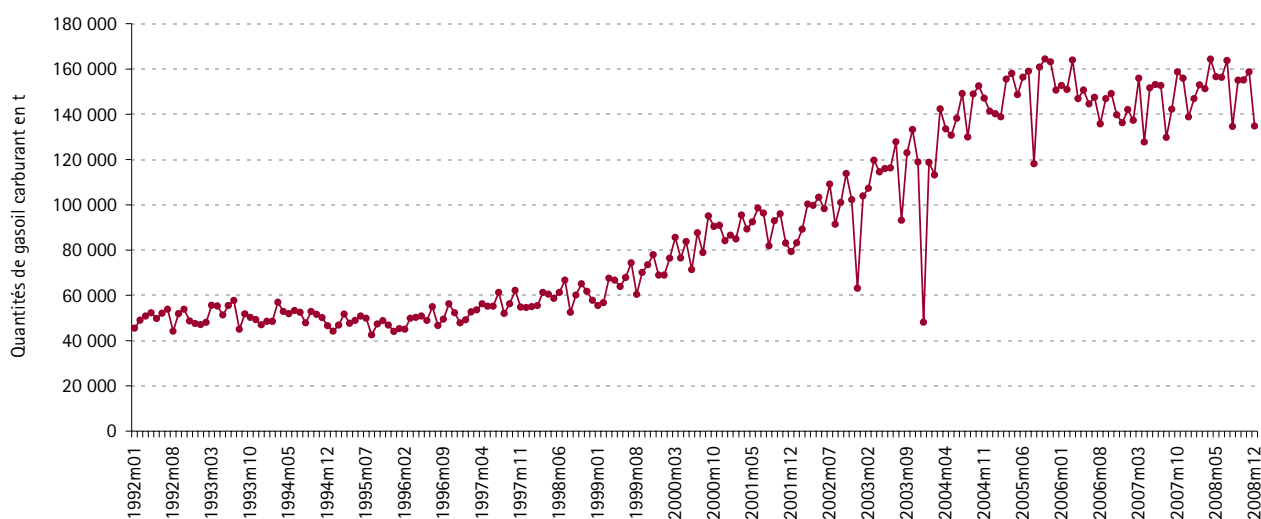
Cette approche présente le réel avantage d'être applicable *a priori*, c'est-à-dire à la fin de la phase

d'encodage. De plus, une fois automatisée cette vérification peut être lancée directement par l'encodeur. Ainsi, dans chacun des fichiers FED utilisé comme fichier de transmission régulière de données entre un partenaire et le STATEC, une macro permettant la vérification des données est implémentée. A la fin de l'encodage, ce module est lancé par l'encodeur et, en cas de valeurs suspectes, une « boîte de dialogue » s'ouvre pour demander la vérification de la valeur. L'encodeur peut soit modifier la donnée en cas d'erreur ou maintenir la valeur. Dans ce cas, il est alors invité à introduire un commentaire justificatif. Le commentaire est directement lié à la valeur et lors du chargement des données transmises dans la base de données de l'énergie du STATEC, ce commentaire est explicitement chargé avec la donnée. Il servira comme information utile pour les analyses ultérieures des séries.

Notons que l'approche par les quartiles présente également l'inconvénient de ne pas tenir compte des seuils (« technologique » ou « lié à de nouvelles installations ») de la série chronologique.

L'exemple des importations en « gasoil carburant » présente typiquement ce problème (cf. figure 6). Depuis 1992, on observe une augmentation croissante de la livraison intérieure de « gasoil carburant ». Si on applique l'approche par les quartiles sur toute la série, aucune valeur aberrante n'est détectée, par contre la simple observation du graphique des évolutions des livraisons ne laisse planer aucun doute sur l'existence d'une valeur aberrante aux mois de décembre 2002 et décembre 2003.

Figure 6: Evolution mensuelle des livraisons intérieures en gasoil carburant en tonnes



Seule une analyse complète de la tendance et de la saisonnalité de la série permet de lever cet inconvénient.

La méthode choisie pour automatiser cette analyse est l'utilisation de la méthode TRAMO dans une approche d'identification des erreurs.

C) La méthode TRAMO – L'approche statistique *a posteriori*

En synthèse, la méthode TRAMO vise à identifier le modèle ARIMA (tendance et saisonnalité) pour détecter et corriger les différents types de valeurs aberrantes. La méthode permet également de tenir compte des effets du calendrier et d'interpoler les données manquantes. L'objectif principal de l'application de la méthode TRAMO est l'application, par la suite de la méthode SEAT, qui permet le calcul prédictif sur base du modèle retenu.

Nous renvoyons le lecteur aux multiples articles de Gomez et Maravall¹ pour une description détaillée de la méthode TRAMO qu'il serait trop long d'exposer dans ce document. Nous mentionnons simplement que la méthode compare l'erreur entre la prédiction du modèle ARIMA retenu à la valeur historique observée et désigne celle-ci comme aberrante si elle dépasse un seuil défini.

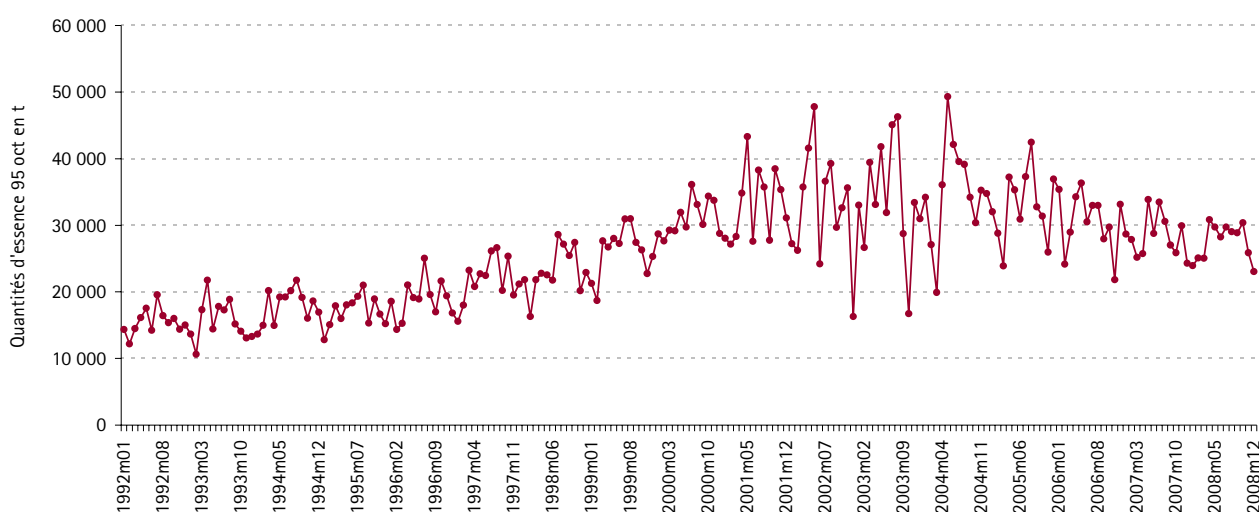
Parmi les séries chronologiques analysées, certaines présentent des valeurs aberrantes liées à une erreur

d'encodage.

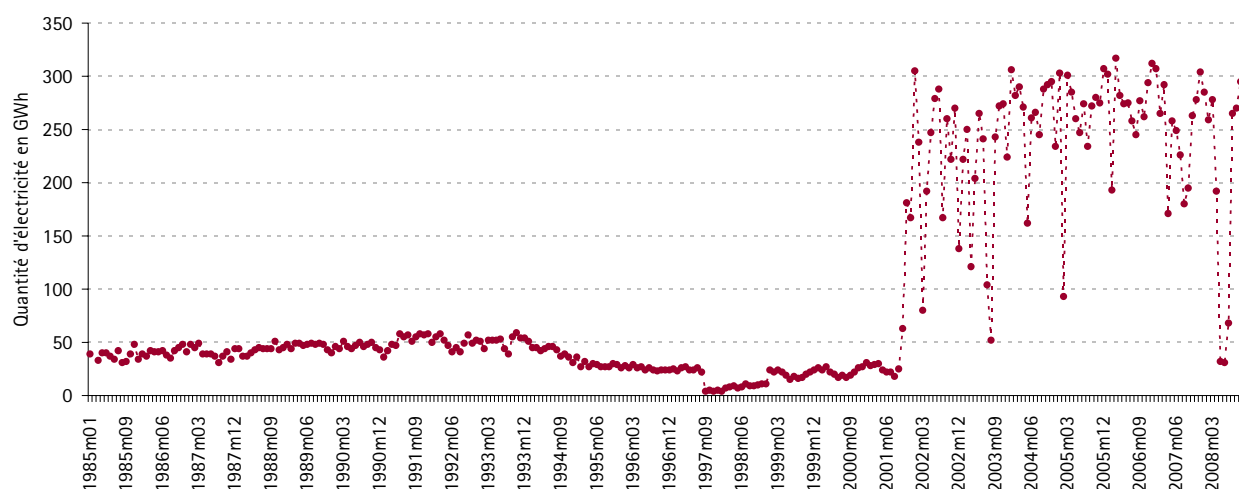
En présence de valeurs aberrantes « très extrêmes », la méthode TRAMO et l'approche par les quartiles détectent ces valeurs. Mais à plusieurs reprises, la méthode TRAMO a détecté plus de valeurs aberrantes que l'approche par les quartiles. Il s'est avéré que le fait de tenir compte de la tendance de la série permet d'identifier des valeurs suspectes qui sinon passeraient inaperçues dans la population. Pour exemple, la série de l'OCRA sur les importations d'essence 95oct (cf. figure 7) présente certainement trois valeurs suspectes (2002m12, 2003m10 et 2004m03). Par la méthode TRAMO, qui tient compte de l'évolution de la série, ces valeurs ont été détectées. Par contre, l'approche des quartiles ne les identifie pas car elles ne semblent pas hors gamme et notamment vis-à-vis des valeurs d'importation de 1992 à 1998. Sur ce point, la méthode TRAMO est réellement un avantage.

Un autre avantage de la méthode TRAMO est de distinguer trois types de valeurs : les valeurs aberrantes (AO), les valeurs seuil (LS) et les valeurs temporaires (TC). Pour expliciter cet avantage, examinons (cf. figure 8) la série de la production brute thermique (série Eurostat). Sur base d'une recherche de valeurs aberrantes via l'approche par les quartiles, aucune valeur n'est détectée. Quant à la méthode TRAMO, elle identifie 9 AO, 4 LS et 6 TC. Au vu du graphique de l'évolution de cette série, il est évident qu'il y a des valeurs remarquables nécessitant une explication.

Figure 7: Evolution mensuelle des importations d'essence 95oct en tonnes



¹ Gomez,V and Maravall, A. (1996), « Programs TRAMO (Time series Regression with Arima noise, Missing observations, and Outliers) and SEATS (Signal Extraction in Arima Time Series). Instructions for the User", Working Paper 9628, Servicio de Estudios, Banco de Espana.

Figure 8: Evolution mensuelle de la production brute thermique en GWh

Sur cette série, la méthode TRAMO détecte le seuil d'octobre-novembre 2001 mais également celui de septembre 1997. La méthode détecte également les valeurs aberrantes comme celles du mois d'août 2008 et de février 2005 bien visibles sur le graphique. En ce qui concerne les valeurs extrêmement basses de la fin 2008, la méthode les identifie comme des valeurs temporaires. C'est-à-dire qu'elle conclut non pas qu'une valeur est anormale mais plutôt qu'au cours d'une certaine période, les valeurs ont été exceptionnelles.

Les informations fournies par la méthode TRAMO sur les valeurs suspectes d'une série chronologique de ce type sont indiscutablement plus importantes et précieuses que celles de l'approche par les quartiles.

Cela dit, ce dernier exemple permet de souligner le fait que les méthodes statistiques permettent de détecter des valeurs suspectes mais que la conclusion finale, qualifiant la valeur d'atypique ou d'aberrante, revient au statisticien. Dans le cas de la production brute thermique, les données de la fin de l'année 2008 sont atypiques, mais correctes ! Elles sont dues à un problème technique de production d'électricité dans une centrale à gaz.

L'utilité d'une méthode statistique à la détection de valeurs suspectes est donc de susciter un questionnement sur certaines valeurs afin que l'opérateur puisse vérifier la véracité de celles-ci. Il est indéniable que dans cette optique la méthode TRAMO est un outil vraiment utile.

Malgré l'option de calcul automatique du modèle ARIMA, l'application de la méthode TRAMO reste complexe (mise en forme des données, choix des options, etc.). Dans notre analyse, nous avons systématiquement appliqué une même procédure de calcul automatique, mais dans certains cas, il semble qu'une procédure semi-automatisée (c'est-à-dire un choix des options adapté à chaque série analysée) mériterait d'être tentée. Deux éléments nous amènent à ce constat :

- Certaines valeurs identifiées comme aberrantes ne le sont pas et il est difficilement compréhensible de savoir pourquoi la méthode TRAMO les a sélectionnées. Nous supposons qu'il existe un problème dans la définition du modèle ARIMA.
- Pour plusieurs séries, la scission de la série en deux séries distinctes permettrait d'améliorer significativement la qualité du modèle (dans le cas, notamment, d'une série qui change radicalement de tendance et de saisonnalité).

La longueur de la série influence le choix du modèle ARIMA sélectionné par la procédure automatique. Elle va également influencer sur la détection des valeurs aberrantes : appliquer la méthode TRAMO sur une même série longue de 10 ans ou de 25 ans donnera des résultats légèrement différents dans la sélection des valeurs aberrantes sur la période commune.

Cette différence est liée à la procédure de calcul de la valeur du critère de détection (Cd) qui tient compte du nombre de valeurs dans la série. Par défaut, la méthode TRAMO se base sur la formule :

$$\begin{aligned} \text{Cd} &= 3 + 0,0025 \cdot (T - 50) \\ \text{et } \text{Cd} &= 3 \quad \text{si } T < 50 \end{aligned}$$

En conclusion, il apparaît que la méthode TRAMO permet de détecter les valeurs aberrantes, les changements de seuil et les changements temporaires sur des séries longues. Il est dès lors très utile dans l'analyse des données *a posteriori*. Il convient cependant de nuancer qu'il ne faut pas se fier totalement à des procédures automatiques, aussi sophistiquées soient-elles, mais que le jugement et l'esprit critique du statisticien sera toujours nécessaire.

Cela dit, au vu de sa mise en place relativement complexe, il semble difficile de pouvoir adapter cette méthode à la détection *a priori* lors de l'encodage des valeurs.

Une solution est tout de même envisageable si on utilise pleinement le potentiel de la méthode. En effet, la méthode permet le calcul de valeur prévisionnelle, il serait dès lors possible de comparer les valeurs prévisionnelles aux valeurs encodées au moment de l'encodage. Il est important de signaler qu'à la vue de l'analyse automatique réalisée ci-avant, cette solution ne serait pas applicable à tous les produits énergétiques (notamment aux séries ne possédant pas de tendance affirmée ou présentant une trop grande variabilité en seraient exclues) et qu'elle demande une étude plus approfondie dans les modèles de tendance et de saisonnalité sélectionnés automatiquement par la méthode.

Encadré 5: La méthodologie d'une analyse de la tendance et saisonnalité

L'approche scientifique généralement appliquée dans le domaine de la modélisation de séries chronologiques est composée de sept étapes, qui sont parfois itératives.

1. La visualisation des données ;
2. La préparation des données ;
3. L'identification du modèle ;
4. L'estimation des paramètres ;
5. La validation du modèle ;
6. La prévision ;
7. L'interprétation des résultats.

Les étapes 6 et 7 ne sont appliquées que si l'objectif du statisticien est d'établir des prévisions à partir des séries historiques.

Les étapes 1 à 5 sont, dans tous les cas, indispensables à la compréhension et l'analyse de séries chronologiques (détermination des tendances, saisonnalités, etc.).

1) La visualisation des données

La visualisation des données permet de se familiariser à la série chronologique et de s'informer sur le domaine d'application, la qualité des données (précision, exactitude), la périodicité de phénomène, l'homogénéité dans le temps, l'existence d'événements qui auraient pu influencer la série.

Elle consiste notamment à l'examen graphique des données visant à repérer les ruptures de séries, les valeurs aberrantes ou suspectes, les tendances et saisonnalités majeures.

2) La préparation des données

La préparation des données consiste à corriger et/ou modifier les données brutes pour obtenir des données épurées qui pourront être utilisées par les méthodes statistiques.

Elle vise notamment à corriger les valeurs aberrantes, à suppléer les données manquantes, à transformer les données (logarithmes, ...), à définir une série homogène (abandon des débuts de série).

3) L'identification du modèle

L'identification du modèle est un préalable à l'application de méthodes statistiques comme ARIMA ou d'autres. Ce modèle peut être déterminé par l'examen des graphiques d'autocorrélation, partielle ou non, de la série.

4) L'estimation des paramètres

L'estimation des paramètres consiste à définir les coefficients des polynômes du modèle. Elle est habituellement réalisée par la minimisation du critère MSE (carré moyen des erreurs), du critère MAE (erreur absolue moyenne) ou du critère MAPE (erreur absolue moyenne en pourcentage).

5) La validation du modèle

La validation du modèle consiste à s'assurer que le modèle est optimal. A défaut, les étapes 3 à 5 sont à réitérer. La validation correspond habituellement à l'application du modèle à une série connue et à l'examen des erreurs résiduelles.

6) La prévision

La prévision consiste à appliquer le modèle pour estimer les valeurs futures. Elle permet également de définir les intervalles de prévision.

7) L'interprétation des résultats

L'interprétation des résultats est l'aboutissement de toute la démarche.

6.4 La diffusion des statistiques

6.4.1 Les principes de publication et regroupement thématique

Une analyse poussée des objectifs de communication des statistiques de l'énergie au grand public et aux entreprises a été menée sur base des missions définies du STATEC et des expériences dans le domaine parmi les pays et organisations membres du Groupe d'Oslo. Le Groupe d'Oslo est un forum issu de la volonté des Nations Unies et doit permettre l'échange d'avis et d'expériences entre les pays membres sur les standards et les méthodes de compilation internationales des statistiques de l'énergie.

Sur base de cette analyse, les lignes directrices suivantes ont été retenues :

- La publication se fera de façon privilégiée à l'aide de tableaux en ligne (Portail statistique) proposant une interactivité à l'utilisateur.
- Afin de garantir une actualité optimale, la mise à jour se fera à rythme mensuel.
- Les statistiques les plus complètes possibles seront mises à disposition des utilisateurs, tout en respectant le secret statistique interdisant la diffusion de séries individualisables pour lesquels le STATEC n'a pas eu un accord explicite autorisant la publication ou qui n'ont pas été publiés autre part (par exemple dans un rapport annuel).
- Les tableaux diffusés seront très largement inspirés du bilan énergétique (aussi bien en termes de produits que de secteurs).

L'objectif premier de la diffusion des statistiques énergétiques étant de répondre aux grands enjeux de la société, il est proposé de structurer la thématique « énergie » comme suit :

- *La conjoncture énergétique* : comprenant des statistiques d'ensemble et des statistiques à court terme.
- *La sécurité/dépendance énergétique* : comprenant des statistiques sur la production (et les capacités de production), les importations énergétiques et la dépendance à certains produits énergétiques.
- *L'efficacité énergétique* : comprenant des statistiques sur les consommations par secteur.
- *Le développement des énergies renouvelables* : comprenant des statistiques spécifiques aux énergies renouvelables en comparaison des volumes totaux produits, importés ou consommés.
- *Le marché de l'énergie* : comprenant des statistiques sur les caractéristiques du marché, les prix des énergies et la part des dépenses énergétiques dans les dépenses totales des ménages.
- *L'impact sur l'environnement* : comprenant des statistiques sur les émissions de gaz à effet de serre (GES) issue de la combustion énergétique.¹

Pour chacune de ces sous-thématiques, le STATEC envisage de présenter d'une part, des tableaux avec les données de base et d'autre part, des indicateurs spécifiques. Chaque tableau ou indicateur sera accompagné d'une note méthodologique reprenant, entre autres, la source des données de base, les définitions des termes, abréviations, symboles et variables présentés et enfin la classification utilisée.²

¹ Au jour de la présente publication, la rubrique « Impact sur l'environnement » ne figure pas encore sur le Portail statistique du Luxembourg.

² Au jour de la présente publication, seuls les tableaux de données pour les cinq premières sous-thématiques sont publiés sur le Portail statistique.

6.4.2 Les supports de publication des tableaux

Pour chacune des sous-thématiques, une liste de tableaux possibles a été définie. Chaque tableau a été ensuite caractérisé sur base de différents critères :

- *Être robuste* : les variations des statistiques d'un tableau doivent effectivement refléter les variations de l'état.
- *Être comparable dans le temps et dans l'espace* : les statistiques, à périodicité mensuelle ou annuelle, doivent être représentatives du niveau national pour de possibles comparaisons internationales ultérieures.
- *Être un outil de communication* : facilement compréhensibles par les utilisateurs (spécialisés ou non) et idéalement facilement interprétables.
- *Être pragmatique* : les statistiques nécessaires à l'établissement d'un tableau doivent être facilement (et rapidement) disponibles à un coût raisonnable (ici dans le sens : nombre d'heures de travail pour l'établir).

Sur base de cette caractérisation, plusieurs tableaux diffusables ont été choisis et intégrés au programme de publication du STATEC :

- *Le portail statistique* présente tous les tableaux disponibles (mensuels et annuels), mis à jour ou historiques par sous-thématiques.
- *L'annuaire statistique* présentera les tableaux annuels (mis à jour) par sous-thématique.
- *Le Luxembourg en chiffres* présente deux tables synthétiques sur la consommation par vecteur énergétique et par usage.

Le STATEC envisage également de créer deux publications récurrentes spécifiques au domaine de l'énergie : l'une portera sur l'analyse du bilan énergétique annuel et l'autre sur la présentation des indicateurs énergétiques.

6.4.3 Le développement d'indicateurs

Suite au « *screening* » des pays et organisations membres du groupe d'Oslo, nous avons également sélectionné différents pays et institutions internationales pour l'exemplarité du travail accompli dans le développement d'indicateurs énergétiques.

A *contrario* de notre démarche dans le cadre de l'établissement d'un bilan énergétique et de la diffusion de tableaux statistiques qui visaient à synthétiser les points forts des exemples nationaux et internationaux pour ensuite les appliquer au contexte luxembourgeois, notre approche pour la thématique des indicateurs a été différente :

- La première étape consistait à définir nos besoins en indicateur (c'est-à-dire identifier les objectifs poursuivis dans l'étude de l'évolution d'un indicateur) ;
- La deuxième étape était de choisir, parmi les indicateurs déjà développés par les autres pays ou institutions internationales, les indicateurs qui correspondent à nos objectifs ;
- Enfin, la troisième étape devait s'assurer de la possibilité du calcul et de la diffusion d'un tel indicateur dans le contexte luxembourgeois.

Pour la première étape, les catégories selon les objectifs de connaissance ont été maintenues car elles sont tout à fait pertinentes et permettent de maintenir leur adéquation avec les données et statistiques existantes dans le domaine de l'énergie au Luxembourg. L'objectif affiché était de produire un maximum d'indicateurs afin de présenter une grande partie des statistiques actuellement produites au sein du STATEC.

Pour la sélection des indicateurs existants prévue à la deuxième étape, il est nécessaire d'évaluer ces indicateurs afin de confirmer qu'ils répondent à nos critères de sélection. Ces critères de sélection doivent tenir compte d'impératifs statistiques mais également d'adaptation au contexte luxembourgeois.

Par définition, un indicateur est une donnée quantitative qui permet de caractériser une situation évolutive, une action ou les conséquences d'une action, de façon à les évaluer et à les comparer à leur état à différentes dates.

Encadré 6: Le bilan énergétique restreint du Luxembourg en 2009

		Combustibles solides	Gaz naturel	Electricité	Chaleur	Produits pétroliers	Energies renouvelables et Energie produite à partir de déchets
2009	Total						
	TJ	TJ	TJ	TJ	TJ	TJ	TJ
Production d'énergie primaire	4 893			683	41		4 169
Importations brutes	187 636	2 801	46 577	21 681		116 554	23
Exportations	10 199			9 376		394	429
Variation des stocks	120					120	
Livraison intérieure brute	182 210	2 801	46 577	12 987	41	116 041	3 763
Entrée en transformation d'énergie	21 320		18 931			49	2 340
Sortie de transformation d'énergie (= production d'énergie secondaire)	11 876			10 663	1 213		
Usages propres (des centrales de transformation, y compris pompage et PAC) + pertes de distribution	1 645			1 645			
Livraison intérieure nette	171 121	2 801	27 646	22 005	1 254	115 992	1 423
Consommation énergétique finale	170 509	2 728	27 645	22 005	1 254	115 380	1 423
Consommation non énergétique finale						612	

Un indicateur doit donc présenter une série de caractéristiques, à savoir : être robuste, être comparable dans le temps et dans l'espace, être un outil de communication et être pragmatique.

La mise en application de ces indicateurs au contexte luxembourgeois, conformément à la troisième étape, implique la création d'une note méthodologique pour chaque indicateur.

Cette note méthodologique reprend les éléments suivants :

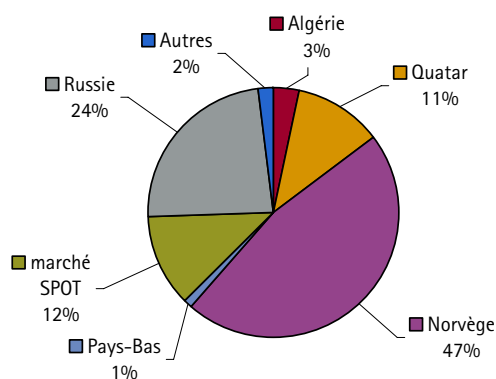
- Thématique couverte par l'indicateur ;
- Code de l'indicateur ;
- Nom de l'indicateur ;
- Périodicité ;

- Disponibilité ;
- Description de l'indicateur ;
- Unités ;
- Source des données de base pour le calcul de l'indicateur ;
- Liens avec d'autres tables et/ou indicateurs ;
- Commentaires sur les définitions et/ou classifications utilisées.

Cela dit, il serait prématuré de publier ces indicateurs sans réaliser une étude plus complète de l'implication de la publication de ces indicateurs. Cette étude devrait notamment comprendre l'établissement de ces indicateurs sur une large période (par exemple, pour les années 2000 à 2009), ainsi que la comparaison de ces indicateurs aux chiffres déjà publiés par d'autres institutions (nationales et internationales).

Encadré 7: Quelques indicateurs mis en valeur

Des indicateurs sous forme graphique, pour exemple :

Indicateur i2h: Répartition des quantités de gaz naturel importé par pays d'origine en 2007


Source: Enovos Luxembourg SA

Des indicateurs sous forme d'indice, pour exemple :

Indicateur i3d:

Consommation final (tous produits) des résidents / PIB (=intensité énergétique)

	2007	2008
Consommation finale** (en GJ)	120 936 872	120 747 349
PIB (en millions d'euro)	37 466	39 348
indice (GJ/ millions d'euro)	3 228	3 069

** la consommation des résidents exclut, par définition, la consommation de carburants par les non-résidents

7. Les conclusions et perspectives

La présente publication clôture deux années de travail sur les « statistiques de l'énergie au Luxembourg », qui ont permis au STATEC de se positionner comme l'interlocuteur central sur cette thématique et ce dans l'intérêt du secteur économique et décisionnel du Luxembourg.

En prenant à sa charge la centralisation, la vérification, la compilation et la transmission des statistiques énergétiques nationales, le STATEC a pris une place prépondérante dans l'établissement et le suivi des statistiques du domaine de l'énergie au Luxembourg. A travers ce nouveau rôle, il devient le garant de la qualité et la fiabilité des statistiques publiées.

Lors de la mise en place du schéma organisationnel nouvellement défini, le STATEC a pu montrer les intérêts :

- de la centralisation des données afin que tous les utilisateurs et décideurs s'appuient sur les mêmes statistiques ;

- de l'automatisation de la collecte et du traitement des données pour un gain d'efficacité et de qualité dans les missions des différentes administrations ;
- de la vérification des données permettant de déceler des incohérences et difficultés dans certains processus de collecte et de traitement ;
- du développement de méthodes d'estimation afin d'établir un bilan énergétique équilibré permettant une meilleure compréhension du domaine de l'énergie ;
- du suivi des discussions internationales afin d'être proactif dans l'établissement des futures statistiques de l'énergie et de sensibiliser les autres Etats membres et Eurostat aux priorités du Luxembourg ;
- de l'analyse des séries chronologiques pour expliciter l'évolution d'autres statistiques sortant du domaine de l'énergie.

Et pourtant, ce travail n'est pas un aboutissement mais une ouverture aux nombreuses possibilités de développements méthodologiques, d'analyses statistiques et de diffusion des données énergétiques.

Les méthodes mises en place pour l'établissement du bilan énergétique de l'année 2009, ont été appliquées à une série plus large, à savoir les années 2000 à 2008. Le principal écueil rencontré a été l'absence de certaines données spécifiques sur les premières années de cette série. Il a été dès lors nécessaire de mettre au point de nouvelles méthodes d'estimation pour le calcul de quelques statistiques spécifiques. Pour les années antérieures à 2000, le STATEC envisage d'établir des bilans simplifiés pour les années 1985, 1990 et 1995.

Une fois les statistiques historiques établies, une optique de projection à court et moyen termes pourra être envisagée. La modélisation énergétique est un sujet fort intéressant et de plus en plus demandé par les instances internationales dans les transmissions annuelles. Une modélisation de qualité nécessite la définition de scénarios plausibles ainsi que la mise au point de modèles fiables permettant de représenter le système dans son ensemble. Deux orientations sont donc nécessaires, d'une part l'analyse statistique des séries chronologiques pour comprendre le mode de fonctionnement du système (ainsi que les spécificités luxembourgeoises), et d'autre part, le couplage du système énergétique à d'autres systèmes comme le système macroéconomique.

Une autre possibilité de développement méthodologique est ainsi dessinée, à savoir l'identification des liens reliant les statistiques en volume et les statistiques en valeur. Un exercice pilote de conversion du bilan énergétique en tableaux ressources-emplois en volume a été testé. Cet exercice doit idéalement être reconduit et prolongé jusqu'au développement de tables hybrides, sortes de compilation des tableaux ressources-emplois en volume et en valeur. Les développements dans ce domaine sont porteurs de beaucoup d'espoirs et sont l'objet de toutes les attentions au niveau international. A court terme, ce type de tableaux seront inévitablement transmis à Eurostat : un nouveau règlement européen portant sur les comptes satellites de l'environnement devrait rentrer en vigueur au cours de l'année 2011. Ce règlement demandera, notamment, la transmission de données environnementales (émissions de polluants atmosphériques, consommations énergétiques, consommations d'eau, productions de déchets,...) dans un format équivalent à celui de la comptabilité nationale. Ces tables standardisées permettront

l'établissement des NAMEA (national accounting matrix with environmental accounts). La matrice NAMEA constitue le socle de base pour l'analyse structurée de la « croissance verte » d'un pays.

Enfin, il ne serait pas correct de passer sous silence l'important sujet de la diffusion des statistiques dans le domaine de l'énergie. La politique choisie par le STATEC est de diffuser un maximum d'informations statistiques sur son Portail Internet et de communiquer un jeu de tableaux plus réduit dans les publications sous format papier. Une série de tableaux ont été récemment publiés (novembre 2010). Ceux-ci conviennent parfaitement à un objectif de mise à disposition des statistiques à un maximum d'utilisateurs. Au sein de cette liste, quelques tableaux ont été sélectionnés pour la diffusion via « L'Annuaire statistique » et la publication « Le Luxembourg en chiffres ». Cette diffusion de statistiques correspond à une communication brute des chiffres et est souvent destinée à des habitués du domaine de l'énergie, mais d'autres indicateurs, comme les indicateurs composites, pourraient également faire l'objet d'une diffusion plus large pour le public néophyte. En plus d'une publication des indicateurs et tableaux, il serait utile de commenter les statistiques fournies dans ces outils de communication afin de fournir des clés d'analyse aux utilisateurs potentiels. La plupart des instituts nationaux de la statistique réalisent ce genre d'exercice par l'intermédiaire de notes explicatives publiées épisodiquement car ils ont compris l'importance de montrer l'intérêt des statistiques et du travail du statisticien au grand public. Le développement d'indicateurs composites et de notes explicatives est également une orientation à privilégier à la suite du présent travail sur les statistiques de l'énergie au Luxembourg.

Quoiqu'il en advienne, le Luxembourg, par la mise en place du système d'information statistique intégré de l'énergie décrit dans la présente publication, s'est d'ores et déjà doté d'un outil performant et efficace. Grâce aux efforts combinés des administrations, instituts, ministères et du STATEC, il est dorénavant possible de dessiner avec précision et qualité les implications des vecteurs énergétiques dans la société moderne luxembourgeoise. Il ne fait dès lors aucun doute que les retombées du système d'information statistique intégré conduiront à une prise à bras le corps des challenges économico-environnementaux de ce début de 21^{ème} siècle.

Annexe 1 : Liste non exhaustive des textes législatifs européens du domaine de l'énergie

Règlement n°1055/72 mis en application par le règlement n°3254/74 du Conseil, du 17 décembre 1974 et par le règlement n°2677/75 de la Commission, du 6 octobre 1975, portant sur la communication à la Commission des importations d'hydrocarbures aux produits pétroliers des sous-positions 27.10 A, B, C I et C II du tarif douanier commun (JO L275 du 27.10.1975, p. 1-7). [Lien vers ce texte](#)

Recommandation 91/141/CECA de la Commission, du 31 janvier 1991, concernant les statistiques du charbon (JO L 74 du 20.3.1991, p. 35-76) et modifié par la recommandation n°2393/96/CECA de la Commission, du 16 décembre 1996 (JO L 326 du 17.12.1996, p. 31-63). [Lien vers ce texte](#)

Règlement n°2964/95 du Conseil, du 20 décembre 1995, instaurant un enregistrement dans la Communauté des importations et des livraisons de pétrole brut (JO L 310 du 22.12.1995, p. 5-6). [Lien vers ce texte](#)

Règlement n°736/96 du Conseil, du 22 avril 1996, concernant la communication à la Commission des projets d'investissement d'intérêt communautaire dans les secteurs du pétrole, du gaz naturel et de l'électricité (JO L 102 du 25.4.1996, p.1-4) et mis en œuvre par le règlement n°2386/96 du Conseil, du 16 décembre 1996, portant application de ce règlement (JO L 326 du 17.12.1996, p. 13-20). [Lien vers ce texte](#)

Décision 1999/280/CE du Conseil, du 22 avril 1999, concernant une procédure communautaire d'information et de consultation sur les coûts d'approvisionnement en pétrole brut et les prix à la consommation des produits pétroliers (JO L 110 du 28.4.1999, p. 8-11) et mis en œuvre par la Décision 1999/566/CE de la Commission, du 26 juillet 1999 (JO L 216 du 14.8.1999, p. 8-12). [Lien vers ce texte](#)

Règlement n°405/2003 du Conseil, du 27 février 2003, relatif à la surveillance communautaire des importations de houille originaire de pays tiers (JO L 62 du 6.3.2003, p. 1-3). [Lien vers ce texte](#)

Directive 2003/30/CE du Parlement européen et du Conseil, du 8 mai 2003, visant à promouvoir l'utilisation de biocarburants ou autres carburants renouvelables dans les transports (JO L123 du 17.5.2003, p. 42-46). [Lien vers ce texte](#)

Directive 2003/54/CE du Parlement européen et du Conseil, du 26 juin 2003, concernant des règles communes pour le marché intérieur de l'électricité (JO L 296 du 14.11.2003, p.34-35) et abrogeant la directive 96/92/CE. [Lien vers ce texte](#)

Directive 2003/55/CE du Parlement européen et du Conseil, du 26 juin 2003, concernant des règles communes pour le marché intérieur du gaz naturel (JO L 176 du 15.7.2003, p.57-78) et abrogeant la directive 98/30/CE du 22 juin 1998. [Lien vers ce texte](#)

Directive n°2004/8/CE du Parlement européen et du Conseil, du 11 février 2004, concernant la promotion de la cogénération sur la base de la demande de chaleur utile dans le marché intérieur de l'énergie (JO L 52 du 21.2.2004, p. 50-60) et modifiant la directive 92/42/CEE. [Lien vers ce texte](#)

Directive 2004/67/CE du Conseil, du 26 avril 2004, relative aux mesures visant à garantir la sécurité d'approvisionnement en gaz naturel (JO L 127 du 29.4.2004, p. 92-96). [Lien vers ce texte](#)

Directive 2006/67/CE du Conseil du 24 juillet 2006 faisant obligation aux Etats membres de maintenir un niveau minimal de stocks de pétrole brute et/ou de produits pétroliers. (JO L 217 du 8.8.2006, p.8-15). [Lien vers ce texte](#)

Directive 2008/92/CE du Parlement européen et du Conseil du 22 octobre 2008 instaurant une procédure communautaire assurant la transparence des prix au consommateur final industriel de gaz et d'électricité (refonte). [Lien vers ce texte](#)

Règlement n°1099/2008/CE du Parlement européen et du Conseil du 22 octobre 2008 concernant les statistiques de l'énergie. [Lien vers ce texte](#)

Directive n°2009/28/CE du Parlement européen et du Conseil du 23 avril 2009 relative à la promotion de l'utilisation de l'énergie produite à partir de sources renouvelables et modifiant puis abrogeant les directives 2001/77/CE et 2003/30/CE. [Lien vers ce texte](#)

Directive n°2009/119/CE du Conseil du 14 septembre 2009 faisant obligation aux Etats membres de maintenir un niveau minimal de stocks de pétrole brut et/ou de produits pétroliers. [Lien vers ce texte](#)

Annexe 2 : Structure de la base de données de l'énergie

