

---

**Datawarehouse « Contrôles des transports routiers »**  
(Politique scientifique fédérale – Agora/01/082)  
**Rapport final et explication du projet de modèle**

---

Christophe Mues  
Promoteur: Jan Vanthienen

K.U.Leuven

Dept. Sciences économiques appliquées

Naamsestraat 69, 3000 Leuven

<http://www.econ.kuleuven.ac.be/tew/academic/infosys/members/mues/member.htm>

e-mail : [christophe.mues@econ.kuleuven.ac.be](mailto:christophe.mues@econ.kuleuven.ac.be)  
[jan.vanthienen@econ.kuleuven.ac.be](mailto:jan.vanthienen@econ.kuleuven.ac.be)

## Table des matières

1.	SITUATION ET INSTANCES PARTICIPANTES .....	3
2.	OBJECTIFS DU PROJET DE DATAWAREHOUSE.....	3
3.	DATAWAREHOUSING : ASPECTS MÉTHODOLOGIQUES.....	5
3.1.	Définition et caractéristiques d’un datawarehouse.....	5
3.2.	Conception d’un datawarehouse .....	7
3.3.	Le processus d’alimentation du datawarehouse.....	11
3.4.	Utilisation du datawarehouse .....	12
4.	DESCRIPTION DE LA SITUATION ACTUELLE .....	13
4.1.	Transport terrestre .....	13
4.2.	Douanes & Accises .....	15
4.3.	Office national de Sécurité sociale (O.N.S.S.).....	15
4.4.	Inspection des Lois sociales (I.L.S.) .....	17
4.5.	Inspection sociale (I.S.).....	18
4.6.	Police fédérale et Police locale .....	19
5.	PROJET DE MODELE ET COMMENTAIRE .....	19
6.	PROOF-OF-CONCEPT : TRAITEMENT DES DONNEES DOUANIERES .....	25
6.1.	Implémentation ETL.....	25
6.2.	Consultation des données cibles – environnement démonstratif .....	30
7.	CONCLUSION : SITUATION ACTUELLE ET RECOMMANDATIONS POUR LA SUITE.....	35
7.1.	Sur le plan législatif et juridique .....	35
7.2.	Sur le plan du contenu et sur le plan technique.....	36
7.3.	Sur le plan organisationnel et budgétaire .....	36
7.4.	Aperçu des étapes suivantes éventuelles .....	37
	BIBLIOGRAPHIE .....	38
	ABREVIATIONS ET ACRONYMES .....	38
	ANNEXES .....	38

### REMARQUE

L’appendice D (proof of concept Douane) et les illustrations suivantes sont en langue néerlandaise: figure 5, figure 9, figure 10, figure 11, figure 12, figure 12, figure 14 et figure 15. Les textes y correspondant donnent les explications nécessaires.

## 1. Situation et instances participantes

Le datawarehouse « Contrôles des transports routiers » est mis en place par la Direction Contrôle et Organismes de Transport de la Direction générale Transport terrestre du SPF Mobilité et Transports, conformément au plan d'action du 20 novembre 2001 relatif à la collaboration entre les différents services de contrôle en vue d'une coordination des contrôles dans le domaine des transports par route de personnes et de choses (tel que paru dans le Moniteur belge du 19 février 2002). Pour l'appui scientifique au projet, il est fait appel au département Sciences économiques appliquées de la K.U.Leuven, dans le cadre du projet Agora AG/01/082 du Service public fédéral de Programmation Politique scientifique. Les partenaires impliqués dans le plan d'action et le projet qui, à ce titre, fournissent et/ou utilisent les données, sont les services suivants :

- Direction générale Transport terrestre;
- Douanes et Accises;
- l'Inspection des Lois sociales (SPF Emploi, Travail et Concertation sociale);
- l'Inspection sociale (SPF Sécurité sociale);
- le service d'inspection de l'Office national de Sécurité sociale;
- la Police fédérale et la Police locale.

Ce plan d'action vise, par un meilleur contrôle du secteur des transports, à contribuer à :

- l'amélioration de la sécurité routière ;
- la promotion d'une concurrence équitable entre les transporteurs ;
- la lutte contre le travail illégal et la fraude sociale ;
- l'amélioration des conditions de travail des chauffeurs ;
- la protection de l'environnement,

et prévoit à cet effet la création d'un certain nombre de nouveaux organes (notamment neuf cellules provinciales « transport »), l'introduction d'actions coordonnées de contrôle (auxquelles les contrôleurs des services susmentionnés participent conjointement) et l'échange réciproque d'informations entre les services de contrôle et de police. Il propose en outre l'élaboration d'un système d'échange d'informations et la création d'une banque de données utile pour sa mise en oeuvre. C'est dans ce contexte que se situe le datawarehouse « Contrôles des transports routiers ».

## 2. Objectifs du projet de datawarehouse

Compte tenu des objectifs généraux du plan d'action précité, du fonctionnement actuel et des besoins d'informations des différents services de contrôle du transport, tels que nous les avons constatés, nous estimons que ce projet de datawarehouse doit poursuivre les objectifs concrets suivants.

**Rapport européen.** Tout d'abord, le datawarehouse doit permettre, d'une manière relativement simple, à la Direction générale Transport terrestre de transmettre les informations les plus pertinentes en matière de constatation d'infractions (graves) en transport à d'autres instances à la demande de celles-ci, conformément à des dispositions réglementaires ou des accords, ou encore en appui à la politique. Ce rapport se déroule en particulier avec d'autres

services de contrôle européens des transports dans le cadre de l'arrangement administratif « Euro Contrôle Route ». L'apport des données nécessaires à cet effet par les autres partenaires et leur introduction dans le datawarehouse permettront dès lors de transmettre plus systématiquement aux instances européennes compétentes les informations relatives aux infractions constatées en Belgique et commises par les transporteurs routiers en provenance d'autres pays participants.

**Profil de l'entreprise.** Deuxièmement, le datawarehouse doit permettre d'esquisser un profil plus précis du comportement d'une entreprise de transport individuelle. Alors qu'auparavant les différents partenaires ne pouvaient se baser que sur les données de contrôle qu'ils avaient eux-mêmes recueillies, le datawarehouse leur donnera désormais accès aux résultats fournis par les autres partenaires ou d'autres pays « Euro Contrôle Route » et permettra de présenter ces informations sous forme agrégée au niveau d'une entreprise de transport individuelle. En cela, le datawarehouse est un instrument approprié pour appuyer la politique de sanctions sur le plan des réglementations contrôlées : il fournira en effet des informations ciblées sur la base desquelles la Direction générale Transport terrestre pourra décider du retrait de la licence d'une entreprise de transport déterminée. En outre, de telles informations peuvent constituer une indication de l'opportunité d'une enquête complémentaire. Si, par exemple, lors de contrôles routiers effectués par la Direction générale Transport terrestre ou les Douanes et Accises, des infractions sont fréquemment constatées pour des véhicules d'une firme donnée, l'ILS ou l'IS pourront réaliser une enquête approfondie sur l'entreprise quant au respect de la réglementation du travail et de la réglementation sociale.

**Coordination réciproque.** Troisièmement, le datawarehouse joue également un rôle dans l'amélioration de la coordination entre les divers services de contrôle, en ce qu'il permet aux contrôleurs de prendre simplement connaissance des résultats des contrôles effectués par d'autres services. Il est à noter que la périodicité de l'apport des données implique naturellement certaines restrictions quant à la disponibilité des contrôles récents.

**Statistiques et examen du secteur.** Bien que la tenue de statistiques ou la réalisation de toutes sortes d'analyses basées sur la vérification ou la découverte (cf. infra : section 3.4) ne constituent pas des objectifs initiaux d'utilisation pour la création du datawarehouse, celui-ci peut néanmoins servir de base appropriée à une suite de projet dans ce domaine. En effet, la conception du projet de banque de données (cf. infra) est telle qu'elle doit permettre d'effectuer des analyses efficaces. Nous pensons ici par exemple au calcul des statistiques relatives au nombre de contrôles, d'infractions et à leur interrelation, le tout classé selon un certain nombre de critères d'agrégation (tels que le transport de personnes versus celui des marchandises, le type d'infraction, l'origine de l'entreprise, etc.), ou à l'identification d'indicateurs potentiels de comportement frauduleux. Le fait que de telles analyses puissent de nouveau porter sur les données recueillies par les différents services de contrôle, doit permettre d'obtenir une image adéquate du fonctionnement du secteur des transports dans sa globalité et pourrait ainsi fournir à l'avenir des informations précieuses en vue de la définition la politique à suivre.

**Accès aux données externes.** Indépendamment des données contenues dans les différentes sources opérationnelles du datawarehouse, il existe d'autres banques de données relatives aux différents acteurs du secteur des transports (employeurs et travailleurs, données concernant les véhicules, etc.), dont l'accessibilité est actuellement limitée, mais qui font l'objet d'un exa-

men dans le cadre de ce projet pour déterminer dans quelle mesure elles peuvent être ouvertes à tous les partenaires. Nous pensons ici en premier lieu :

- à la banque de données des licences Transis;
- aux données relatives aux plaques d'immatriculation de la Direction pour l'Immatriculation des véhicules (D.I.V.);
- à la banque de données DIMONA (déclaration immédiate de l'emploi).

Elles sont susceptibles d'enrichir le datawarehouse, parce qu'elles permettraient entre autres de valider les données de contrôle fournies.

### 3. Datawarehousing : aspects méthodologiques

Dans cette section, nous expliquons brièvement ce que nous entendons par datawarehouse et quels sont les principes et méthodes d'élaboration proposés par la littérature de la recherche applicables dans ce domaine.

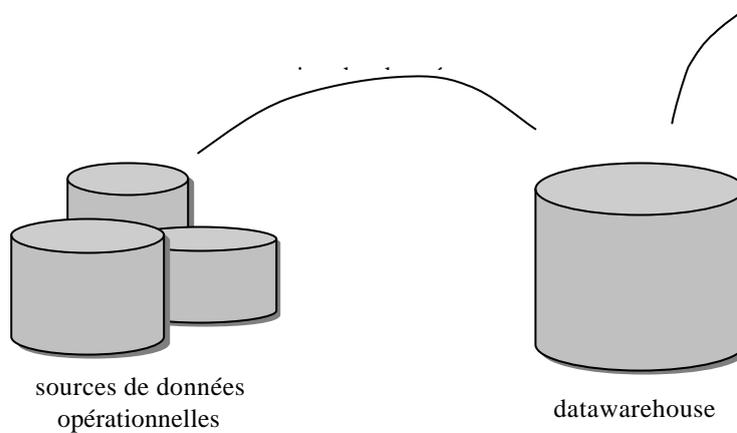
#### 3.1. Définition et caractéristiques d'un datawarehouse

En nous basant sur la définition de Inmon (1996)<sup>1</sup>, nous considérons un datawarehouse comme une collection de données intégrées, orientées sujet, historiées et non variables, susceptibles d'appuyer le processus décisionnel dans une entreprise ou organisation.

**Données intégrées.** Ce qui caractérise peut-être le plus un datawarehouse est le fait qu'il collecte et intègre de la manière la plus automatisée possible des données provenant d'un ensemble hétérogène de systèmes sources. Comme ceux-ci possèdent généralement leur propre méthode de conception et évoluent indépendamment les uns des autres, il n'y a en règle générale pas la moindre cohérence dans l'utilisation des modes de codification, la dénomination et la mesure des attributs, la structure-clé, etc. En vue de permettre des consultations ou des analyses globales néanmoins fiables au travers de l'ensemble des systèmes, il est nécessaire de traiter la structure et le contenu des différents apports de données de telle sorte qu'ils soient rendus cohérents avec ceux du datawarehouse global.

---

<sup>1</sup> Inmon, W. H. (1996), Building the Data Warehouse, second edition, Wiley, N.Y., 401 pp.



*Figure 1: intégration des données dans le datawarehouse*

**Données orientées sujet.** Une deuxième caractéristique d'un datawarehouse est d'être spécifiquement structuré autour de quelques sujets bien déterminés issus de la réalité organisationnelle, à propos desquels on souhaite récolter des informations et des connaissances (ex. : clients, produits, ventes, ..., ou, dans notre contexte, entreprises de transport, contrôles effectués, etc.), au lieu de constituer une image de la structure interne de l'organisation et de ses différents domaines d'application.

**Données historiques.** Alors qu'une banque de données opérationnelle est d'abord conçue pour obtenir un instantané de la situation actuelle et qu'elle doit dès lors être tenue à jour en permanence – et de préférence en temps réel – (ex. : quelles sont les enquêtes en cours ? Pour quelle entreprise travaille un conducteur de poids lourd déterminé ?), l'objectif primaire d'un datawarehouse est de tenir un historique des faits (ex. : quelles sont les entreprises qui ont été contrôlées l'année passée ?). Bien que cette distinction ne soit pas toujours aussi stricte (il arrive souvent qu'une banque de données conserve des données historiques, mais pour une période plus limitée), ces points de départ différents influencent naturellement la conception de la banque de données. Ainsi, la plupart des tables du datawarehouse comporteront un élément temporel. De même, le fait de disposer en permanence des informations les plus récentes n'est pas aussi vital et fait que l'apport des données est moins fréquent.

**Données non variables.** Une quatrième caractéristique connexe d'un datawarehouse consiste à ne plus modifier en principe les données qui y sont stockées; seules de nouvelles données y sont ajoutées. Supposons, par exemple, qu'il ressort d'un nouveau contrôle que la nationalité d'un conducteur de poids lourd a changé et que nous actualisons directement ces données dans un enregistrement précédemment mémorisé pour cette personne. Dans ce cas, tous les contrôles antérieurs ayant trait à cette même personne se réfèreraient aussi indirectement à cette donnée modifiée et entraîneraient la perte évidente d'informations historiques (et l'altération d'analyses statistiques éventuelles). Plutôt que de modifier un tel enregistrement, la préférence est souvent donnée, dans le cadre d'un datawarehouse, à l'incorporation d'un élément temporel dans la clé d'identification (par exemple valable du ... au ...). Ainsi, dans cet exemple, la table des conducteurs pourrait contenir deux enregistrements pour la même personne, chacun ayant sa propre période de validité.

**Aide à la décision.** Enfin, un datawarehouse a un objectif et un groupe d'utilisateurs spécifiques : il doit constituer la base d'un système d'aide à la décision. C'est la raison pour laquelle il n'est pas nécessaire d'y stocker toutes les données. Seules celles susceptibles de contribuer à l'amélioration des décisions prises suffisent. Il s'agit la plupart du temps de décisions d'ordre tactique et/ou stratégique. Dans notre cas, cela concerne (certainement au cours d'une première phase ne générant pas encore de statistiques sectorielles) tout autant des aspects opérationnels tels que l'évaluation du comportement d'une entreprise de transport individuelle en support à une politique de sanctions.

Sur la base de ce qui précède, nous estimons que la fonctionnalité et les caractéristiques de la banque de données à développer pour les contrôles des transports répondent suffisamment au profil d'un datawarehouse pour nous baser également, lors de sa conception, sur la littérature de recherche en matière de datawarehouse. Dans la sous-section suivante, nous abordons brièvement quelques aspects pertinents de la conception.

### **3.2. Conception d'un datawarehouse**

Une conséquence intéressante du fait qu'un datawarehouse ne doit en principe faire l'objet d'aucune modification, est que la duplication et la dénormalisation des données ne provoquent pas les mêmes problèmes d'incohérence et de performance que dans le cas des banques de données opérationnelles à partir desquelles elles sont alimentées. En outre, les nouvelles données sont spécifiquement chargées en batch, ce qui génère un tout autre modèle d'utilisation que celui des nombreuses petites transactions caractéristiques d'un système opérationnel. Par ailleurs, lors de l'accès en lecture, l'accent n'est généralement pas mis sur la recherche fréquente de petites quantités de données détaillées, mais sur des traitements moins nombreux mais de plus grande ampleur (ex. : l'agrégation de données chiffrées trimestrielles ou régionales, etc.). En résumé, la performance de mise à jour des systèmes de datawarehouse est moins critique; ceux-ci doivent néanmoins pouvoir traiter de grandes quantités de données ainsi que des consultations complexes. Il n'est dès lors pas étonnant que la conception d'un datawarehouse soit souvent différente de celle d'une banque de données opérationnelle.

**Schéma en étoile.** Une méthode courante de conception de datawarehouse est basée sur la définition d'un schéma dit *en étoile*. Un tel schéma n'est rien d'autre qu'un modèle relationnel dans lequel une *table de faits* centrale est reliée à une ou plusieurs *tables de dimension*. La Figure ci-après donne un exemple fictif de schéma en étoile pour les ventes d'une chaîne de magasins. La Figure 2 montre les lignes complétées pour un exemple de vente (d'un casier de bouteilles d'eau pétillante de la marque Spa à monsieur Lens de Heverlee, le 22 juillet 2003, dans la supérette de Louvain).

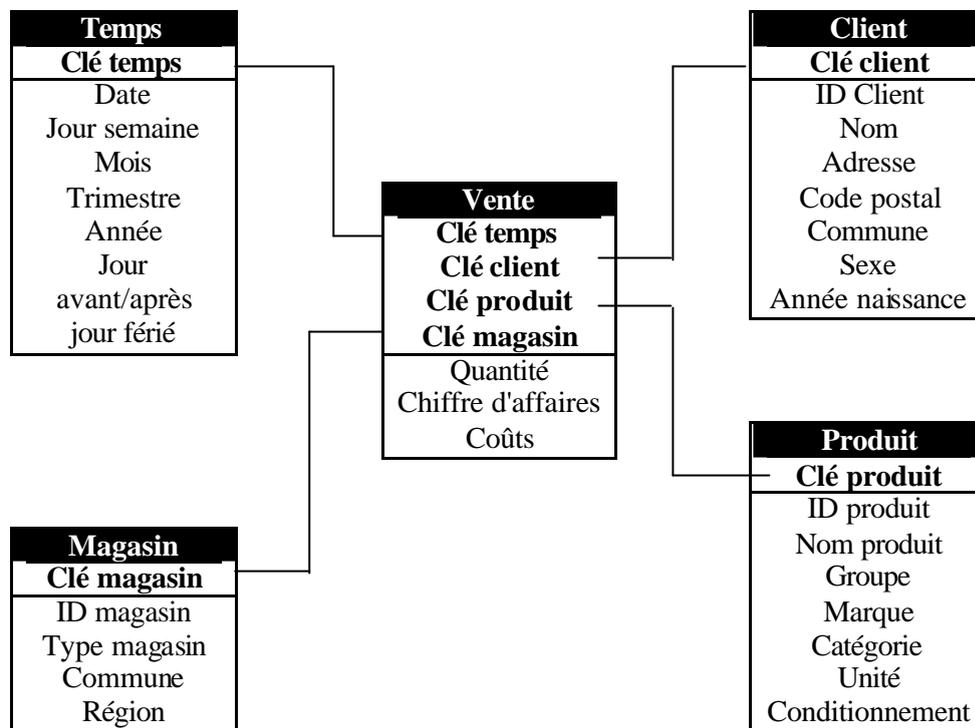


Figure2: exemple de schéma en étoile pour les ventes d'une chaîne de magasins<sup>2</sup>

La table de faits constitue une table de référence centrale permettant d'accéder aux événements ou activités archivés et inhérents à un processus déterminé. Dans l'exemple ci-dessus, il s'agit d'une chaîne de magasin, qui, à partir d'un processus de vente, enregistre des faits relatifs à un achat : qui, quoi, où et quand (cf. la table « ventes »). Une table de faits contient essentiellement des informations numériques : une clé composée permettant de se référer aux lignes des tables de dimension (cf. les tables « temps », « client », « produit » et « magasin ») et un certain nombre de valeurs mesurées susceptibles d'être agrégées et pouvant être attribuées à un fait déterminé (en l'occurrence : quantité, chiffre d'affaires et coûts). Ici, une décision conceptuelle importante porte sur la définition du maillage fin de la table de faits : à quel niveau de détail ou d'agrégation définissons-nous un fait atomique ? Tant que la capacité nécessaire de stockage ne constitue pas un problème insurmontable, il est peut-être préférable de garder un maillage aussi dense que possible. En effet, l'agrégation peut toujours être appliquée ultérieurement (un certain nombre de tables d'agrégation, dans lesquelles une ou plusieurs dimensions disparaissent complètement ou partiellement, est souvent dérivé de la table de faits).

Une table de dimension se compose spécifiquement d'une clé non significative établissant un lien avec les lignes de la table de faits (ex. : « clé produit »), une clé significative reprise d'une source de données opérationnelle ou externe (ex. : « ID produit ») et d'un nombre d'attributs permettant de caractériser la dimension (ex. : nom du produit, groupe, marque, ...).

<sup>2</sup> Vandenbulcke, J.A. et Lemahieu, W., Databasesystemen voor de praktijk, septième édition, Ten Hagen et Stam, Den Haag, 560 pp.

Ces derniers fournissent les critères de sélection sur la base desquels les valeurs mesurées figurant dans la table de faits peuvent être agrégées. Ainsi, dans l'exemple donné, nous pourrions demander le rapport entre les quantités vendues et le chiffre d'affaires par groupe de produits et (au sein de chaque groupe) par région de magasins.

<b>Ventes</b>						
<i>Clé temps</i>	<i>Clé client</i>	<i>Clé produit</i>	<i>Clé magasin</i>	<i>Quantité</i>	<i>Chiffre d'affaires</i>	<i>Coûts</i>
<b>203</b>	<b>55</b>	<b>68</b>	<b>07</b>	1	7,00	6,25

<b>Temps</b>							
<i>Clé temps</i>	<i>Date</i>	<i>Jour semaine</i>	<i>Semaine</i>	<i>Mois</i>	<i>Trimestre</i>	<i>Année</i>	<i>Jour avant/après jour férié</i>
<b>203</b>	22/07/03	2	30	Juillet	3	2003	Oui

<b>Client</b>							
<i>Clé client</i>	<i>ID client</i>	<i>Nom</i>	<i>Adresse</i>	<i>Code postal</i>	<i>Commune</i>	<i>Sexe</i>	<i>Année naissance</i>
<b>55</b>	K100567	Lens	Burgstraat 45	B3001	Heverlee	M	1975

<b>Produit</b>							
<i>Clé produit</i>	<i>ID produit</i>	<i>Nom produit</i>	<i>Groupe</i>	<i>Marque</i>	<i>Catégorie</i>	<i>Unité</i>	<i>Conditionnement</i>
<b>68</b>	P658942	Eau	Pétillante	Spa	Produit alimentaire	6	Verre

<b>Magasin</b>				
<i>Clé magasin</i>	<i>ID magasin</i>	<i>Type magasin</i>	<i>Commune</i>	<i>Région</i>
<b>07</b>	W894	Supérette	Louvain	Brabant flamand

Figure 2 : ventes d'une chaîne de magasins; enregistrement d'un exemple de transaction de vente

A propos de ce qui précède, il convient de remarquer que les tables de dimension sont souvent dénormalisées.<sup>3</sup> Si les structures hiérarchiques sont néanmoins explicitées directement au sein des dimensions, il s'agit alors d'un « schéma en flocon de neige ». Ainsi, la dépendance fonctionnelle (transitive) entre les attributs « Commune » et « Région » pourrait être reprise dans une table distincte, à laquelle il est fait référence à partir de la table magasin. Une telle stratégie permet certes des économies d'espace mais influencera souvent négativement la performance de la consultation.

<sup>3</sup> Comme signalé précédemment : étant donné que la banque de données ne fait généralement pas l'objet d'instructions de modification, aucune anomalie de mise à jour ne peut dès lors survenir, alors que la performance peut à ce niveau être améliorée (en évitant de coûteuses opérations « join »).

**Traitement de l'historique.** L'exemple de schéma décrit ci-dessus permet d'intégrer directement l'historique des ventes et facilite une série d'analyses où la dimension temps joue un rôle. Au fil du temps, la situation reflétée par les attributs dans les tables de dimension peut cependant changer. Ainsi, un client peut à un moment donné changer d'adresse. Si nous voulons que de telles modifications soient maniées correctement dans le cadre des analyses, il y a lieu d'en tenir compte dans le modèle. Une première méthode de travail consiste à écraser l'enregistrement concerné dans la table de dimension. Ceci entraînerait toutefois une perte d'informations historiques : si nous voulons examiner le lien de dépendance entre les ventes et le domicile du client, cette solution pourrait alors aboutir à des conclusions erronées. En effet, les anciennes transactions d'achat concernant le client récemment déménagé pourraient être involontairement liées à sa nouvelle adresse. Une meilleure solution consiste à « accumuler » les modifications. Dans l'exemple susmentionné, cela revient à ajouter un deuxième enregistrement pour le client contenant sa nouvelle adresse, pour lequel nous créons en outre une nouvelle clé client (voir Figure 3). Des champs additionnels peuvent également être prévus pour les dates de début et de fin de la période de validité. La clé client (non significative) n'identifie dès lors plus un client déterminé mais bien un client déterminé au cours d'une période déterminée; la même ID client (significative) apparaît donc plusieurs fois.

Client									
Clé client	ID client	Date début	Date fin	Nom	Adresse	Code postal	Commune	...	
55	K100567	01/01/1998	21/07/2003	Lens	Burgstraat 45	B3001	Heverlee	...	
78	K100567	22/07/2003	31/12/2999	Lens	Naamsestraat 85	B3000	Louvain	...	

Figure 3 : ventes d'une chaîne de magasins; maniement des changements dans la dimension client

Une autre solution consiste à distinguer des *mini-dimensions*. Celles-ci apparaissent lorsque, dans une vaste table de dimension, un sous-ensemble d'attributs intercorrélés est détaché de la dimension initiale. Supposons que, dans l'exemple donné, nous définissions une table distincte reprenant des données démographiques relatives aux clients (sexe, catégories d'âge et de revenu, état civil, etc.), pourvues à chaque fois d'une clé propre. Il est à chaque fois fait référence à cette dernière à partir des enregistrements de la table client (voir Figure 4). La démographie fait ainsi office de mini-dimension du client. Si les données démographiques d'un client changent, un nouvel enregistrement est inséré dans la table démographie et la référence dans la table client est mise en concordance avec les données les plus récentes. Toutefois, en mentionnant la référence-clé démographie non seulement dans la table client mais aussi dans la table ventes centrale (comme dans la Figure 4), nous pouvons faire en sorte que les faits de vente du passé restent liés aux anciennes données démographiques adéquates.

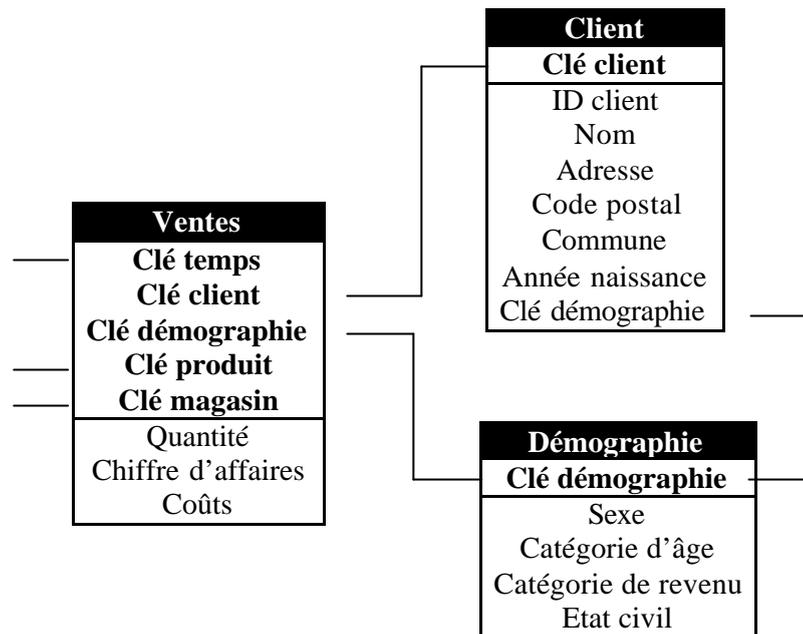


Figure 4 : ventes d'une chaîne de magasins; mini-dimension avec historique

### 3.3. Le processus d'alimentation du datawarehouse

L'alimentation d'un datawarehouse consiste à y introduire des données correctes et pertinentes selon des procédures convenues et à les conserver. Les quatre étapes intermédiaires suivantes peuvent y être distinguées :

- l'extraction de données;
- la transformation de données;
- le (re)chargement du datawarehouse;
- le post-traitement.

**Extraction de données.** On entend par extraction de données le prélèvement des données nécessaires d'une source opérationnelle (au moyen d'instructions SELECT ou d'utilitaires UNLOAD). Le processus d'extraction est souvent asynchrone périodique. Un fichier de données est créé à des moments déterminés (ex. : quotidiennement ou mensuellement) et contient uniquement les modifications incrémentales dans les données sources (c'est-à-dire depuis l'extraction précédente). Si le système source prévoit une indication de temps (*time stamping*), l'extraction pourra s'en servir utilement. Si non, d'autres méthodes plus complexes seront nécessaires, telles que le scannage de fichiers log ou la comparaison d'images « before » et « after ».

**Transformation de données.** Le processus de transformation de données qui suit, comprend le traitement des données extraites en vue de la préparation du chargement. Ceci sous-entend les activités suivantes :

- le formatage de données : transformation dans le format exigé (ex. : comma-delimited text-files versus formats import/export spécifiques aux produits);
- le nettoyage de données : élimination ou correction de données polluées (entre autres rendre les codes appliqués cohérents; le sexe pourrait être mentionné sous différentes formes : 'H'/'F', 'homme'/'femme' ou au moyen d'un code chiffré);

- l'agrégation et la jonction de données : si le maillage des faits dans le datawarehouse est moindre que celui dans le système source (ex. : les ventes du magasin pourraient être cumulées sur une base journalière plutôt que transactionnelle);
- l'enrichissement de données : complémentation des données propres par des données externes.

Nous observons que ces activités peuvent se dérouler (partiellement) avant ou après le transport physique des données, donc sur la plate-forme source ou cible.

**(Re)chargement du datawarehouse.** Les fichiers de données obtenus au cours des étapes précédentes doivent ensuite être chargés (en vrac) dans les tables de faits et de dimension du datawarehouse. Les clés non significatives sont générées durant le processus de chargement.

**Post-traitement.** Le post-traitement a lieu après le chargement. Pendant ce processus, des données dérivées sont générées, des tables d'agrégats sont complétées (ex. : tables reprenant les données trimestrielles, directement consultables), etc.

Il va de soi que l'organisation du processus global ainsi que l'implémentation et la documentation des procédures requises est un travail de longue haleine qui ne peut certainement pas être sous-estimé. Le développement se déroule dès lors souvent selon un processus itératif et débute avec seulement quelques sujets importants; l'extension du contenu du datawarehouse se poursuit progressivement.

### **3.4. Utilisation du datawarehouse**

Finalement, un datawarehouse est destiné à être consulté et, donc, à aider l'utilisateur dans l'exécution de ses tâches. Dans nombre de cas, le datawarehouse constitue la source principale de données pour un système d'aide à la décision. A ce niveau, une distinction est souvent opérée entre les méthodes d'analyse *basées sur la vérification*, qui permettent à l'utilisateur de tester une hypothèse à partir de données issues du datawarehouse, et celles *basées sur la découverte*, qui visent à mettre en lumière de nouveaux modèles de connaissances.

**Méthodes d'analyse basées sur la vérification.** Dans le premier cas, l'utilisateur doit avoir la possibilité d'analyser les données à souhait via les dimensions disponibles (par exemple en permettant d'agréger les données de ventes aux catégories de produit, aux emplacements des magasins, etc.). Des outils visuels modernes de consultation et de rapport appuient une telle approche. Divers outils OLAP (On-Line Analytical Processing) offrent en outre de puissantes formes de traitement.

**Méthodes d'analyse basées sur la découverte.** Dans le deuxième cas, l'objectif consiste à découvrir des nouvelles connaissances à l'aide de techniques dites de *data mining*, comme par exemple les relations insoupçonnées entre les données démographiques et le comportement d'achat, en fonction desquelles des actions spécifiques pourront être entreprises par la suite. Les résultats obtenus dépendront naturellement beaucoup de la qualité et de la complétude des données contenues dans le datawarehouse.

## 4. Description de la situation actuelle

Chacun des partenaires précités effectue certains contrôles et inspections dans le domaine du transport de personnes et de marchandises par route. La nature et le volume des données rassemblées qui en résultent dépendent évidemment des compétences<sup>4</sup> (en partie chevauchantes, en partie communes) et des moyens du service spécifique de contrôle. Un autre point de divergence est la technologie au moyen de laquelle et la forme définitive dans laquelle ces données sont conservées et rendues accessibles. Nous donnons ci-après un aperçu de la situation actuelle.<sup>5</sup>

Nous ferons une distinction concernant la nature du contrôle, et plus particulièrement entre un *contrôle sur la route* et un *contrôle dans une entreprise*. Le contrôle sur la route :

- concerne un véhicule qui est arrêté sur la voie publique par un contrôleur ou une équipe de contrôleurs;
- est limité en durée;
- peut être limité à un seul véhicule (articulé ou non) et à un conducteur spécifique.

Le second type de contrôle :

- concerne une entreprise (et a lieu la plupart du temps dans l'entreprise);
- demande un traitement (administratif) qui prend souvent du temps;
- concerne indirectement plusieurs véhicules et/ou travailleurs.

### 4.1. Transport terrestre

**Missions de contrôle effectuées.** Ce service de contrôle, qui fait partie du SPF Mobilité & Transports, effectue des contrôles en matière de transport routier de marchandises – y compris les marchandises dangereuses (cf. ADR) – et de personnes, tant au niveau du transport professionnel que du transport pour compte propre. Il veille au respect des dispositions de diverses conventions internationales, de règlements CE (notamment les temps de conduite et de repos, l'utilisation du tachygraphe, cf. CE 3820/85 et 3821/85) et de la législation nationale, au moyen de contrôles sur la route et dans les entreprises qui peuvent donner lieu au retrait des autorisations de transport. Dans cette capacité, il utilise une banque de données des entreprises belges autorisées, appelée *Transis* (consultable depuis peu sur le site internet du SPF : <http://mobilit.fgov.be>). En outre, il est l'instance responsable pour l'échange de données dans le cadre de l'arrangement administratif 'Euro Contrôle Route' (cf. supra: section 2). En outre, il recense sous forme de statistiques mensuelles, trimestrielles et annuelles, les résultats des contrôles qu'il a effectués. Il agrège le nombre de contrôles, de disques de tachygraphe contrôlés, d'infractions constatées, de procès-verbaux et de perceptions immédiates selon une série de critères dont l'origine du conducteur (Belge vs résident UE vs résident non appartenant à l'UE), le type de transport (marchandises vs personnes, professionnel vs compte pro-

---

<sup>4</sup> En ce qui concerne les différents paquets de compétences, nous renvoyons au vade-mecum qui a été établi dans le cadre du Plan d'Action du 20 novembre 2001.

<sup>5</sup> En ce qui concerne les partenaires de la Police fédérale et locale, il faut d'abord un examen du cadre légal de l'échange éventuel de données (autres que celles anonymes); c'est pourquoi les systèmes d'alimentation (qui sont pour une part encore en cours de développement) ne sont pas encore repris dans cet aperçu.

pre), la nature de l'infraction, etc. Enfin, il prend les initiatives dans le cadre du Plan d'Action et le projet Agora qui en résulte en vue de la création d'une banque de données pour le transport routier.

**Sources de données opérationnelles.** Si le contrôleur constate une infraction, il établit un formulaire avec les détails suivants:

- lieu, date, heure, contrôleur;
- *travailleur(s)/conducteur*: nom, nationalité, date et lieu de naissance, adresse, numéro de carte d'identité, numéro de permis de conduire;
- *employeur/transporteur*: nom et adresse;
- *véhicule* : (e.a.) marque, châssis, numéro de plaque, numéro du certificat de visite (du tracteur et du semi-remorque), lieu de départ et de destination, propriétaire;
- *constatations*: infractions à quelles dispositions (marquées d'une croix dans la liste proposée – formulaires différents selon transport de personnes ou de marchandises), nombre de PV, montant éventuel de la perception immédiate, autres sanctions prises.

Les données de ce formulaire ne sont pas traitées systématiquement dans une banque de données électronique. L'élément correspondant du *Contris*-software qui a été développé au sein du SPF et dans lequel les contrôleurs peuvent retrouver la législation qui les intéresse, n'est pas utilisé couramment pour l'instant, mais pourrait devenir dans le futur une source d'alimentation pour le datawarehouse. Toutefois, sans phase préalable au cours de laquelle une telle banque de données opérationnelle est développée et utilisée (et dans laquelle les données de contrôle du Service de Contrôle du Transport routier de la Direction générale Transport terrestre pourraient être d'abord systématiquement saisies), l'implémentation d'un datawarehouse commun est à notre avis difficilement réalisable.

Les données statistiques précitées sont établies périodiquement en MS Excel. Un second formulaire plus concis sert de base pour la collecte des données brutes. Etant donné qu'il s'agit là d'informations chiffrées agrégées qui font donc abstraction d'entreprises individuelles, ce processus n'est pas d'utilité directe pour l'alimentation du datawarehouse. Toutefois, dans le futur, il devrait être possible de générer les mêmes statistiques de manière automatique à partir du datawarehouse, ce qui pourrait constituer un allègement administratif. Cela ne sera certainement pas un élément de la présente phase de projet mais nous tiendrons déjà compte de cette éventuelle application future dans le projet de banque de données.

**Sources additionnelles/externes.** Nous avons déjà abordé ci-dessus la banque de données des autorisations, Transis. C'est une source précieuse d'information pour le datawarehouse, parce qu'elle fournit une liste complète des entreprises belges de transport. En reprenant les numéros de TVA, il est possible d'établir un lien avec d'autres sources de données, permettant ainsi d'enrichir plus avant les données.<sup>6</sup> En outre, elle permet de valider les données de contrôle émanant d'autres services ; l'introduction des données permettrait de tirer une liste d'employeurs contrôlés posant problème qui ne figurent pas dans la liste des transporteurs autorisés. Enfin, le système informatique permet aussi l'accès aux données des véhicules inscrits au nom d'un transporteur. L'introduction des données des plaques d'immatriculation

---

<sup>6</sup> Nous n'oublions pas non plus la Banque-Carrefour des Entreprises (BCE) qui aura aussi son influence sur le système Transis (cf. SPF Economie, PME, Classes moyennes et Energie, site internet : <http://mineco.fgov.be>).

dans le datawarehouse permet une validation supplémentaire des données entrées. Ces données sont rassemblées au sein du même SPF mais par la Direction pour l'Immatriculation des Véhicules (D.I.V.).

## 4.2. **Douanes & Accises**

**Missions de contrôle effectuées.** L'Administration des Douanes & Accises du SPF Finances exerce plusieurs missions de contrôle en matière de transport routier, tant au niveau de particuliers (des voitures sont aussi contrôlées) qu'au niveau d'entreprises. Le personnel de la brigade effectue à cet effet des contrôles le long de la route et chez les garagistes. Leurs compétences principales englobent:

- le contrôle des mouvements de marchandises sur le plan de la douane, des accises et/ou de la TVA;
- le contrôle du diesel routier par la prise d'échantillons;
- le contrôle de la taxe de circulation et de l'Eurovignette.

A cela s'ajoute un certain nombre d'autres missions (même non fiscales) (p.ex drogue, immigration, ...).

**Sources de données opérationnelles.** Au moyen d'un rapport de contrôle, des données sont recueillies sur le terrain :

- lieu, date, heure, brigade et équipe d'intervention;
- *nature du contrôle*: contrôle statique/dynamique/de routine/ciblé/opération;
- *employeur/transporteur*: numéro de TVA (éventuellement aussi celui du donneur d'ordre);
- *véhicule*: type (voiture, camionnette, camion, autobus, mobilhome, tracteur et/ou semi-remorque), plaque d'immatriculation et nationalité (séparés pour une combinaison tracteur/semi-remorque);
- *constatations/suite* : dispositions contrôlées (marquées d'une croix dans une liste de choix); indication si oui ou non il y a infraction et établissement d'un PV (si d'application).

Ensuite, ces rapports de contrôle sont traités électroniquement à la brigade et envoyés tous les mois au S.C.G.I. (Service central de gestion de l'information) où les données sont utilisées pour le suivi des brigades. Cette banque de données est implémentée en MS Access. Nous remarquons que l'identité du conducteur n'est pas enregistrée.

**Sources additionnelles/externes.** Le service dispose d'informations externes sur les entreprises provenant de l'administration centrale de la TVA et de l'application *Bel-first* du Bureau van Dijk. Il y a également un accès (limité) aux données relatives aux plaques d'immatriculation de la D.I.V.

## 4.3. **Office national de Sécurité sociale (O.N.S.S.)**

**Missions de contrôle effectuées.** L'O.N.S.S. est le gestionnaire d'une banque de données de tous les employeurs belges, appelé le répertoire des employeurs. Les services d'inspection de l'O.N.S.S. ont principalement une mission de contrôle administratif portant sur l'exactitude de la déclaration des travailleurs et des salaires par les employeurs (y compris ceux du secteur du transport) ; la plupart du temps, un tel contrôle est effectué à la demande des propres services internes et parfois à la demande d'une instance externe (notamment aussi l'I.L.S. et

l’I.S.). Pour l’instant, l’O.N.S.S. n’effectue pas (encore) lui-même de contrôles routiers. Les catégories de contrôles qui sont potentiellement des données importantes pour le datawarehouse sont :

- le contrôle de l’assujettissement (travailleurs): vrais ou faux indépendants ;
- le contrôle de l’assujettissement (salaires): indemnités de frais comme salaire déguisé;
- la non-introduction de la déclaration (“muets”).

**Sources de données opérationnelles.** A l’administration centrale de Bruxelles, il y a une banque de données et une application Oracle-/Powerbuilder, appelée *DBEO*, qui permet de suivre les demandes de contrôle et la suite (majoritairement administrative) qui y est donnée.

On y conserve les données suivantes:

- date de mise en route et de clôture, contrôleur exécutant;
- *nature du contrôle*: origine (ex. instance qui a demandé le contrôle);
- *employeur/transporteur* (lié au fichier des employeurs): numéro ONSS, nom, adresse et secrétariat social;
- *constatations*: objet du contrôle (exemples : voir ci-dessus) et suite (notamment, déclaration, avis rectificatif – éventuellement avec mention du montant –, pro justitia, avertissement, enquête complémentaire).

Chaque contrôleur dispose d’une application Lotus Notes pour le support de ses activités, la rédaction des rapports et la gestion des dossiers ; ce dernier point n’a pas d’intérêt pour le datawarehouse.

**Sources additionnelles/externes.** Le *répertoire des employeurs de l’O.N.S.S.* est une banque de données importante qui contient les données d’identification de base de chaque employeur et qui indique la catégorie d’employeurs à laquelle il appartient. Cette information est donnée par l’employeur lors de son inscription auprès de l’O.N.S.S. ; l’historique des modifications est également conservé. Ce fichier des employeurs constitue une source de données importante pour le réseau géré par la Banque-Carrefour de la Sécurité sociale (BCSS). Tant l’I.L.S. que l’I.S. l’utilisent lors de leurs activités de contrôle. Nous remarquons également que le numéro d’entreprise unique est inscrit dans le fichier, de sorte qu’un lien peut être établi avec la Banque-Carrefour des Entreprises (BCE).

Une deuxième banque de données importante est la déclaration DIMONA et le registre électronique du personnel qui en résulte. La déclaration DIMONA a pour but de transmettre immédiatement par voie électronique à l’O.N.S.S. le début et la fin d’une relation de travail. DIMONA est l’acronyme de Déclaration Immédiate/Onmiddellijke Aangifte et est accessible via le portail de la Sécurité sociale, [www.securitesociale.be](http://www.securitesociale.be). La déclaration contient des informations concernant l’employeur, le travailleur (si possible identifié par son numéro d’identification unique pour la sécurité sociale), le lieu d’occupation d’un étudiant, le numéro de la commission paritaire et la date à laquelle le travailleur est entré en fonction ou a quitté son employeur. Grâce à DIMONA, certaines obligations incombant à l’employeur en matière de documents sociaux ont été simplifiées ou supprimées (notamment suppression du registre papier du personnel). L’ensemble de la banque de données DIMONA est évidemment consultable par les contrôleurs de l’O.N.S.S., de l’I.L.S. et de l’I.S. Une des activités des contrôleurs de l’I.L.S. et de l’I.S. est de vérifier systématiquement si les conducteurs visés lors d’un contrôle routier sont bien inscrits comme travailleur dans le système DIMONA (cf. infra).

Une source externe additionnelle et intéressante à laquelle tant l’O.N.S.S. que l’I.L.S. et l’I.S ont accès, est la *Banque-Carrefour du droit pénal social*. Elle permet de consulter des données concernant les sanctions pénales appliquées en matière de droit du travail. Ce système doit permettre aux contrôleurs de suivre comment les dossiers transmis sont traités par les auditeurs du travail. A ce propos, nous insistons sur le fait qu’une telle information n’apparaît pas dans aucune des sources de données opérationnelles étudiées dans cette section et les suivantes : on se limite toujours à la mention “pro justitia”.

#### **4.4. Inspection des Lois sociales (I.L.S.)**

**Missions de contrôle effectuées.** Le service Inspection des Lois sociales (I.L.S.) fait partie du SPF Emploi, Travail et Concertation sociale (l’ancien Ministère fédéral de l’Emploi et du Travail), et est un des trois services d’inspection du travail. Il effectue des contrôles dans le domaine du droit du travail. L’I.L.S. compte une administration centrale et 24 directions externes (13 dans la Région flamande, 10 dans la Région wallonne et 1 dans la Région de Bruxelles-Capitale). Deux directions ont reçu spécifiquement l’ordre de s’occuper du secteur du transport, à savoir celles de Malines et de Huy. Quelques infractions typiques qui peuvent être constatées par le service concernent : le travail en noir (non-communication à DIMONA (voir plus haut), paiement en noir des prestations), faute dans le paiement du salaire, des jours fériés ou des indemnités de déplacement, la mise à jour incorrecte des feuilles de prestations, l’établissement incorrect des comptes individuels, l’emploi de travailleurs de nationalité étrangère sans permis de séjour ou sans carte ou permis de travail et le non-respect des règlements européens en matière de temps de conduite et de repos et d’usage du tachygraphe, une compétence que l’Inspection partage avec la plupart des autres services.

**Sources de données opérationnelles.** Les données suivantes sont collectées:

- date de mise en route et de clôture, inspecteur;
- *nature du contrôle*: motif, origine, protocole;
- *travailleur(s)/conducteur* : pas disponible directement (cf. infra); mais bien le nombre concerné d’hommes/femmes, d’ouvriers/employés, de travailleurs à temps partiel, etc. par le contrôle;
- *employeur/transporteur*: numéro ONSS, code NACE, nom, adresse, forme juridique, etc.;
- *véhicule*: pas disponible directement (cf. infra); mais bien le nombre de véhicules contrôlés (1 pour un contrôle routier) et le nombre de disques contrôlés;
- *constatations*: pour chaque dossier, les différentes visites et leurs résultats sont enregistrés, avec mention du lieu (entreprise, secrétariat social ou autre – dans cette catégorie, on trouve les contrôles sur la route), code des dispositions légales contrôlées, la suite donnée (avertissement, régularisation, pro justitia, pas d’irrégularités ou transmission à une autre direction) et, si d’application, le montant de régularisation ou le numéro de PV.

Nous avons remarqué que la banque de données ne comprend actuellement pas de champs distincts qui permettraient d’identifier automatiquement un conducteur et son véhicule qui a été contrôlé le long de la route : le nom, le numéro d’identification unique pour la sécurité sociale (ou numéro du registre national), la plaque d’immatriculation et d’autres données supplémentaires sont bien notés par le contrôleur sur le terrain, mais ces éléments peuvent tout au plus être inscrits ultérieurement dans un champ de texte de commentaire et ne sont donc pas directement extractibles de la banque de données définitive (celle-ci n’étant pas ciblée exclusivement sur le domaine du transport).

Le système actuel est implémenté en Omnis (précédemment du Blythe Software anglais, maintenant chez Raining Data; voir <http://www.omnis.net>), un environnement intégré de développement d'applications et de banque de données qui prévoit aussi les facilités d'extractions nécessaires. Une difficulté est que les données ne sont pas actuellement centralisées dans une grande banque de données; des fichiers semestriels sont recueillis toutefois sur bandes à partir des diverses directions locales. C'est pourquoi, il est proposé de saisir dans une première phase de datawarehousing, uniquement les données provenant de Malines et de Huy. Dans une deuxième phase, on pourrait instaurer un système qui extrait périodiquement les fichiers pertinents de toutes les directions pour le datawarehouse, données qui seraient ensuite transmises électroniquement à l'administration centrale qui les fusionnerait et les enverrait à la Direction générale Transport terrestre.

**Sources additionnelles/externes.** L'I.L.S. utilise le répertoire des employeurs de l'O.N.S.S., la banque de données DIMONA et a également accès à la Banque-Carrefour du Droit pénal social ; ces systèmes ont été décrits dans la section 4.3.

#### **4.5. Inspection sociale (I.S.)**

**Missions de contrôle effectuées.** L'Inspection sociale fait partie du SPF Sécurité sociale et a pour mission de contrôler le respect de la législation en matière de sécurité sociale des travailleurs. A cet effet, elle enquête auprès des employeurs, des travailleurs et des organismes de sécurité sociale. En ce qui concerne plus spécifiquement le secteur du transport, elle effectue également des contrôles routiers. Ses compétences de contrôle s'étendent aux catégories suivantes:

- cotisations de sécurité sociale: déclaration correcte travailleurs (DIMONA) et salaires; enquête sur indépendants apparents;
- octroi vacances annuelles et paiement pécule de vacances;
- non-délivrance documents assurance maladie-invalidité;
- obligations en matière d'allocations familiales;
- mise à jour et conservation des documents sociaux (registre du personnel, registre de présence, comptes individuels);
- accidents de travail : assurance et déclaration accident de travail;
- travail à temps partiel : communication et respect de l'horaire de travail;
- occupation illégale/séjour de travailleurs de nationalité étrangère;
- empêchement du contrôle.

Nous remarquons que les contrôles des tachygraphes n'y figurent pas.

**Sources de données opérationnelles.** Bien que des données additionnelles concernant les conducteurs et leurs véhicules puissent être recueillies sur le terrain, et plus particulièrement lors de contrôles routiers, la banque de données actuelle (*SIS*) est axée particulièrement sur le suivi des missions de contrôles au niveau des employeurs. On retient éventuellement le nom, le numéro de registre national et l'adresse d'un intéressé individuel mais aucune donnée concernant le véhicule. L'employeur qui fait l'objet d'une enquête est identifiable sur la base de son numéro d'O.N.S.S. et de TVA. Les données de l'employeur sont complétées avec des données d'exploitation et de contact officielles ; en outre, on mentionne les fonds, la compagnie d'assurances, le secrétariat social, etc., auxquels l'employeur est affilié. En ce qui concerne la nature du contrôle, on indique s'il a été demandé par une autre instance (si oui,

laquelle), à la suite d’une plainte ou à l’initiative de l’inspection sociale, par quelle cellule il a été effectué (ex. cellule transport d’un bureau précis), et où d’application, dans le cadre de quel protocole (ex. “transport”, en ce qui concerne spécifiquement les contrôles routiers).

Les constatations et la suite qui y est donnée sont encodées comme suit. Une distinction est faite entre un nombre de catégories principales (– dans la sous-section précédente : missions de contrôle effectuées –), et ensuite entre un nombre de sous-catégories. Pour chacune des catégories principales, on indique : 1 (vérifié et en ordre); 2 (vérifié et régularisé ou avertissement); 3 (vérifié et PV établi). Pour les constatations du type 2 et 3, on indique selon la catégorie précise: oui/non; le montant de la régularisation; le nombre de personnes concernées; une certaine combinaison des éléments précités. Enfin, il est établi une liste des services qui sont mis au courant.

Nous avons remarqué qu’une nouvelle version de système est actuellement en cours de développement et implémentée en Oracle/Business Objects. Des informations complémentaires à ce sujet doivent encore être recueillies.

**Sources additionnelles/externes.** Les sources de données externes auxquelles l’I.S. a accès sont les mêmes que celles pour l’I.L.S. (voir plus haut).

#### **4.6. Police fédérale et Police locale**

En ce qui concerne les partenaires de la Police fédérale et de la Police locale, il convient d’abord d’examiner de façon plus approfondie le cadre légal dans lequel un échange de données (autres que les données rendues anonymes) serait possible. L’objectif du datawarehouse suppose, en effet, que certaines données d’identification (p. ex. la plaque d’immatriculation) soient communiqués. C’est la raison pour laquelle les systèmes (partiellement encore en cours de développement) de la Police fédérale et de la Police locale ne sont pas repris en détail dans le présent document. Il est toutefois évident que pour ces deux partenaires, le registre des PV pourrait être une source d’information utile pour certaines catégories de constats relatifs au transport de marchandises et de personnes par la route. Nous pensons ici par exemple à des données de PV relatives au tachygraphe, au contrôle technique, au chargement, à l’ADR et au permis de conduire.

### **5. Projet de modèle et commentaire**

Malgré le fait que certains systèmes sources ne sont pas ciblés spécifiquement sur le domaine du transport et ne peuvent donc fournir que peu d’informations concernant le véhicule ou le conducteur, nous sommes toutefois d’avis que la surcharge administrative générée par le transfert manuel de telles données ne semble pas contrebalancer les données plus complètes qui pourraient ainsi être obtenues. C’est pourquoi, nous voudrions opter pour un système dans lequel le datawarehouse est alimenté principalement par l’extraction automatique à partir d’un certain nombre de systèmes sources décrits dans la section 4 (uniquement en ce qui concerne les données de contrôle du Transport terrestre et celles fournies dans le cadre de l’arrangement administratif 'Euro Contrôle Route', il faudra peut-être recourir (provisoirement) à une introduction manuelle – ou dans le premier cas, à un développement plus poussé d’un système de banques de données). Dans notre proposition modèle, nous tentons toutefois

de tenir compte au maximum de l’extension future éventuelle des données disponibles dans les systèmes sources opérationnels.

Nous voulons aussi éviter des redondances inutiles par rapport aux systèmes existants. Les divers services disposent déjà chacun d’une large compilation de données d’entreprises pertinentes pour leurs activités. Il ne nous semble pas indiqué de dupliquer toutes ces données dans le datawarehouse. Nous préférons proposer de prévoir les champs de données nécessaires qui permettront d’identifier sans équivoque une entité (entreprise, véhicule, conducteur) dans les systèmes actuels, ainsi que ceux qui sont souhaités comme critères de sélection pour les recherches et les analyses, afin d’éviter des redondances inutiles.

Le projet de modèle proposé est basé dans les grandes lignes sur les méthodes décrites dans la section 3. Il distingue deux tables de faits centrales, respectivement pour les contrôles effectués (*controles*) et leurs résultats (*vaststellingen*), autour desquelles s’agglutinent un certain nombre de tables pour les dimensions temps (*tijd*), véhicule (*voertuig*), conducteur (*bestuurder*), entreprise (*onderneming*) et disposition légale (*bepaling*); à cette dernière table de dimension sont liées hiérarchiquement les tables Disp. Catégories Agora (*Bep\_AgoraCategorieen*) et Disp. Catégories Euro (*Bep\_EuroCategorieen*) (cf. schéma flocon de neige, p.9). La structure relationnelle qui en découle est illustrée dans la figure 6 (elle montre une impression écran d’un prototype simple établi en MS Access, voir fichier dans annexe : ‘AgoraDB.mdb’).

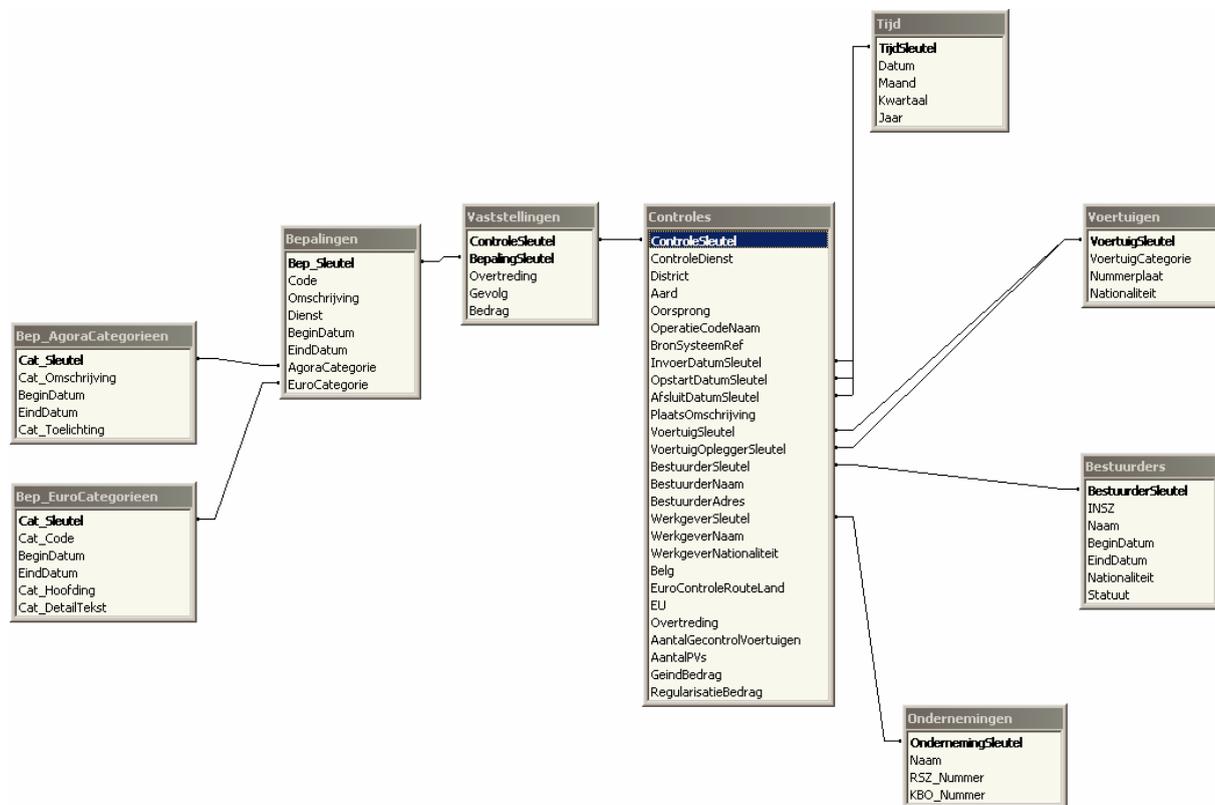


Figure 5 : projet de proposition datawarehouse; structure relationnelle

Nous commentons plus avant quelques-uns des choix, et en première instance ceux qui dérogent quelque peu au cadre méthodologique décrit dans la section 3. Une spécification plus complète des tables et de leurs attributs se trouve en annexe (voir *Appendice A: fichier ‘Agora-01-082\_Appendice A\_DW-Model.xls’*).

**Contrôles vs constatations.** L’enregistrement des contrôles effectués et des enquêtes n’est pas groupé dans une seule table de faits mais dans deux tables : contrôles (*controles*) et constatations faites à cette occasion (*vaststellingen*). Prenons l’exemple d’un contrôle : un conducteur  $x$  dans un véhicule  $y$  travaillant pour l’entreprise  $z$  est arrêté par un contrôleur du service  $d$  à la date  $t$ . Pour chacune des dispositions légales contrôlées (ayant pour résultat la constatation d’une infraction, cf. infra), il y a une ligne dans la table des constatations, avec mention de la suite qui y est donnée. Sans choisir une telle solution normalisée et sans scinder la table des faits, il faudrait, pour chaque constatation effectuée lors du contrôle, répéter les clés ‘temps’, ‘conducteur’, ‘véhicule’ et ‘employeur’ ainsi qu’un certain nombre d’autres données non clés. En outre, beaucoup d’analyses qui sont concevables ne doivent pas se faire au niveau des constatations individuelles mais uniquement au niveau des contrôles effectués. Pour modéliser ensuite le rapport 1-sur- $n$  entre les tables résultantes, nous définissons dans la table de contrôle une clé primaire non significative, appelée la « clé de contrôle » (*Controle-Sleutel*), à laquelle nous référons dans la table des constatations (*vaststellingen*).

**Contrôles dans les entreprises vs contrôles sur les routes; rapport avec travailleurs/véhicules.** Pour garder la structure la plus simple possible, nous proposons de grouper les contrôles sur les routes et dans les entreprises dans la même table de faits (comme c’est le cas dans la plupart des systèmes sources). Cette table contient un champ « Nature » (*Aard*), qui indique de quel type de contrôle il s’agit. Pour les contrôles routiers, la possibilité d’identifier le conducteur et son véhicule est prévue. Les contrôles dans les entreprises concernent toutefois en règle générale plusieurs travailleurs et/ou véhicules. En principe, ils font l’objet de constatations spécifiques (i.e., on pourrait pour chaque infraction constatée indiquer de quel travailleur individuel et/ou de quel véhicule il s’agit). Dans la pratique, cette information n’est toutefois pas disponible dans les systèmes sources. On ne retrouve des informations que sur le nombre de travailleurs et de véhicules concernés. L’I.L.S. et l’I.S. enregistrent tant le nombre total de travailleurs concernés (classés ensuite par statut) que les nombres partiels par constatation. Le nombre total de véhicules contrôlés (donc pas par constatation) n’apparaît que dans la banque des données de l’I.L.S. On opterait donc au maximum pour la solution suivante, à savoir, là où les données sont disponibles : le nombre total de véhicules et de travailleurs contrôlés au niveau de la table “contrôles”, et plus spécifiquement, chaque fois aussi le nombre de travailleurs concernés lors de constatations individuelles. Si dans le futur, davantage de données détaillées étaient disponibles, le projet de datawarehouse conçu peut être adapté ultérieurement afin de les intégrer à toutes fins utiles.

**Dimension ‘entreprise’.** Comme expliqué dans la section 3.2, l’information textuelle est en principe reprise dans les tables de dimension. Dans certains cas, la duplication dans la table de faits est admise, par exemple pour des raisons d’efficacité ou comme solution pour le modelage historique des modifications de dimensions. Dans notre cas, il se pose toutefois un problème supplémentaire, à savoir la délimitation et l’identification unique du contenu des tables de dimension. Prenons par exemple la population d’entreprises de transport. On peut identifier sans équivoque les entreprises belges au moyen du numéro de TVA (ou de la BCE) ou éventuellement aussi à partir du numéro O.N.S.S. Identifier une entreprise par son nom

s'avère plus risquée parce que le nom peut être noté de manière inconsistante. En ce qui concerne les entreprises étrangères, ces moyens d'identification ne sont pas disponibles ; certains services ne ciblent d'ailleurs que des employeurs belges. Toutefois, le rapportage 'Euro Contrôle Route' prévoit la communication du nom de l'entreprise lorsque des infractions (graves) commises par des transporteurs routiers étrangers sont constatées. Plutôt que d'essayer d'enregistrer toutes les entreprises sous forme de lignes dans la table de dimension (avec le risque d'avoir des doublons et par conséquent des analyses fautives – la majorité des données ne peut en outre être utile que pour les entreprises belges), nous proposons d'enregistrer uniquement un sous-ensemble bien défini d'entreprises contrôlables dans la table de dimension « entreprises » (à savoir les entreprises de transport autorisées telles que celles reprises dans la banque de données Transis) ; en ce qui concerne les autres entreprises, on indiquerait directement que le nom et la nationalité dans la table des faits « contrôles », et ce surtout à des fins de rapportage et non d'analyse.<sup>7</sup> Dans le futur, un lien peut être établi avec diverses sources externes d'informations sur les entreprises via le numéro de l'entreprise, par exemple à des fins d'enquête.

**Dimension 'conducteur' (contrôles routiers).** Un problème analogue se pose pour les conducteurs contrôlés. En ce qui concerne les travailleurs assujettis au régime de sécurité sociale belge, le numéro d'identification unique pour la sécurité sociale (ce qui correspond pour les personnes habitant en Belgique au numéro du registre national) peut servir dans le futur de moyen d'identification. En ce qui concerne les travailleurs d'entreprises étrangères de transport, d'autres modes d'identification sont évidemment d'application. En outre, il faut signaler que tous les services ne notent pas (ou ne peuvent pas noter) le numéro unique ou le numéro du registre national et qu'actuellement aucun des services ne peut offrir automatiquement cette donnée à partir de la banque de données opérationnelle disponible. La protection de la vie privée est un aspect complémentaire qu'il faut encore examiner dans ce contexte. Dans l'optique des finalités d'utilisation du datawarehouse (où l'analyse du comportement de l'entreprise prime sur celle des travailleurs individuels), nous proposons, dans une première phase, de commencer sans la table de dimension « conducteur » et de l'intégrer plus tard dès que les objections pratiques et légales auront été balayées. Pour satisfaire aux exigences du rapportage 'Euro Contrôle Route', nous prévoyons déjà deux champs pour le nom et l'adresse du conducteur contrôlé dans la table des faits. Ils ne doivent pas être adaptés ultérieurement mais reflètent la situation au moment du contrôle.

**Dimension 'véhicule' (contrôles routiers).** Pour les contrôles routiers, deux champs sont prévus dans la table des faits « contrôles » en ce qui concerne le véhicule contrôlé (deux tels champs sont nécessaires pour les combinaisons tracteur/semi-remorque). L'identification unique peut se faire sur la base de la plaque d'immatriculation (et de la nationalité). En ce qui concerne les véhicules immatriculés en Belgique, cette table de dimension peut être remplie sur la base de la banque de données des plaques d'immatriculation de la DIV et ensuite être complétée périodiquement (pas de suppressions, cf. conservation de l'historique des contrôles). Cette procédure permet de valider les données de contrôle recueillies où apparaît l'information relative à la plaque d'immatriculation. Pour tout véhicule étranger non contrôlé auparavant, une ligne peut chaque fois être ajoutée. En ce qui concerne les données de l'I.L.S.

---

<sup>7</sup> Une autre solution serait de prévoir une table de dimension supplémentaire pour la catégorie restante d'entreprises.

et l'I.S., il faut examiner la possibilité d'un système permettant de transmettre systématiquement aussi les données d'identification des véhicules.<sup>8</sup>

**Dimension 'temps'.** Pour les contrôles dans les entreprises, on note la date de début et de clôture du contrôle ; pour les contrôles routiers, il s'agit du même jour. Pour faciliter la consultation, quelques champs redondants sont prévus dans la table temps pour le mois, le trimestre et l'année. Il faut souligner que des contrôles de longue durée dans une entreprise entraînent une difficulté additionnelle d'extraction : des contrôles qui ne sont pas encore clôturés peuvent subir des modifications dans la période suivante (par ex. nouvelles constatations à la suite de visites supplémentaires). Les procédures d'extraction doivent tenir compte de cette possibilité et ne peuvent pas recueillir les enregistrements ajoutés ou modifiés dans la nouvelle période sans certaines adaptations des enregistrements précédents. A moins de décider de charger les données qu'après la clôture du contrôle, il faut par la force des choses déroger au principe que seul des nouveaux faits sont ajoutés de préférence à un datawarehouse, sans modification des existants.

**Dimension 'Dispositions'.** Chaque service a son propre schéma de codage pour les dispositions légales dont il contrôle le respect. Pour rendre les résultats compréhensibles et permettre l'analyse d'infractions relevant de plusieurs services, il est indiqué de répertorier ces schémas dans un schéma de classification commun. Toutefois, il reste peut-être nécessaire de conserver aussi les codages propres à chaque service, surtout pour des considérations de flexibilité : la répartition peut être ainsi adaptée simplement sans rupture avec les anciennes données ou des classifications additionnelles peuvent être prévues. Dans la table 'dispositions' (*bepalingen*), on garde les dispositions légales spécifiques à chaque service auxquelles il est fait référence dans la table des constatations.<sup>9</sup> Deux champs dans la table des dispositions réfèrent à deux sortes de classification que nous estimons actuellement utiles pour le datawarehouse, à savoir le codage utilisé dans le cadre de l'arrangement Euro Contrôle Route (voir table Disp. catégories Euro - *Bep\_Eurocategorieen*), et une nouvelle classification pour le projet même de datawarehouse (voir table Disp. catégories Agora - *Bep\_Agoracategorieen*).

**'Bep\_Eurocategorieen' (dispositions catégories Euro).** Cette première table contient les codes prescrits pour l'échange d'informations entre les pays participant à Euro Contrôle Route.

**'Bep\_Agoracategorieen' (dispositions catégories Agora).** La classification ci-dessus ne couvre pas tout l'éventail des contrôles auxquels sont soumises les entreprises belges de transport par les partenaires du projet de datawarehouse, d'où la nécessité d'une classification

---

<sup>8</sup> Nous voulons souligner que la non-saisie ou la saisie non structurée d'un certain nombre de données importantes d'identification dans les systèmes sources font que le datawarehouse, dans le contexte actuel, ne contiendra inévitablement que des données incomplètes. Par conséquent il serait recommandé que tous les partenaires, dans la mesure du possible et dans les limites de leurs compétences spécifiques, s'efforcent d'enregistrer systématiquement certaines de ces données additionnelles d'identification (ex. numéro de sécurité sociale du conducteur contrôlé et plaque d'immatriculation) dans leurs versions de système futures.

<sup>9</sup> Des accords clairs doivent être pris avec chacun des partenaires en ce qui concerne la maintenance (i.e., la transmission de codes nouvellement introduits).

complémentaire, sur la base de laquelle on aurait une image plus large du type d'infractions constatées et de leur fréquence. Sur la base des informations recueillies auprès des divers services, nous voulons proposer la classification Agora initiale suivante :

1. Absence d'autorisation de transport;
2. Lettre de voiture/ document de transport;
3. (Sur)charge;
4. Réglementation A.D.R. ;
5. Transport exceptionnel : autorisation;
6. Transport de personnes: attestation compte propre;
7. Transport de personnes: autorisation communautaire / contrôle de qualité/ permis d'exploitation/ feuille de route;
8. Contrôle technique/ certificat de visite;
9. Tachygraphe: temps de conduite et de repos & interruptions;
10. Tachygraphe: utilisation appareil;
11. Code de la route;
12. Permis de conduire/ sélection médicale;
13. Attestation de conducteur pour ressortissant d'un pays tiers;
14. Immatriculation véhicule/ plaque d'immatriculation;
15. Assurance;
16. Taxe de circulation;
17. Eurovignette;
18. Contrôle du gasoil routier;
19. Documents de douane : carnet TIR/ documents T/ document d'accompagnement;
20. Accises/ T.V.A.;
21. Déclaration DIMONA début/fin relation de travail;
22. Déclaration salaires ONSS (DMFA);
23. Assujettissement : faux indépendants/ salaires déguisés ;
24. Règlement du travail : établissement/ conservation;
25. Travail partiel: horaire;
26. Conservation documents sociaux : registre du personnel/registre de présences/ comptes individuels/ ...;
27. Paiement du salaire et indemnités diverses;
28. Jours fériés : paiement/ récupération;
29. Vacances annuelles/ pécule de vacances ;
30. Accidents de travail : assurance/ déclaration;
31. Occupation illégale/séjour travailleurs étrangers;
32. Assurance maladie-invalidité;
33. Allocations familiales;
34. Empêchement contrôle;
35. Décret sur l'emploi des langues;
36. Autres.

Lors de leur adhésion respective au système, les services participants doivent valider cette liste et si nécessaire l'adapter ensuite en fonction de l'évolution des besoins. Un exercice important auquel chaque partenaire doit s'astreindre est de déterminer à quelle catégorie appartient chaque disposition individuelle (telle qu'elle est utilisée dans leur propre banque de données). Cette information de classification peut ensuite être introduite dans la table de dispositions du datawarehouse (comme illustré dans la section suivante). Un autre élément concerne

l'incorporation éventuelle de mécanismes de filtrage qui n'autorisent l'approvisionnement en informations détaillées que pour les matières relevant de chaque service en particulier.

## **6. Proof-of-concept : traitement des données douanières**

Pour illustrer la présente proposition de datawarehousing, un processus unique de ETL (extraction, transformation, loading) a été implémenté pour un des sous-trajets, plus particulièrement pour un ensemble partiel des données de contrôle recueillies par les Douane & Accises (cf. section 4.2). Un tel exercice peut d'ailleurs mettre davantage en lumière le degré de difficulté de l'ensemble du projet ou les problèmes imprévus avec les banques de données sources. En outre, il est utile de développer une interface simple à l'intention de l'utilisateur pour la consultation des résultats. Un tel moyen de démonstration permet à l'utilisateur final de se faire une meilleure idée des possibilités de query et d'analyse et des limitations du datawarehouse final.

### **6.1. Implémentation ETL**

Nous utilisons à titre de test les données des contrôles routiers recueillies durant le premier trimestre 2003. La table 1 donne la liste des champs de données disponibles dans la table source. La table 2 indique le codage utilisé (cf. champs 'SORTE DE CONTROLE' et 'CONS-TATATION') et indique la référence de la classification Agora proposée dans la section 5 (cf. numéros d'ordre) et de Euro Contrôle Route (avec lequel il y a peu de points communs ; on n'a pu établir un lien qu'avec D03 – "Absence of technical inspection"). Vu qu'il n'entrait pas encore dans nos intentions d'implémenter une procédure à répéter dans le temps, il n'a pas été prévu de mécanisme pour saisir lors d'une extraction ultérieure éventuelle que les données plus récentes ou modifiées. Ces données sources ont été fournies dans un format MS Access et sont chargées après traitement intensif dans le prototype précité (également en MS Access). Le code ETL a été développé au moyen de Borland Delphi 7 (voir: <http://www.borland.com/delphi/>), vu que nous ne disposions pas d'un purpose tool spécial.

Attribut	Type	Description
<b>Contrôles (routiers) brigades de douane</b>		
NUMERO D'ORDRE	Nombre	Numéro d'ordre contrôle dans le mois (par code fors)
CODE FORS	Texte	Code Fors de la brigade
DATE INPUT	Date	Date à laquelle le rapport de contrôle a été entré
DAT CONTROLE	Date	Date à laquelle le contrôle a été effectué
HEURE	Texte	Indication heure contrôle
EQUIPE	Texte	Plaque d'immatriculation du véhicule de service ou nom d'appel du team fixe
POINT DE CONTROLE	Texte	Description lieu du contrôle (ex, "Gand-Bruges E17")
IMPORTATEUR	Texte	Numéro TVA importateur/destinataire*
EXPORTATEUR	Texte	Numéro TVA exportateur/expéditeur*
TRANSPORTEUR	Texte	Numéro TVA transporteur*
VEHICULE	Texte	Information structurée concernant véhicule contrôlé; type1/plaque d'immatriculation1/nationalité1;type2/plaque d'immatriculation2/nationalité2 (types : 1: voiture; 2: camionnette; 3: camion; 4: autobus; 5: mobile home; 6: tracteur; 7: semi-remorque; 8: autres)
NATIONALITE	Texte	Description nationalité*
TYPE DE CONTROLE	Nombre	Type de contrôle: 1: dynamique (interception du véhicule); 2: statique 1 (le long de la route; sélection); 3: statique 2 (le long de la route; tous)
NATURE DU CONTROLE	Nombre	Nature contrôle: 1: routine; 2: ciblé (sur une disposition spécifique, cf. infra); 3: opération (evt. coordonnée)
NOM DE CODE	Texte	Detail nature contrôle: détermination contrôle ciblé ou nom de code opération (ex. 210 ou "ZW VERV LOKPOL ZOTTEGEM")
SORTE DE CONTROLE	Texte	Liste codes dispositions contrôlées (chaque sous-code est précédé d'un point -virgule ex, ";210;320;340"); voir ailleurs pour aperçu des codes utilisés
CONSTATATION	Texte	liste codes dispositions pour lesquelles une infraction a été constatée (doit être une sous-liste de la liste ci-dessus)
A115	Texte	Info détaillée disposition 115 (transport produits d'accises - autres: ...)
A122	Texte	Info détaillée disposition 122 (droits d'importation - produits visés: ...)
A132	Texte	Info détaillée disposition 132 (contrôle TVA - produits visés: ...)
A145	Texte	Info détaillée disposition 145 (missions non fiscales - autres: ...)
A380	Texte	Info détaillée disposition 380 (contrôle moyen de transport - autres: ...)
NUMERO DE DOSSIER	Texte	(numéro de référence pour usage interne)
BATCH	Oui/non	(usage interne)

\*pas disponible ou seulement de manière non systématique dans le set de test

Table 1: spécification table source données de contrôle douane (voir aussi Appendice B)

<b>Code</b>	<b>Description</b>	<b>Catég. Agora.</b>	<b>Catég. Euro.</b>
111	Transport produits d'accises – Tous les produits	20	–
112	Transport produits d'accises – Alcool	20	–
113	Transport produits d'accises – Huiles minérales	20	–
114	Transport produits d'accises – Tabac	20	–
115	Transport produits d'accises – Autres	20	–
121	Droits d'importation – Tous les produits	19	–
122	Droits d'importation – Produits visés (...)	19	–
131	Contrôle TVA – Tous les produits	20	–
132	Contrôle TVA – Produits visés (...)	20	–
133	Contrôle TVA – rédaction PV 108	–	–
134	Contrôle TVA – rédaction PV 109	–	–
141	Missions non fiscales – Drogues	36	–
142	Missions non fiscales – Immigration	36	–
143	Missions non fiscales – CITES	36	–
144	Missions non fiscales – Patrimoine culturel	36	–
145	Missions non fiscales – Autres	36	–
210	Contrôle carburant – Sur la voie publique	18	–
220	Contrôle carburant – Dans les stations-services	18	–
230	Contrôle carburant – Petit échantillon diesel pour labo	18	–
240	Contrôle carburant – Grand échantillon diesel pour labo	18	–
310	Contrôle moyen de transport – Taxe de circulation	16	–
311	Contrôle moyen de transport – Taxe de circulation complémentaire	16	–
320	Contrôle moyen de transport – Trafic rémunéré	36	–
330	Contrôle moyen de transport – Assurance	15	–
340	Contrôle moyen de transport – Eurovignette	17	–
350	Contrôle moyen de transport – Certificat d'immatriculation	14	–
360	Contrôle moyen de transport – Contrôle technique	8	D03
370	Contrôle moyen de transport – Permis de conduire	12	–
380	Contrôle moyen de transport – Autres	36	–
410	PV concernant douane et accises – rédaction PV	–	–
510	Non-poursuite du contrôle – Moyen de transport vide (camion, coffre, ...)	–	–
520	Non-poursuite du contrôle – Autres produits que ceux ciblés	–	–
530	Non-poursuite du contrôle – Empêchement service, délit de fuite, etc.	34	–

Table 2: schéma codage; mapping vers classification Agora et Euro Contrôle Route

Vu l'objectif de l'exercice, la transformation des données et le data-cleaning a demandé le plus de travail.

L'analyse (parsing) et le nettoyage des informations sur les véhicules (la figure 7 montre un exemple simple de la transposition), qui servent de base pour remplir la table de dimension 'Véhicules' demandent pas mal de codes pour un certain nombre de raisons. Premièrement, le codage du type de véhicule (1: voiture, etc.) n'est pas toujours bien suivi (on a également retrouvé des descriptions en texte). C'est pourquoi notre code distingue chaque fois un certain

nombre de synonymes possibles pour chaque type de véhicule. Deuxièmement, les immatriculations ont d'abord dû être traitées pour éviter des doublons, notamment en éliminant les espaces, les traits d'union, etc. Troisièmement, la nationalité du véhicule n'est pas enregistrée de manière conséquente. Nous trouvons par exemple 'belge', 'belgië', 'belgique', 'B', etc. pour indiquer la nationalité belge. Pour résoudre ce problème, nous avons rassemblé dans un tableau ('NationMapping.mdb') toutes les descriptions (environ 130) apparues dans le set de test ; chaque description a été reliée à un des codes standards des pays. Deuxièmement, nous avons recherché une nationalité correspondante dans la liste standardisée des pays (abréviation ISO de trois lettres ou nom complet du pays). Si l'information sur la nationalité manque dans le champ 'VEHICULE', on tente d'utiliser le contenu du champ 'NATIONALITE' pour l'identification. Enfin, nous avons opté pour la nationalité belge lorsque aucune indication n'est donnée et si l'immatriculation correspond au format utilisé en Belgique (trois lettres suivies de trois chiffres). Nous constatons que le traitement des données est souvent davantage la règle que l'exception lors de la réalisation de projets de datawarehousing, et que par conséquent, 100% d'informations exactes peuvent rarement être garanties. La qualité de l'information recueillie peut être améliorée à terme en appliquant ces aspects de cleaning le plus possible à la source même, mais la littérature sur le datawarehousing nous apprend qu'il ne faut pas attendre cela la plupart du temps.

VEHICULE: ";6/HIX220/België;7/UJJ317/België" (information source douane)

⇒

Contrôles			
...	Clé Véhicule	Clé Véhicule Semi-remorque	...
...	82	83	...

Véhicules			
Clé Véhicule	Catégorie Véhicule	Plaque d'immatriculation	Nationalité
82	1	HIX220	150
83	1	UJJ317	150

Figure 7 : exemple transposition données véhicule

L'identification de l'employeur fut un autre problème (pour remplir la table de dimension 'entreprises'). Bien qu'un champ soit prévu dans la table source des données pour l'enregistrement du numéro de TVA, cette information n'a pas été recueillie jusqu'à présent sur le terrain au cours de la période de test (premier trimestre 2003). Cette limitation fait que dans le contexte de cette étude, il a été impossible de lier les contrôles et les infractions constatées aux entreprises individuelles. Si on faisait appel à une seconde source, à savoir la banque de données des plaques d'immatriculation de la DIV, l'information recueillie sur les véhicules (voir plus haut) permettrait bien d'identifier – quoique indirectement - les transporteurs belges. Ce n'est évidemment pas une solution pour les entreprises étrangères (et le rapportage européen). Les partenaires concernés doivent étudier le problème pour déterminer dans quelle mesure il peut être résolu dans l'avenir par une adaptation des procédures sur le terrain et une extension de la banque de données de la douane.

Plus haut, nous avons déjà souligné que la douane, comme la plupart des autres services, n'enregistre pas de données concernant l'identité du conducteur contrôlé. Les champs correspondants du datawarehouse restent donc vides.

En annexe, nous donnons un aperçu des données du datawarehouse extraites du set douane et leur origine (cf. Appendice C), ainsi que le code Delphi-datamodule requis à cet effet (cf. Appendice D).<sup>10</sup> Veuillez remarquer que vu l'objectif du datawarehouse, on n'a retenu que les données concernant le trafic de personnes et de marchandises. Cette sélection a lieu sur la base des informations disponibles concernant les véhicules : on retient les catégories 'camionnette', 'camion', 'tracteur', 'semi-remorque' et 'autobus', pas les catégories voitures, mobile homes et autres.

L'implémentation ETL décrite ci-dessus, a été intégrée dans une interface graphique simple (l'application 'AgoraETLDouane'), qui permet d'en tester la fonctionnalité (cf. figure 9). Cette application Delphi établit le lien entre la banque de données source et la banque de données résultante (respectivement 'TestDataDouane.mdb' et 'AgoraDB.mdb', à mettre dans l'application directory). Pour permettre à l'utilisateur de réaliser un nouveau processus ETL, vers une banque de données résultante sans données de contrôle, le menu 'Maintenance' contient une série de sujets permettant d'effacer à nouveau le contenu de 'AgoraDB.mdb' ('Delete Agora Table': 'Controles', 'Vaststellingen', 'Tijd', 'Voertuigen') et de remplir à nouveau la table 'temps' ('Auto-fill tijd'). La logique ETL peut être testée soit sur l'enregistrement sélectionné dans la banque de données douane (entrée menu : ETL – Test), soit dans son ensemble (entrée menu : ETL – Execute). Au cours de ce dernier processus, une série d'avertissements est consignée en outre dans un fichier texte ('Log.txt'), qui peut par exemple être utilisé ultérieurement à des fins de validation (voir figure 8). Les numéros de référence du système source mis entre parenthèses permettent d'identifier l'enregistrement original (la composition de ce champ est décrite dans Appendice C).

```
Execution started at 12/27/2003 0:42:07.  
Warning (2-255182300-2003-02-136): replaced missing country code with  
Belgium for nummerplaat "SPF224"  
Warning (2-255180500-2003-02-822): replaced missing country code with  
Belgium for nummerplaat "QAZ014"  
Warning (2-255180500-2003-02-821): replaced missing country code with  
Belgium for nummerplaat "GME946"  
(...)
```

*Figure 8 : fragment fichier de journalisation*

L'exécution de l'ETL prend évidemment un certain temps et c'est pourquoi une fenêtre de progression est montrée. A titre indicatif, la banque de données sources contient 44657 enregistrements de contrôle, dont après sélection (voir plus haut), 14602 contrôles ont été retenus dans la banque résultante et liés à 56194 constatations individuelles (positives ou négatives). Le formulaire permet de surfer dans les données sources et résultats. Pour un mode de consultation convivial, nous vous renvoyons à la section suivante.

---

<sup>10</sup> Il nous semblait peu judicieux de reprendre dans l'annexe textuelle le code interface de l'utilisateur ou le code de l'application décrite dans la section suivante. L'annexe ne contient donc qu'une partie de l'implémentation.

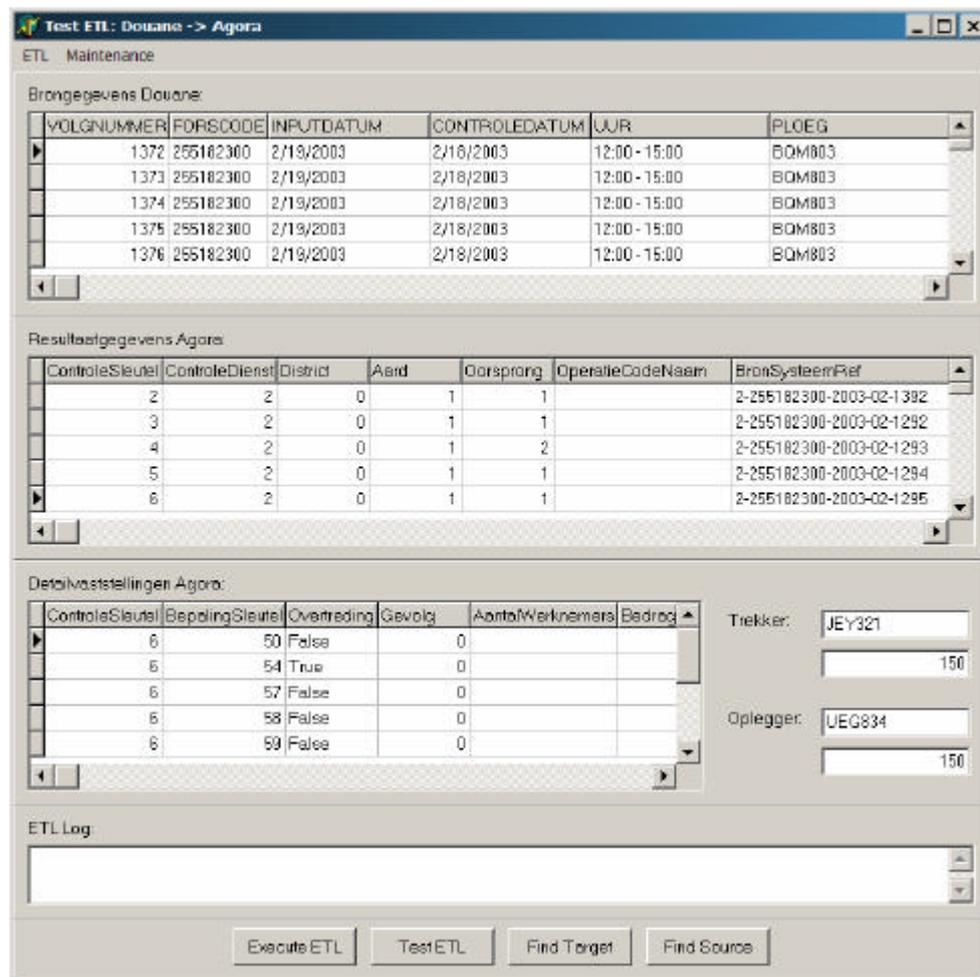


Figure 9: exemple écran interface utilisateur implémentation ETL

## 6.2. Consultation des données cibles – environnement démonstratif

Parallèlement au processus ETL décrit ci-haut, nous avons développé une interface graphique pour consulter et analyser les données de résultat. Cette dernière n’a pas la prétention d’être complète mais elle sert surtout à avoir un meilleur aperçu des possibilités et des limitations du datawarehouse et à valider les données de résultat.

L’écran principal (voir figure 10) est conçu comme un formulaire de recherche : sur la base de certains critères de recherche à préciser par l’utilisateur (par exemple, la date, transport de choses/personnes, la nationalité du véhicule, la plaque d’immatriculation<sup>11</sup>, le type d’infraction), une sélection est opérée dans les constatations de contrôle recueillies.<sup>12</sup> Un join-query

<sup>11</sup> Un critère incomplet de sélection est aussi possible au moyen de jokers, à savoir ‘%’ pour remplacer plusieurs signes ou ‘\_’ pour un signe.

<sup>12</sup> Nous nous sommes limités à la base des données disponibles. Par exemple, nous n’avons pas prévu de critère de recherche pour l’entreprise de transport concernée (par ex. numéro de TVA ou nom) pour la bonne et simple raison que ces données ne sont pas encore disponibles dans les données sources.

est réalisé en run-time dans les tables de faits et de dimensions du datawarehouse. Le résultat de la recherche est donné sous la forme d’un spreadsheet.

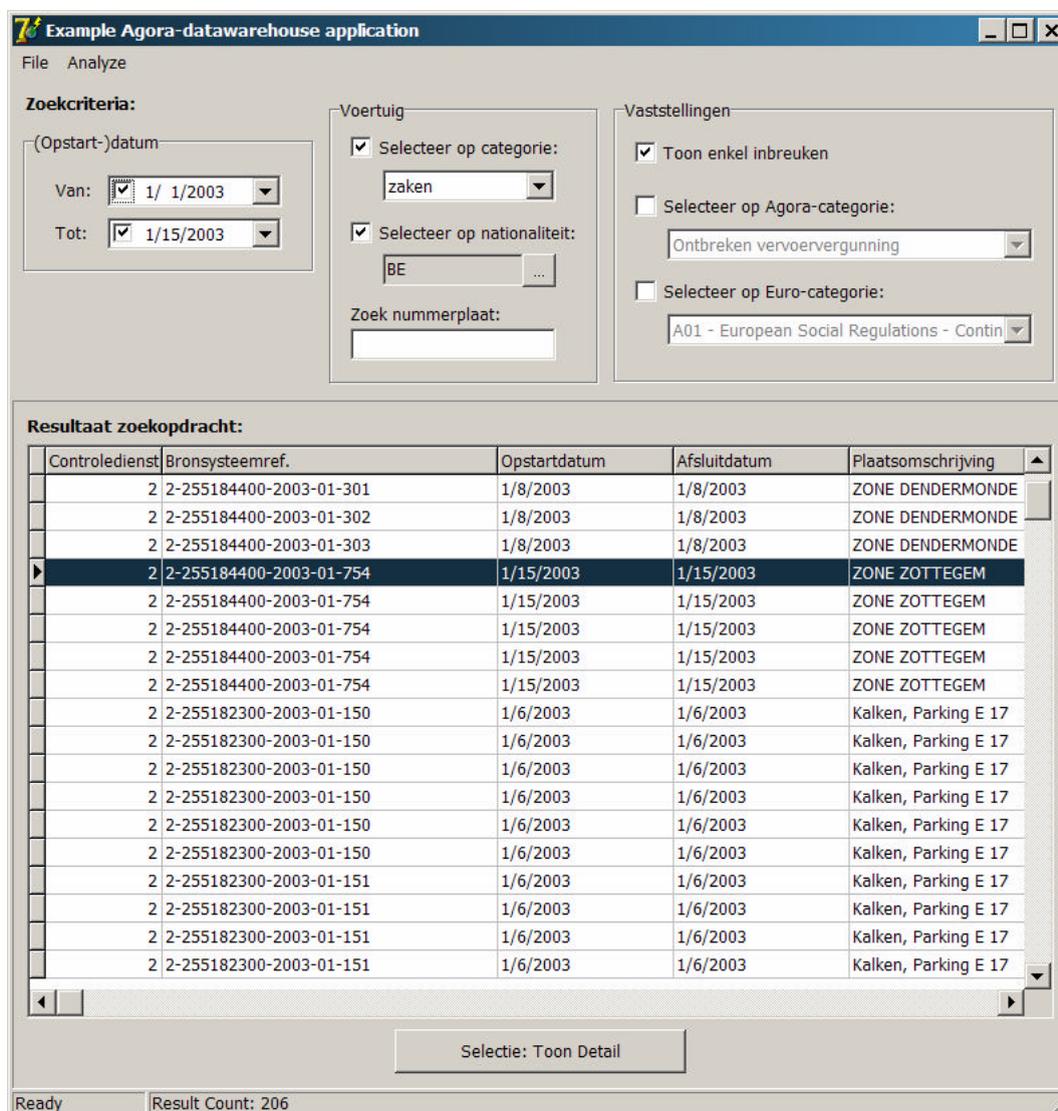


Figure 10 : exemple écran; écran de recherche

Plus de détails sur le contrôle sélectionné dans l’aperçu et sur les constatations y afférentes (voir figure 11) peuvent être demandés en cliquant sur le bouton situé en dessous de la fenêtre. Dans la fenêtre suivante, l’utilisateur peut choisir entre trois codages pour les constatations : le codage spécifique au service, la classification Agora (cf. exemple illustré) et le codage de l’arrangement administratif 'Euro Contrôle Route'.

Dans une implémentation définitive, un système de recherche comparable pourrait être offert via le web aux divers services, moyennant introduction des mécanismes de sécurisation et de filtrage requis.

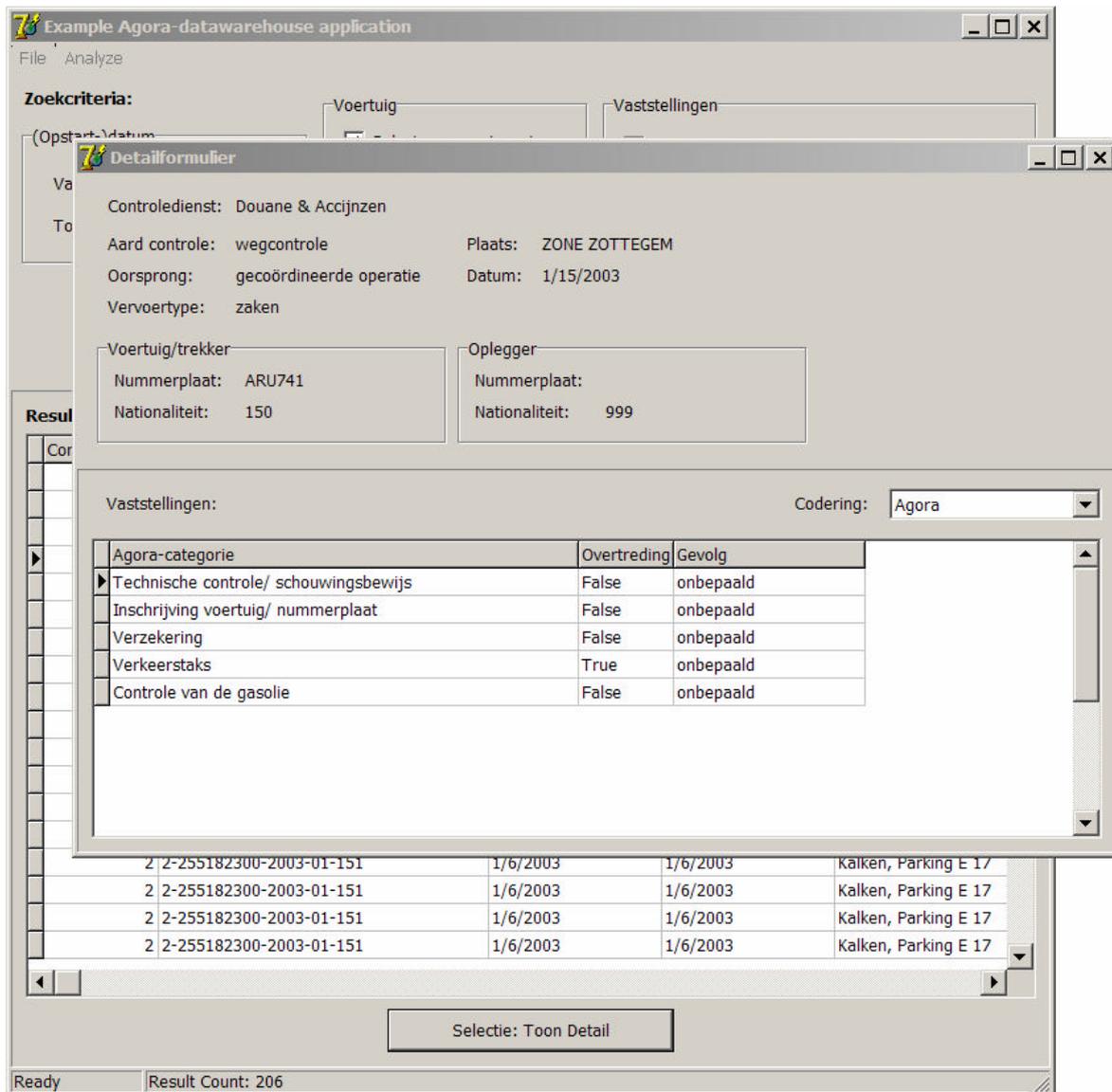


Figure 11: exemple écran; détail écran

Parallèlement à ces possibilités de recherche, l’environnement développé permet également d’établir un certain nombre de rapports d’exemple à partir des données recueillies. Ils peuvent être sélectionnés au moyen du menu situé en haut de la fenêtre principale. Un premier rapport d’exemple est la répartition des fréquences des nationalités de tous les véhicules repris dans le datawarehouse (pour autant que les nationalités soient disponibles) sous la forme d’un classement et d’un diagramme circulaire. Un second rapport d’exemple donne les fréquences des infractions constatées au cours du premier trimestre 2003 et enregistrées dans le datawarehouse selon les catégories Agora. La figure 12 montre une impression de ce rapport. Un troisième rapport d’exemple donne l’évolution, pour le premier trimestre 2003, du nombre de contrôles effectués versus le nombre de contrôles positifs (voir figure 13).

Si lors de l’implémentation définitive, on peut utiliser un outil de datawarehousing plus puissant, de tels rapports pourraient éventuellement être composés de manière relativement simple par l’utilisateur au moyen d’une interface point-and-click. Dans le présent exercice, nous

avons dû nous limiter à l'établissement manuel des queries nécessaires (qui ne peuvent pas être personnalisés davantage par l'utilisateur).

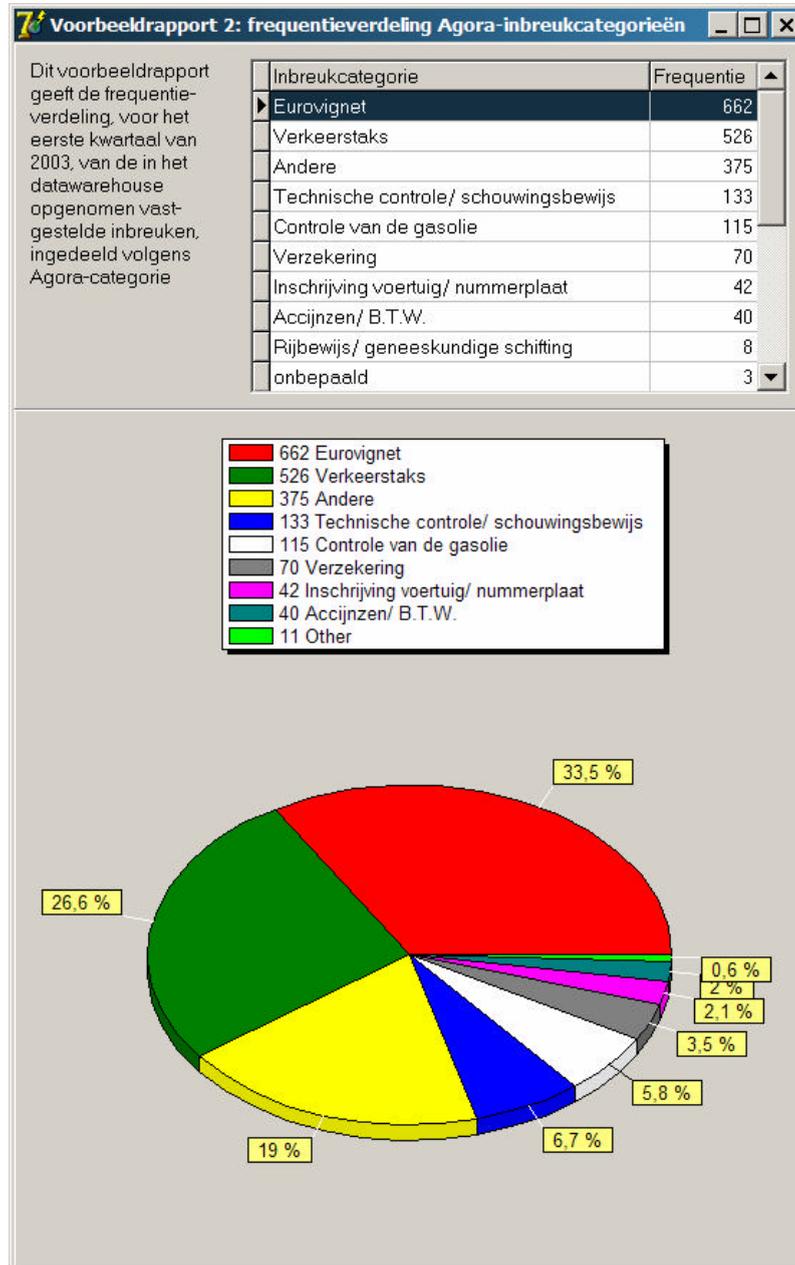


Figure12 : exemple écran; répartition fréquence catégories d'infractions Agora

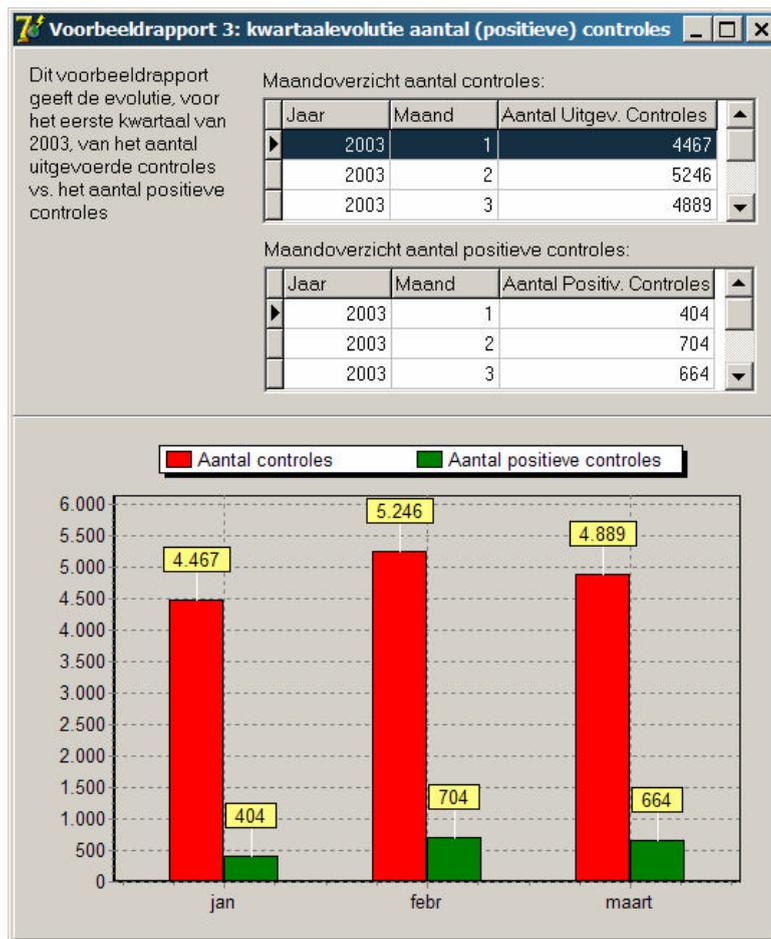


Figure13 : exemple écran; évolution trimestrielle nombre de contrôles (positifs)

Pour terminer, les données recueillies permettent des analyses pluridimensionnelles, au moyen de tableaux croisés montrant des (sous-) totaux et des moyennes pour un certain nombre de dimensions sélectionnables. Une telle présentation permet à l'utilisateur de reconnaître activement les tendances et les relations possibles. A cet effet, nous utilisons les classes de composantes de TDecisionCube qui font partie des composantes de support de décision livrées avec l'environnement Delphi. Les figures 14 et 15 montrent deux exemples d'écran. Dans la figure 14, le nombre de contrôles a été agrégé par mois (colonnes) et par catégorie de transport (lignes); la somme de chaque catégorie (pour les mois indiqués) apparaît dans la dernière colonne. En cliquant sur le signe +, ces totaux peuvent être scindés selon qu'une infraction a été constatée ou non. La table résultante est illustrée dans la figure 15. Cette application permet également de demander la nationalité. Ensuite, toutes les dimensions peuvent être réarrangées entre elles via drag-and-drop, tant au niveau de l'ordre de succession qu'au niveau des lignes et des colonnes, afin d'établir différents rapports possibles. Enfin, l'utilisateur peut aussi choisir une autre valeur, plus particulièrement le nombre de procès-verbaux (en cliquant à droite et en choisissant le sujet dans le menu qui apparaît alors). Ceci n'est donné qu'à titre d'illustration; dans un environnement de support de décision complètement développé, il devrait être possible de choisir à la carte entre tous les attributs des tables de dimensions disponibles (pour la répartition en lignes et colonnes) et les valeurs (pour les cellules des valeurs).

7 Meerdere dimensionale Voorbeeldanalyse: Controles/PV's uitgesplitst per dimensie					
		Datum			
Vervoercateg.		jan. 2003	feb. 2003	mrt. 2003	som
onbepaald		10	7	14	31
personen		8	18	12	38
zaken		4449	5221	4863	14533
som		4467	5246	4889	14602

Figure14 : exemple écran; nombre de contrôles par mois/catégorie de transport

7 Meerdere dimensionale Voorbeeldanalyse: Controles/PV's uitgesplitst per dimensie						
			Datum			
Vervoercateg.	Overtreding?		jan. 2003	feb. 2003	mrt. 2003	som
onbepaald	nee		8	5	9	22
	ja		2	2	5	9
	som		10	7	14	31
personen	nee		8	18	7	33
	ja				5	5
	som		8	18	12	38
zaken	nee		4047	4519	4209	12775
	ja		402	702	654	1758
	som		4449	5221	4863	14533
som			4467	5246	4889	14602

Figure15 : exemple écran ; nombre de contrôles par mois/catégorie de transport

## 7. Conclusion : situation actuelle et recommandations pour la suite

### 7.1. Sur le plan législatif et juridique

La loi relative à la protection de la vie privée est d'application sur les données à caractère personnel qui sont recueillies. Par conséquent, il faudra demander l'avis de la Commission de la protection de la vie privée pour mettre en oeuvre et développer la banque de données.

Il faudra examiner avec le parquet et les auditorats dans quelle mesure l'enregistrement de données relatives à des infractions s'accorde ou peut être accordé avec le secret de l'instruction en matière pénale.

Pour effectuer la liaison avec les banques de données opérationnelles gérées par l'Inspection des Lois sociales et l'Inspection sociale et plus tard, avec les sources additionnelles de données dans le cadre du réseau géré par la Banque-Carrefour des Entreprises (BCE) (comme le registre du personnel généré électroniquement par la déclaration DIMONA), il sera indiqué à l'avenir de noter dans la banque de données, le numéro d'identification unique de sécurité sociale du travailleur contrôlé. Il est possible que dans une première phase, on note aussi le

numéro O.N.S.S. de l'employeur. Une demande sera introduite pour ces deux données auprès du Comité de gestion de la BCE. Cette demande peut par exemple être élaborée en collaboration avec les partenaires de O.N.S.S.

Dans l'optique de la reprise du numéro d'entreprise (Banque-Carrefour Entreprises), des informations complémentaires doivent être demandées au gestionnaire du système à propos d'exigences éventuelles de procédure.

Pour rendre opérationnel le plan d'action du 20 novembre 2001, à savoir l'échange de données entre les partenaires susmentionnés, des arrêtés d'exécution complémentaires sont nécessaires. Il faut souligner qu'il n'existe pas de précédent en ce qui concerne l'échange de données autres que celles anonymes avec la police fédérale et locale.

## **7.2. Sur le plan du contenu et sur le plan technique**

Lors de l'inventaire des sources de données disponibles dans les divers services, nous avons constaté quelques manquements importants qui entravent le développement d'un système de datawarehouse pour les contrôles des transports routiers. En premier lieu, il n'existe pas actuellement encore de banque de données opérationnelle au Transport terrestre, ce qui est une condition impérative pour mettre en oeuvre et alimenter dans une phase ultérieure un datawarehouse. Pour réaliser rapidement cette phase préalable, un support additionnel serait indiqué ; ce support serait aussi utile lors du développement du datawarehouse.

Des problèmes ont aussi été constatés dans les systèmes des autres partenaires et compliquent la saisie automatique des données. A l'I.L.S., les données ne sont pas centralisées dans une seule grande banque de données. C'est pourquoi, nous avons proposé, de ne prendre en compte que les données de Malines et de Huy dans une première phase de datawarehousing, avant de développer un système d'approvisionnement plus étendu.

En outre, nous avons constaté que les systèmes sources disponibles ne peuvent fournir que très peu de données d'identification relatives au véhicule et au conducteur. Par conséquent, le datawarehouse ne pourra contenir, dans le contexte actuel que des données incomplètes. Il est donc recommandé que tous les partenaires s'efforcent, dans la mesure du possible et dans les limites de leurs compétences spécifiques, d'ajouter systématiquement certaines de ces données additionnelles d'identification dans leurs versions de système futures.

A la police locale et fédérale, les nouveaux systèmes qui peuvent servir à alimenter le datawarehouse prévu sont toujours partiellement en cours de développement. Une concertation plus étroite s'impose, notamment pour faire connaître nos besoins et pour avoir un meilleur aperçu du cadre légal.

## **7.3. Sur le plan organisationnel et budgétaire**

Un projet aussi important de datawarehouse n'a de chances de réussir que si les moyens nécessaires y sont affectés. Il ressort de la présente étude que l'implémentation des routines d'extraction des données ainsi que l'organisation du processus d'alimentation sera une tâche conséquente, vu le caractère hétérogène des systèmes sources disponibles. En outre, un tel datawarehouse doit, même après le développement d'une première version, continuer à évo-

luer en fonction des besoins et des sources d'alimentation disponibles ; il faut disposer d'un staff continu pour l'alimentation périodique, la correction des fautes, l'analyse, le rapportage et l'archivage des vieilles données.

Nous voulons insister sur le fait que des efforts ne sont pas uniquement requis du service qui a pris l'initiative mais aussi des autres partenaires, pour pouvoir mettre en oeuvre et (surtout) entretenir un tel système. Des accords clairs doivent être pris concernant l'organisation de la fourniture des données, le renvoi et le traitement des données si des fautes sont constatées lors de la saisie, la communication des nouveaux codes spécifiques aux services pour les infractions, etc. Le développement plus avant du datawarehouse devrait en plus constituer un facteur important d'input pour la prise de décisions concernant la propre infrastructure de banque de données, vu que l'offre actuelle en données livrables est encore très limitée.

#### **7.4. Aperçu des étapes suivantes éventuelles**

L'inventaire des données disponibles, un projet de modèle pour la structure du datawarehouse et un 'proof-of-concept' ont été faits. Il faudrait maintenant approfondir le volet législatif et juridique ainsi que le volet développement. En ce qui concerne ce second volet, il faudrait d'abord développer et mettre en oeuvre une banque de données opérationnelle au sein du Transport terrestre. Le développement du datawarehouse proprement dit se fait typiquement selon un processus itératif ; on commencerait avec le développement complet du trajet pour un des partenaires, sur la base duquel le projet initial peut éventuellement être adapté. Dans une première phase, nous songeons surtout à l'Inspection des Lois sociales (avec laquelle il existe déjà une collaboration étroite) et/ou à la douane pour laquelle l'essai d'implémentation ci-haut a été développé. L'inspection sociale et la police fédérale peuvent suivre dans une deuxième phase. Les données de la police locale et des services de contrôle de l'O.N.S.S. sont provisoirement moins pertinentes pour le contrôle des transports routiers. Chacun des sous-trajets suppose le développement des étapes suivantes : (1) extraction des données; (2) transformation des données; (3) alimentation du datawarehouse; (4) post traitement. En outre, il faut prévoir les facilités de consultation et de rapportage nécessaires.

Pour terminer, nous faisons remarquer que, mis à part le présent projet de datawarehousing pour les contrôles et les infractions, il existe d'autres opportunités pour un meilleur échange des informations, à savoir l'amélioration de l'accès aux banques de données existantes, par exemple celles des plaques d'immatriculation et des autorisations d'une part et de la DIMONA d'autre part.

## **Bibliographie**

Inmon, W. H. (1996), Building the Data Warehouse, second edition, Wiley, N.Y., 401 pp.

Vandenbulcke, J.A. en Lemahieu, W., Databasesystemen voor de praktijk, zevende druk, Ten Hagen en Stam, Den Haag, 560 pp.

## **Abréviations et acronymes**

S.C.G.I.– Administration de la Douane et Accises, Service Central de Gestion de l'Information (C.D.I.B. – Administratie der Douane & Accijnzen, Centrale Dienst voor Informatiebeheer)

DIMONA - La Déclaration IMMédiate de l'emploi (De ONmiddellijke Aangifte van tewerkstelling)

D.I.V. - Direction pour l'Immatriculation des Véhicules (Dienst voor Inschrijving van de Voertuigen)

I.L.S. – Inspection des Lois Sociales (I.S.W. – Inspectie Sociale Wetten)

O.N.S.S – Office National de Sécurité Sociale (R.S.Z. – Rijksdienst voor Sociale Zekerheid)

I.S. – Inspection Sociale (S.I. – Sociale Inspectie)

## **Annexes**

Les annexes peuvent être obtenues auprès de l'équipe de recherche à l'adresse suivante:

KU Leuven

Département Sciences économiques appliquées

Naamsestraat 69

3000 Leuven

e-mail: christphe.mues@econ.kuleuven.ac.be ou jan.vanthienen@econ.kuleuven.ac.be

TABLEAU/ATTRIBUT	TYPE	DESCRIPTION
<b>Contrôles</b>		tableau central des faits comprenant les contrôles routiers et enquêtes en entreprise effectués par divers services.
ContrôleClé	Nombre attribué automatiquement	Numéro contrôle – identification, attribué automatiquement
ServiceContrôle	Nombre	Service qui a effectué le contrôle; voir tableau ContrôleServiceCodes (0: non défini; 1: Transport terrestre; 2: Douane; 3: ILS.; 4: I.S.; 5: ONSS.; 6: Police fédérale; 7: Police locale; 8: Euro Contrôle Route)
District	Nombre	District au sein du service de contrôle (0 = si non d'application); voir tableau DistrictCodes (p.ex. 1: ILS-Mechelen)
Genre	Nombre	Genre du contrôle; voir tableau ContrôleGenreCodes (1: contrôle routier; 2: inspection en entreprise)
Origine	Nombre	Origine du contrôle; voir tableau OrigineCodes (0: non défini; 1: contrôle de routine; 2: contrôle ciblé; 3: opération coordonnée; 4: demande d'une instance externe; 5: plainte; 6: propre enquête)
OpérationCodeNom	Texte	Nom-code e l'opération au sein de laquelle le contrôle a eu lieu (le champ reste vierge si non disponible ou non applicable).
SourceSystèmeRéf	Texte	Serie-caractère qui permet d'identifier le record-contrôle original dans le système-source opérationnel.
DateEnregistrementClé	Nombre	Référence nombre clé relative à la date d'introduction du record-contrôle dans le datawarehouse; voir tableau DimensionTemps
DateDébutClé	Nombre	Référence nombre relative à la date du début du contrôle routier / de l'inspection en entreprise; voir tableau DimensionTemps
DateClôtureClé	Nombre	Référence nombre clé relative à la date de clôture du contrôle routier / de l'inspection en entreprise (à savoir identique à la date de début dans le cas d'un contrôle routier); voir tableau DimensionTemps
LieuDescription	Texte	Endroit où le contrôle a eu lieu (p.ex. "Gent-Brugge E17"; uniquement d'application dans le cas d'un contrôle routier; non standardisé; vierge si non disponible ou non d'application).
CléVéhicule	Nombre	Référence nombre clé relative au véhicule contrôlé ou, pour un véhicule articulé, au tracteur (uniquement d'application lors d'un contrôle routier; 0 lorsque non d'application ou non disponible); voir tableau DimensionVéhicules
SemiremorqueClé	Nombre	Référence nombre clé relative à la semi-remorque (uniquement d'application en cas de contrôle routier d'un véhicule articulé; 0 lorsque non d'application ou non disponible); voir tableau DimensionVéhicules
ChauffeurClé	Nombre	Référence nombre clé relative au conducteur, si celui-ci est identifiable de façon unique sur base de son NISS (uniquement d'application lors d'un contrôle routier, 0 lorsque non d'application ou non disponible); voir tableau DimensionConducteurs
NomChauffeur	Texte	Nom et prénom du conducteur (le champ reste vierge lorsque non disponible / non d'application; pas de contrôle sur l'identification unique)
AdresseChauffeur	Texte	Adresse du conducteur (le champ reste vide si non disponible ou non applicable)
EmployeurClé	Nombre	Référence nombre clé à l'entreprise de transport (belge) contrôlée; voir tableau DimensionEntreprises; 0 si non identifiable dans ce dernier.

TABLEAU/ATTRIBUT	TYPE	DESCRIPTION
NomEmployeur	Texte	Nom de l'entreprise contrôlée; permet p.ex. le rapportage concernant les contrôles routiers d'entreprises étrangères (qui ne sont pas reprises dans le tableau DimensionEntreprises); vide si non disponible
EmployeurNationalité	Nombre	Code nombre pour la nationalité de l'entreprise contrôlée; voir tableau Codes Pays (p.ex. 150 = Belgique; 999 = non défini)
Belge	Oui/non	Variable oui/non: oui: entreprise belge; non: non belge; vide: lorsque l'information n'est pas disponible
EuroContrôleRoutePays	Oui/non	Variable oui/non: oui: entreprise d'un pays Euro Contrôle Route; non: autre; vide: lorsque l'information n'est pas disponible
UE	Oui/non	Variable oui/non: oui: entreprise UE; non: non UE ; vide: si l'information n'est pas disponible
Infraction	Oui/non	Variable oui/non qui indique si une infraction a été constatée lors du contrôle; vide lorsque l'information n'est pas disponible
NombreVéhiculesContrôlés	Nombre	Valeur de mesure: nombre de véhicules contrôlés; toujours 1 lors d'un contrôle routier; 0, 1 ou supérieur pour une enquête en entreprise; vide lorsque l'information n'est pas disponible
NombreEmployésContrôlés	Nombre	Valeur de mesure: nombre d'employés contrôlés; vide lorsque l'information n'est pas disponible
NombrePVx	Nombre	Valeur de mesure: nombre de PV'x établis dans le cadre du contrôle; vide lorsque l'information n'est pas disponible
PontantPerçu	Nombre	Valeur de mesure: montant (total) perçu en Euro; 0 lorsque il n'y a pas de perception immédiate; vide lorsque l'information n'est pas disponible
MontantRégularisation	Nombre	Valeur de mesure: montant (total) régularisé en Euro; 0 s'il n'y a pas de régularisation; vide lorsque l'information n'est pas disponible
<b>Temps</b>		tableau dimension contenant l'information de temps relative aux contrôles exécutés
TempsClé	Nombre (attribué automatiquement)	Numéro d'identification du record attribué automatiquement
Date	Date	Date (p.ex., 31/5/2003)
Mois	Nombre	Mois – sous forme de chiffre (p.ex 5)
Trimestre	Nombre	Trimestre (p.ex. 2)
Année	Nombre	Année (p.ex. 2003)
<b>Véhicules</b>		tableau dimension contenant des informations détaillées relatives aux véhicules contrôlés
VéhiculeClé	Nombre (attribué automatiquement)	Numéro d'identification du record véhicule à attribuer automatiquement (0 est réservé pour les renvois zéro dans le tableau des FaitsContrôles)
CatégorieVéhicule	Nombre	Catégorie du véhicule: choses / personnes; voir tableau CodesCatégorieVéhicule (0 = non défini; 1 = choses; 2= personnes)
PlaquelImmatriculation	Texte	Numéro de plaque d'immatriculation (max. 10 caractères)
Nationalité	Nombre	Code nombre Nationalité; voir tableau Codes Pays (p.ex. 150 = Belgique; 999 = non défini)
<b>Conducteurs</b>		tableau dimension contenant des informations détaillées relatives aux conducteurs contrôlés identifiés sur base de leur NISS
ConducteurClé	Nombre (attribué automatiquement)	Numéro d'identification du record conducteur à attribuer automatiquement (0 est réservé aux références zéro dans le tableau des FaitsContrôles)
NISS	Texte	Numéro d'identification de la sécurité sociale (NISS); vide lorsque non d'application ou non disponible

TABLEAU/ATTRIBUT	TYPE	DESCRIPTION
Nom	Texte	Nom et prénom du conducteur
DateDébut	Date	Date de début de la période de validité des données conducteurs ci-dessous
DateFin	Date	Date de fin de la période de validité des données conducteurs ci-dessous; 12/31/2999 tant que non modifié
Nationalité	Nombre	Code nombre pour la nationalité du conducteur; voir tableau CodesPays (p.ex. 150 = Belgique; 999 = non défini)
Statut	Nombre	Statut social du conducteur; voir tableau CodesStatutConducteur (0 = non défini; 1 = employé; 2 = indépendant; 3 = intérimaire)
<b>Entreprises</b> tableau dimension contenant des informations détaillées relatives aux entreprises de transport belges		
EntrepriseClé	Nombre (attribué automatiquement)	Numéro d'identification du record de l'entreprise de transport (belge) à attribuer automatiquement (0 est réservé aux références zéro dans le tableau des FaitsContrôle)
Nom	Texte	Nom officiel de l'entreprise
Numéro_ONSS	Texte	Numéro ONSS
Numéro_BCE	Texte	Numéro BCE (Banque Carrefour des Entreprises); remplace le numéro de TVA
<b>Constations</b> tableau des faits contenant les constatations faites lors des contrôles		
ContrôleClé	Nombre	Référence nombre clé au numéro d'identification du contrôle du tableau des faits attribué automatiquement
DéfinitionClé	Nombre	Référence nombre clé aux dispositions respectives (propres au service) dont l'application a été contrôlée lors du contrôle; voir tableau Dispositions
Infraction	Oui/non	Variable Oui/non qui indique si lors du contrôle il y a eu oui ou non constat d'une infraction à la disposition précitée (remarque: en cas de non-infraction il n'y aura pas toujours un record de disponible à ce sujet dans la source opérationnelle)
Suite	Nombre	Façon dont il a été donné suite à la constatation; voir tableau CodesSuite (0 = non défini; 1 = aucune; 2 = avertissement; 3 = régularisation; 4 = perception immédiate; 5 = pro-justitia; 6 = transmission; 7 = autres)
NombreEmployés	Nombre	Nombre d'employés concernés; vide lorsque non disponible
Montant	Nombre	Montant perçu/régularisé (en fonction de la valeur "suite") en Euro pour la constatation en question; 0 si pas de perception immédiate ou de régularisation; vide lorsque l'information n'est pas disponible
<b>Définitions</b> tableau reprenant les codifications utilisées par les services en ce qui concerne les dispositions légales contrôlables		
Déf_Clé	Nombre attribué automatiquement	Numéro d'identification du record pour la disposition légale (propre au service) attribué automatiquement
Code	Texte	Codification de la disposition propre au service (p.ex. "D20811")
Description	Texte	Libellé de la disposition (p.ex. "temps de repos insuffisant – moins de 11 heures – art. 8.1., 1 <sup>er</sup> alinéa – 62 Euro par période de 30 minutes)
Service	Nombre	Service de contrôle qui applique la disposition / le code; voir tableau CodesContrôleService (1 = Transport terrestre; 2 = Douane et Accises; 3 = I.L.S.; 4 = I.S.; 5 = O.N.S.S.; 6 = Police fédérale; 7 = Police locale; 8 = Euro Contrôle Route)
DateDébut	Date	Date d'introduction (valeur par défaut: 1/1/1800, si déjà d'application avant entrée en vigueur du datawarehouse)

TABLEAU/ATTRIBUT	TYPE	DESCRIPTION
DateFin	Date	Date de radiation lorsque le code n'est plus utilisé; 12/31/2999 tant qu'utilisé
CatégorieAgora	Nombre	Référence nombre clé à des catégories de contrôles communes encore à convenir par les partenaires; voir tableau Déf_CatégoriesAgora
CatégorieEuro	Nombre	Référence nombre clé au code Euro Contrôle Route correspondant; voir tableau Déf_CatégoriesEuro (0 si la disposition ne fait pas partie de ces catégories)
<b>Déf_CatégoriesAgora</b>		tableau avec la catégorisation commune des dispositions à définir par les services
Cat_Clé	Nombre (attribué automatiquement)	Numéro code d'identification à attribuer automatiquement (à convenir en commun) pour la catégorie d'ensemble des dispositions
Cat_Description	Texte	Description (p.ex. Tachygraphe: temps de conduite et de repos & interruptions)
DateDébut	Date	Date d'introduction
DateFin	Date	Date de suppression lorsque le code n'est plus utilisé; 12/31/2999 tant qu'utilisé
Cat_Explication	Texte	Explication optionnelle
<b>Déf_CatégoriesEuro</b>		tableau reprenant la catégorisation des dispositions utilisées dans le cadre d'Euro Contrôle Route
Cat_Clé	Nombre attribué automatiquement	Numéro de code d'identification à attribuer automatiquement pour les catégories de dispositions utilisées dans le cadre de l'Euro Contrôle Route
Cat_Code	Texte	Codification de la catégorie (p.ex. "A04")
DateDébut	Date	Date d'introduction
DateFin	Date	Date de suppression lorsque le code n'est plus utilisé; 12/31/2999 tant qu'utilisé
Cat_Entête	Texte	Entête sous laquelle la catégorie est reprise (p.ex. European Social Regulations)
Cat_DetailTexte	Texte	Description de la catégorie (en détail) ("Daily rest")
<b>ConducteurStatutCodes</b> <b>ContrôleGenreCodes</b> <b>ContrôleServiceCodes</b> <b>DistricCodes</b> <b>SuiteCodes, PaysCodes</b> <b>OrigineCodes</b> <b>VéhiculeCatégorieCodes</b>		tableaux de codes du datawarehouse

ATTRIBUT	TYPE	DESCRIPTION
Contrôles (routiers) brigades de douane		
NUMERO D'ORDRE	Nombre	Numéro d'ordre contrôle dans le mois (par code fors)
CODE FORS	Texte	Code Fors de la brigade
DATE INPUT	Date	Date à laquelle le rapport de contrôle a été entré
DATE CONTROLE	Date	Date à laquelle le contrôle a été effectué
HEURE	Texte	Indication heure contrôle
EQUIPE	Texte	Plaque d'immatriculation du véhicule de service ou nom d'appel du team fixe
POINT DE CONTROLE	Texte	Description lieu du contrôle (ex, "Gand-Bruges E17")
IMPORTATEUR	Texte	Numéro TVA importateur/destinataire*
EXPORTATEUR	Texte	Numéro TVA exportateur/expéditeur*
TRANSPORTEUR	Texte	Numéro TVA transporteur*
VEHICULE	Texte	Information structurée concernant véhicule contrôlé; type1/plaque d'immatriculation1/nationalité1;type2/plaque d'immatriculation2/nationalité2 (types : 1: voiture; 2: camionnette; 3: camion; 4: autobus; 5: mobile home; 6: tracteur; 7: semi-remorque; 8: autres)
NATIONALITE	Texte	Description nationalité*
TYPE DE CONTROLE	Nombre	Type de contrôle: 1: dynamique (interception du véhicule); 2: statique 1 (le long de la route; sélection); 3: statique 2 (le long de la route; tous)
NATURE DU CONTROLE	Nombre	Nature contrôle: 1: routine; 2: ciblé (sur une disposition spécifique, cf. infra); 3: opération (evt. coordonnée)
NOM DE CODE	Texte	Detail nature contrôle: détermination contrôle ciblé ou nom de code opération (ex, 210 ou "ZW VERV LOKPOL ZOTTEGEM")
SORTE DE CONTROLE	Texte	Liste codes dispositions contrôlées (chaque sous-code est précédé d'un point -virgule ex, ";210;320;340"); voir ailleurs pour aperçu des codes utilisés
CONSTATATION	Texte	liste codes dispositions pour lesquelles une infraction a été constatée (doit être une sous-liste de la liste ci-dessus)
A115	Texte	Info détaillée disposition 115 (transport produits d'accises - autres: ...)
A122	Texte	Info détaillée disposition 122 (droits d'importation - produits visés: ...)
A132	Texte	Info détaillée disposition 132 (contrôle TVA - produits visés: ...)
A145	Texte	Info détaillée disposition 145 (missions non fiscales - autres: ...)
A380	Texte	Info détaillée disposition 380 (contrôle moyen de transport - autres: ...)
NUMERO DE DOSSIER	Texte	(numéro de référence pour usage interne)
BATCH	Oui/non	(usage interne)

\* pas disponible ou seulement de manière non systématique dans le set de test

TABLEAU/ATTRIBUT	TYPE	ATTRIBUTION/ORIGINE
<b>Contrôles</b>		
ContrôleClé	Nombre	Valeur clé attribuée automatiquement: numéro précédent + 1 (procédure implémentation: voir CléTGénérateur-classe); uniquement les contrôles qui concernent le transport de personnes et de marchandises sont pris en compte. La sélection se fait sur base du type de véhicule (voir tableau "Véhicules").
ServiceContrôle	Nombre	2 (code pour le service de contrôle Douanes et Accises)
District	Nombre	0 (pas d'application)
Genre	Nombre	1 (contrôle routier)
Origine	Nombre	Champ-source: GENRECONTROLE; 1-> 1 (contrôle de routine); 2 -> 2 (contrôle ciblé); 3 -> 3 (opération coordonnée)
CodeNomOpération	Texte	Champ-source: CODENOM
SourceSystèmeRéf	Texte	Champs -source: CODE FORS; DATE DU CONTROLE; NUMERO D'ORDRE Exemple: 2-255182300-2003-02-1391, avec - 255182300: code fors brigade motorisée - 2003: année - 02: février - 1391: numéro d'ordre dans le mois pour la brigade
DateEncodageClé	Nombre	Valeur clé (cf. tableau "temps") date exécution processus ETL
DateDémarrageClé	Nombre	Champ-source: DATE CONTROLE; transposée vers la valeur clé correspondante dans le tableau "temps".
DateClôtureClé	Nombre	Identique à DateCléDémarrage (d'ailleurs uniquement contrôles routiers)
DescriptionLieu	Texte	Champ-source: LIEU CONTRÔLE
VéhiculeClé	Nombre	Champ-source: VEHICULE; procédure cleansing/parsing – implémentation,: ParseVéhiculeField; procédure-implémentation recherche record existant dans le tableau "Véhicules (cf. infra) ou réalisation d'un nouveau record en cas de non-identification: UpdateTableauVéhicules (clé d'identification secondaire: Plaque d'immatriculation, Nationalité). Le résultat est une référence nombre clé au véhicule contrôlé ou, pour un véhicule articulé, au tracteur (0 lorsque non disponible).
VéhiculeSemiremorqueClé	Nombre	Champ-source: VEHICULE; transformation similaire à ci-dessus; référence nombre clé à la semi-remorque si VEHICULE contient une deuxième donnée à ce sujet; si non 0 (notez que ces semi-remorques forment un record distinct dans le tableau "Véhicules").
ConducteurClé	Nombre	0 (non disponible dans le système source)
NomConducteur	Texte	Vide (id.)
AdresseConducteur	Texte	Vide (id.)
EmployeurClé	Nombre	0 (non disponible systématiquement dans le système source)
NomEmployeur	Texte	Vide (id.)
NationalitéEmployeur	Nombre	999 (id.)
Belge	Oui/Non	Non implémenté
PaysEuroControleRoute	Oui/Non	Non implémenté
UE	Oui/Non	Non implémenté
Infraction	Oui/Non	Valeur 'Oui' si, après traitement des champs-source GENRE CONTRÔLE et CONSTATATION (cf. tableau "Constata-tions") une infraction a été enregistrée dans au moins une des constata-tions relatées au contrôle; dans le cas contraire: 'Non'

TABLEAU/ATTRIBUT	TYPE	ATTRIBUTION/ORIGINE
<b>Contrôles</b>		
NombreVéhiculesContrôlés	Nombre	1 (contrôle routier)
NombreEmployeursContrôlés	Nombre	Vide (non disponible dans le système source)
NombrePVx	Nombre	Nombre de PV'x établis dans le cadre du contrôle; complété après traitement des constatations (cf. champs-source GENRE CONTROLE et CONSTATATION); dans le cadre de cette implémentation-test nous avons considéré que si 133 (PV 108), 134 (PV 109) ou 410 (PV Douane et Accises) apparaissent comme codes partiels, un PV doit être établi; une validation complémentaire est souhaitée.
MontantPerçu	Nombre	Vide (non disponible dans le système source)
MontantRégularisé	Nombre	Vide

<b>Temps</b>		
TempsClé	Nombre	Numéro d'identification-record attribué automatiquement; dans notre implémentation-test: le nombre de jours depuis le 1er janvier 1995; une procédure distincte (FillTableauTemps) a été prévue pour compléter ce tableau.
Date	Date	Date (p.ex., 31/5/2003)
Mois	Nombre	Mois sous forme de chiffre (p.ex., 5)
Trimestre	Nombre	Trimestre (p.ex., 2)
Année	Nombre	Année (p.ex., 2003)

<b>Véhicules</b>		
VéhiculeClé	Nombre	Numéro d'identification-record véhicule attribué automatiquement: numéro précédent + 1 (procédure-implémentation: voir TcléGénérateur-classe); 0 est réservé pour les renvois zéro.
CatégorieVéhicule	Nombre	Champ-source: VEHICULE; procédure partielle cleansing/parsing: ParseCatégorieStr - remplacement des synonymes code - transposition: 2 (camionnette.), 3 (camion.), 6 (tracteur.), 7 (sémi-remorque.) -> 1 (choses); 4 (autobus) -> 2 (personnes); 1 (voiture.), 5 (mobilhome) ou autres: non retenus
PlaqueImmatriculation	Texte	Champ-source: VEHICULE; procédure partielle cleansing/parsing: ParsePlaqueStr - transposition en majuscules - enlever les signes contrôle (espaces, traits d'union, ...) - contrôle selon longueur min. et max.
Nationalité	Nombre	Champ-source: VEHICULE; procédure partielle cleansing/parsing: ParseNationStr; décrite dans la section 6.1. du texte du rapport; fait usage de la banque de données "NationMapping.mdb".

<b>Conducteurs</b>		
Non complété (non disponible dans le système source)		

<b>Entreprises</b>		
Non complété		

TABLEAU/ATTRIBUT	TYPE	ATTRIBUTION/ORIGINE
<b>Constatation</b>		
ContrôleClé	Nombre	Référence au numéro d'identification du contrôle attribué automatiquement au départ du tableau Contrôles -faits Constatés.
DéfinitionClé	Nombre	Champ-source: GENRE CONTROLE; la liste séparée ";"-est d'abord subdivisée (cf. procédure ParseListString); pour chaque subcode qui en résulte, un record est réalisé, la valeur-clé correspondante est recherchée dans le tableau "Définitions", ensuite il y est fait référence; procédure-implémentation: UpdateTableauConstatations.
Infraction	Oui/Non	Champs -source: GENRE CONTROLE / CONSTATATION; valeur 'Oui' en cas de présence de la définition dans les deux listes de codes; dans le cas contraire 'Non'; le traitement de la deuxième liste est similaire à ci-dessus.
Suite	Nombre	0 (indéfini): non disponible à ce niveau de détail
NombreEmployés	Nombre	Vide
Montant	Nombre	Vide

<b>Définitions</b>
Contenu réalisé préalablement à l'exécution du processus ETL; décrit dans la section 6.1. du texte du rapport.

<b>Def_CatégoriesAgora</b>
Contenu réalisé préalablement à l'exécution du processus ETL; décrit dans la section 5 du texte du rapport.

<b>Def_CatégoriesEuro</b>
Contenu réalisé préalablement à l'exécution du processus ETL

<b>ConducteurStatutCodes, ContrôleGenreCodes, ServiceContrôleCodes, DistrictCodes, SuiteCodes, PaysCodes, OrigineCodes, VéhiculeCatégorieCodes</b>
Contenu réalisé préalablement à l'exécution du processus ETL.

---

**Appendix D: ETL-datamodulecode proof-of-concept douane**

---

```

unit ETLDataMod;

interface

uses
  Windows, Messages, SysUtils, Classes, Graphics, Controls, Forms, Dialogs,
  Db, ADOdb, ProgressDlgs, DateUtils, Variants, StrUtils;

type
  TControleDienst = (cdVervoerTeland, cdDouane, cdISW, cdSI, cdRSZ, cdFedPol,
    cdLokPol, cdEurControlRout);
  TControleAard = (caWegControle, caBedrijfsOnderzoek);
  TOorsprong = (osRoutineControle, osGerichteControle, osGecoordOperatie,
    osAanvraagExtInst, osKlacht, osEigenOnderzoek);
  TGevolg = (gGeen, gWaarschuwing, gRegularisatie, gOnmidInning, gProJustitia,
    gDoorverzending, gAnder);
  TBestuurderStatuut = (bsWerknemer, bsZelfstandige, bsInterim);

  TVoertuigTypeDouane = (vtPersonenWagen, vtLichteVrachtwagen, vtVrachtwagen,
    vtAutobus, vtMobilHome, vtTrekker, vtOplegger, vtAndere);
  TVoertuigCategorie = (vcZaken, vcPersonen, vcAndere);

  TVoertuigData = class
public
  DouaneCategorie: TVoertuigTypeDouane;
  AgoraCategorie: TVoertuigCategorie;
  NummerPlaat: string;
  NationCodeSleutel: Integer;
  Oplegger: TVoertuigData;
constructor Create;
destructor Destroy; override;
end; //TVoertuigData

  TSleutelGenerator = class
private
  ADOTable: TADOTable;
  SleutelName: string;
  NewSleutelValue: Integer;
public
constructor Create(AnADOTable: TADOTable; ASleutelName: string);
procedure Reset;
function GetNewSleutelValue: Integer;
end;

const
  OnbepaaldCode = 0;
  OnbepaaldPlaatStr = '';
  OnbepaaldLandCode = 999;
  BelgiumLandCode = 150;
  ControleDienstCodes: array[TControleDienst] of Integer = (1, 2, 3, 4, 5, 6, 7, 8);
  ControleAardCodes: array[TControleAard] of Integer = (1, 2);
  OorsprongCodes: array[TOorsprong] of Integer = (1, 2, 3, 4, 5, 6);
  GevolgCode: array[TGevolg] of Integer = (1, 2, 3, 4, 5, 6, 7);
  BestuurderStatuutCodes: array[TBestuurderStatuut] of Integer = (1, 2, 3);
  VoertuigCategorieCodes: array[TVoertuigCategorie] of Integer = (1, 2, 0);
  FromYear = 1995;
  ToYear = 2010;
  MaxNummerPlaatLength = 10;

type
  TETLDataModule = class(TDataModule)
  DouaneConnection: TADOConnection;
  AgoraConnection: TADOConnection;

```

```

    DouaneTable: TADOTable;
    AgoraControles: TADOTable;
    AgoraVaststellingen: TADOTable;
    AgoraTijd: TADOTable;
    AgoraVoertuigen: TADOTable;
    NationMapConnection: TADOCConnection;
    DouaneLandenMap: TADOTable;
    LandCodes: TADOTable;
    AgoraBepalingen: TADOTable;
    procedure DataModuleCreate(Sender: TObject);
    procedure DataModuleDestroy(Sender: TObject);
private
    { Private declarations }
    FromDate: TDateTime;
    WritingToLog: Boolean;
    LogFile: TextFile;
    CurrentDouaneBronSysteemRef: string; //used for logging
    function GetDouaneBronSysteemRef: string;
    function GetTijdSleutelValue(ADate: TDateTime): Integer;
    function GetMaxSleutelValue(ADOTable: TADOTable; SleutelName: string): Integer;
    function UpdateVoertuigenTabel(VoertuigData: TVoertuigData;
        var SleutelValue: Integer): Boolean;
    function ParseVoertuigField(VoertuigStr: string): TVoertuigData;
    procedure UpdateVaststellingenTabel(ControleSleutelValue: Integer;
        ControlSoortStr, VaststellingStr: string; var Overtreding: Boolean;
        var PVCCount: Integer);
    procedure ExtractCurrentControleRecord;
public
    { Public declarations }
    procedure FillTijdTabel;
    procedure DeleteADOTable(ADOTable: TADOTable);
    procedure TestETL;
    procedure ExecuteETL;
    function LocateAgoraTargetRecord: Boolean;
    function LocateDouaneSourceRecord: Boolean;
end;

var
    ETLDataModule: TETLDataModule;
    VoertuigSleutelGenerator, ControleSleutelGenerator: TSleutelGenerator;

{=====}

implementation

{$R *.DFM}

{---- TVoertuigData ----}

constructor TVoertuigData.Create;
begin
    inherited Create;
    Oplegger := nil;
end;

destructor TVoertuigData.Destroy;
begin
    if Oplegger <> nil then
        Oplegger.Destroy;
    inherited Destroy;
end;

{---- TETLDataModule ----}

```

```

procedure TETLDataModule.DataModuleCreate(Sender: TObject);
const
    ADOConnectStr = 'Provider=Microsoft.Jet.OLEDB.4.0;'
                  + 'Data Source=%s;'
                  + 'Persist Security Info=False';
var
    ExeFilePath: string;
begin
    FromDate := EncodeDate(FromYear, 1, 1); //base date for time key value calculation
    WritingToLog := False;
    CurrentDouaneBronSysteemRef := '';
    VoertuigSleutelGenerator := TSleutelGenerator.Create(AgoraVoertuigen,
        'VoertuigSleutel');
    ControleSleutelGenerator := TSleutelGenerator.Create(AgoraControles,
        'ControleSleutel');
    ExeFilePath := ExtractFilePath(Application.ExeName);
    DouaneConnection.Close;
    AgoraConnection.Close;
    NationMapConnection.Close;
    DouaneConnection.ConnectionString := Format(ADOConnectStr, [ExeFilePath +
        'TestDataDouane.mdb']);
    DouaneConnection.Open;
    DouaneTable.Open;
    AgoraConnection.ConnectionString := Format(ADOConnectStr, [ExeFilePath +
        'AgoraDB.mdb']);
    AgoraConnection.Open;
    AgoraControles.Open;
    AgoraVaststellingen.Open;
    AgoraBepalingen.Open;
    AgoraTijd.Open;
    AgoraVoertuigen.Open;
    NationMapConnection.ConnectionString := Format(ADOConnectStr, [ExeFilePath +
        'NationMapping.mdb']);
    NationMapConnection.Open;
    DouaneLandenMap.Open;
    LandCodes.Open;
end;

```

```

procedure TETLDataModule.DataModuleDestroy(Sender: TObject);
begin
    DouaneTable.Close;
    DouaneConnection.Close;
    AgoraControles.Close;
    AgoraVaststellingen.Close;
    AgoraBepalingen.Close;
    AgoraTijd.Close;
    AgoraVoertuigen.Close;
    AgoraConnection.Close;
    DouaneLandenMap.Close;
    LandCodes.Close;
    NationMapConnection.Close;
    VoertuigSleutelGenerator.Free;
    ControleSleutelGenerator.Free;
end;

```

```

{---- Sleutelberekening ----}

```

```

function TETLDataModule.GetDouaneBronSysteemRef: string;
begin
    Result := Format('%d-%s-%s-%s', [ControleDienstCodes[cdDouane],
        DouaneTable.FieldByName('FORSCODE').AsString, //code motorbrigade
        FormatDateTime('yyyy"-mm', DouaneTable.FieldByName('CONTROLEDATUM').AsDateTime),

```

```

        //year and month of CONTROLEDATUM (not INPUTDATUM!)
        DouaneTable.FieldByName('VOLGNUMMER').AsString]); //order within month
end;

function TETLDataModule.GetTijdSleutelValue(ADate: TDateTime): Integer;
begin
    Result := DaysBetween(ADate, FromDate) + 1;
end;

function TETLDataModule.GetMaxSleutelValue(ADOTable: TADOTable; SleutelName:
    string): Integer;
var
    Q: TADOQuery;
begin
    Q := TADOQuery.Create(Self);
    try
        Q.SQL.Clear;
        Q.SQL.Add(Format('SELECT MAX(%s) AS MaxSleutel', [SleutelName]));
        Q.SQL.Add(Format('FROM %s', [ADOTable.TableName]));
        Q.Connection := AgoraConnection;
        Q.Open;
        if Q.RecordCount = 0 then
            Result := 0
        else
            Result := Q.FieldByName('MaxSleutel').AsInteger;
        Q.Close;
    finally
        Q.Free;
    end;
end;

constructor TSleutelGenerator.Create(AnADOTable: TADOTable; ASleutelName: string);
begin
    ADOTable := AnADOTable;
    SleutelName := ASleutelName;
    NewSleutelValue := 0;
end;

procedure TSleutelGenerator.Reset;
begin
    NewSleutelValue := ETLDataModule.GetMaxSleutelValue(ADOTable, SleutelName);
end;

function TSleutelGenerator.GetNewSleutelValue: Integer;
begin
    Inc(NewSleutelValue);
    Result := NewSleutelValue;
end;

procedure TETLDataModule.FillTijdTabel;
var
    FillDate, ToDate: TDateTime;
    SleutelValue: Integer;
begin
    AgoraTijd.DisableControls;
    Screen.Cursor := crHourGlass;
    try
        ToDate := EncodeDate(ToYear, 12, 31);
        FillDate := FromDate;
        while FillDate <= ToDate do
            begin
                SleutelValue := GetTijdSleutelValue(FillDate);
                with AgoraTijd do

```

```

if not Locate('TijdSleutel', SleutelValue, []) then
begin
    Append;
    FieldByName('TijdSleutel').AsInteger := SleutelValue;
    FieldByName('Datum').AsDateTime := FillDate;
    FieldByName('Maand').AsInteger := MonthOf(FillDate);
    case MonthOf(FillDate) of
        1..3: FieldByName('Kwartaal').AsInteger := 1;
        4..6: FieldByName('Kwartaal').AsInteger := 2;
        7..9: FieldByName('Kwartaal').AsInteger := 3;
        10..12: FieldByName('Kwartaal').AsInteger := 4;
    end; //case
    FieldByName('Jaar').AsInteger := YearOf(FillDate);
    Post;
end; //if
    FillDate := IncDay(FillDate);
end;
finally
    AgoraTijd.EnableControls;
    Screen.Cursor := crDefault;
end;
end;

procedure TETLDataModule.DeleteADOTable(ADOTable: TADOTable);
var
    Q: TADOQuery;
    //using query here because DeleteRecords-method doesn't work on ADOTable
begin
    ADOTable.Close;
    Q := TADOQuery.Create(Self);
    try
        Q.SQL.Clear;
        Q.SQL.Add(Format('DELETE FROM %s', [ADOTable.TableName]));
        if ADOTable = AgoraVoertuigen then
            Q.SQL.Add('WHERE VoertuigSleutel > 0'); //prevent deletion of zero-record
        Q.Connection := ADOTable.Connection;
        Q.ExecSQL;
    finally
        ADOTable.Open;
        Q.Free;
    end;
end;

function TETLDataModule.UpdateVoertuigenTabel(VoertuigData: TVoertuigData;
var SleutelValue: Integer): Boolean;
begin
    with VoertuigData, AgoraVoertuigen do
        if not Locate('Nummerplaat;Nationaliteit', VarArrayOf([NummerPlaat,
            NationCodeSleutel]), [loCaseInsensitive]) then
            begin
                Append;
                SleutelValue := VoertuigSleutelGenerator.GetNewSleutelValue;
                FieldByName('VoertuigSleutel').AsInteger := SleutelValue;
                FieldByName('VoertuigCategorie').AsInteger := VoertuigCategorieCodes[
                    AgoraCategorie];
                FieldByName('Nummerplaat').AsString := NummerPlaat;
                FieldByName('Nationaliteit').AsInteger := NationCodeSleutel;
                Post;
                Result := True;
            end
        else
            begin
                SleutelValue := FieldByName('VoertuigSleutel').AsInteger;

```

```

    Result := False;
end;
end;

function TETLDataModule.ParseVoertuigField(VoertuigStr: string): TVoertuigData;
var
    PosSemiColon: Integer;

procedure ParseVoertuig(AStr: string; VoertuigData: TVoertuigData);
var
    Pos1stSlash, Pos2ndSlash: Integer;

procedure ParseCategorieStr(CatStr: string);
begin
    {Cleanse:}
    if (CatStr = '1') or SameText(CatStr, 'Voiture')
        or SameText(CatStr, 'Personenwagen') then
        VoertuigData.DouaneCategorie := vtPersonenWagen
    else
    if (CatStr = '2') or SameText(CatStr, 'Camionette')
        or SameText(CatStr, 'Lichte vrachtwagen') then
        VoertuigData.DouaneCategorie := vtLichteVrachtWagen
    else
    if (CatStr = '3') or SameText(CatStr, 'Camion')
        or SameText(CatStr, 'Vrachtwagen') then
        VoertuigData.DouaneCategorie := vtVrachtWagen
    else
    if (CatStr = '4') or SameText(CatStr, 'Autobus') then
        VoertuigData.DouaneCategorie := vtAutobus
    else
    if (CatStr = '5') or SameText(CatStr, 'Mobilhome') then
        VoertuigData.DouaneCategorie := vtMobilHome
    else
    if (CatStr = '6') or SameText(CatStr, 'Tracteur')
        or SameText(CatStr, 'Trekker') then
        VoertuigData.DouaneCategorie := vtTrekker
    else
    if (CatStr = '7') or SameText(CatStr, 'Remorque')
        or SameText(CatStr, 'Oplegger') then
        VoertuigData.DouaneCategorie := vtOplegger
    else
    if (CatStr = '8') then
        VoertuigData.DouaneCategorie := vtAndere
    else
    begin
        VoertuigData.DouaneCategorie := vtAndere;
        if WritingToLog then
            Writeln(LogFile, Format('Warning (%s): voertuigtype "%s" not recognized',
                [CurrentDouaneBronSysteemRef, CatStr]));
        end;
    }Convert to Agora code:}
    with VoertuigData do
        case DouaneCategorie of
            vtLichteVrachtWagen, vtVrachtWagen, vtTrekker, vtOplegger:
                AgoraCategorie := vcZaken;
            vtAutobus: AgoraCategorie := vcPersonen;
            vtPersonenWagen, vtMobilHome, vtAndere: AgoraCategorie := vcAndere;
        end; //case
    end; //ParseCategorieStr

procedure ParsePlaatStr(PlaatStr: string);
begin
    {Cleanse:}

```

```

PlaatStr := UpperCase(PlaatStr); //convert to capitals
PlaatStr := StringReplace(PlaatStr, ' ', '', [rfReplaceAll]); //remove spaces
PlaatStr := StringReplace(PlaatStr, '-', '', [rfReplaceAll]); //remove dashes
PlaatStr := StringReplace(PlaatStr, '_', '', [rfReplaceAll]); //remove underscores
if Pos('+', PlaatStr) > 0 then
begin
    if WritingToLog then Writeln(LogFile, Format(
        'Warning (%s): data following "+" in nummerplaat "%s" removed',
        [CurrentDouaneBronSysteemRef, PlaatStr]));
    PlaatStr := LeftStr(PlaatStr, Pos('+', PlaatStr) - 1);
end;
if Length(PlaatStr) > MaxNummerPlaatLength then
begin
    if WritingToLog then
        Writeln(LogFile, Format('Warning (%s): nummerplaat "%s" is truncated to %d chars',
            [CurrentDouaneBronSysteemRef, PlaatStr, MaxNummerPlaatLength]));
    PlaatStr := LeftStr(PlaatStr, MaxNummerPlaatLength);
end;
if PlaatStr = '' then
begin
    PlaatStr := OnbepaaldPlaatStr;
    if WritingToLog then
        Writeln(LogFile, Format('Warning (%s): nummerplaat data missing',
            [CurrentDouaneBronSysteemRef]));
end;
    VoertuigData.NummerPlaat := PlaatStr;
end; //ParsePlaatStr

procedure ParseNationStr(NationStr: string);
begin
    VoertuigData.NationCodeSleutel := OnbepaaldLandCode;
    if VoertuigData.NummerPlaat <> OnbepaaldPlaatStr then
begin
        NationStr := Trim(NationStr); //remove leading and trailing spaces if any
        {first try looking for it in the provided mapping table:}
        if DouaneLandenMap.Locate('LandBeschrijving', NationStr, [loCaseInsensitive])
        then
            VoertuigData.NationCodeSleutel := DouaneLandenMap.FieldName('LandCode').
                AsInteger
        else {if it has two chars, look for it in the ISO-code list}
        if Length(NationStr) = 2 then
begin
            if LandCodes.Locate('ISO', NationStr, [loCaseInsensitive]) then
                VoertuigData.NationCodeSleutel := LandCodes.FieldName('LandCode').AsInteger
            end
            else {look for it in full description}
            if LandCodes.Locate('Naam', NationStr, [loCaseInsensitive]) then
begin
                VoertuigData.NationCodeSleutel := LandCodes.FieldName('LandCode').AsInteger
            end
            else {if the string is empty, try two more things...}
            if NationStr = '' then
begin
                {try looking up NATIONALITEIT-field next to VOERTUIG-field instead}
                if DouaneTable.FieldName('NATIONALITEIT').AsString <> '' then
                    if DouaneLandenMap.Locate('LandBeschrijving', DouaneTable.FieldName(
                        'NATIONALITEIT').AsString, [loCaseInsensitive]) then
begin
                        VoertuigData.NationCodeSleutel := DouaneLandenMap.FieldName('LandCode').
                            AsInteger;
                        if WritingToLog then
                            Writeln(LogFile, Format('Warning (%s): replaced missing country code ' +
                                'with %s for nummerplaat "%s" based on adjacent NATIONALITEIT field',

```

```

        [CurrentDouaneBronSysteemRef, DouaneLandenMap.FieldByName(
            'LandBeschrijving').AsString, VoertuigData.Nummerplaat]);
    end
    else //give up
    else
    begin
        {if no nationality data are present, assume Belgian nationality if
        nummerplaat has the right format (3 characters followed by 3 digits)}
        with VoertuigData do
            if (Length(Nummerplaat) = 6) then
                if (Nummerplaat[1] in ['A'..'Z'])
                    and (Nummerplaat[2] in ['A'..'Z'])
                    and (Nummerplaat[3] in ['A'..'Z'])
                    and (Nummerplaat[4] in ['0'..'9'])
                    and (Nummerplaat[5] in ['0'..'9'])
                    and (Nummerplaat[6] in ['0'..'9'])
                then
                begin
                    NationCodeSleutel := BelgiumLandCode;
                    if WritingToLog then
                        Writeln(LogFile, Format('Warning (%s): replaced missing country code ' +
                            'with Belgium for nummerplaat "%s"',
                            [CurrentDouaneBronSysteemRef, Nummerplaat]));
                    end;
                end;
            end;
        end;
        if (VoertuigData.NationCodeSleutel = OnbepaaldLandCode) and WritingToLog then
            Writeln(LogFile, Format('Warning (%s): "%s" not identified as valid country code ' +
                'for nummerplaat "%s"', [CurrentDouaneBronSysteemRef, NationStr, VoertuigData.
                Nummerplaat]));
        end;
    end; //ParseNationStr

begin //ParseVoertuig
    Pos1stSlash := Pos('/', AStr);
    Pos2ndSlash := PosEx('/', AStr, Pos1stSlash + 1);
    if Pos2ndSlash = 0 then
        Pos2ndSlash := Length(AStr) + 1;
        //if second slash is missing, assume remainder is nummerplaat, not nationality
    ParseCategorieStr(LeftStr(AStr, Pos1stSlash - 1));
    if VoertuigData.AgoraCategorie <> vcAndere then
        //we're only interested in transport of goods or persons (cf. infra)
    begin
        ParsePlaatStr(Copy(AStr, Pos1stSlash + 1, Pos2ndSlash - Pos1stSlash - 1));
        ParseNationStr(RightStr(AStr, Length(AStr) - Pos2ndSlash));
    end
    else //save time by not parsing other types of vehicles
    begin
        VoertuigData.NummerPlaat := OnbepaaldPlaatStr;
        VoertuigData.NationCodeSleutel := OnbepaaldLandCode;
    end;
end; //ParseVoertuig

begin //ParseVoertuigField
    Result := TVoertuigData.Create;
    Result.Oplegger := nil;
    if LeftStr(VoertuigStr, 1) = ';' then Delete(VoertuigStr, 1, 1); //remove leading ';'
    PosSemiColon := Pos(';', VoertuigStr);
    if PosSemiColon = 0 then //geen geleed voertuig
        ParseVoertuig(VoertuigStr, Result)
    else
        begin //geleed voertuig
            ParseVoertuig(LeftStr(VoertuigStr, PosSemiColon - 1), Result);
        end;
    end;
end;

```

```

    Result.Oplegger := TVoertuigData.Create;
    ParseVoertuig(RightStr(VoertuigStr, Length(VoertuigStr) - PosSemiColon),
        Result.Oplegger);
    if Result.Oplegger.Nummerplaat = '' then
    begin //if parsing of second vehicle fails, remove reference
        Result.Oplegger.Free;
        Result.Oplegger := nil;
    end;
end;
end; //ParseVoertuigField

function ParseListString(ListStr: string; Separator: Char): TStringList;
{parses ListStr into a stringlist (e.g., ParseListString('a; b;c', ';')}
returns a stringlist containing: 'a', 'b', 'c'; returns a one-element list
containing the empty string, if ListStr is empty; do not forget to free the
newly created result-stringlist after its use}
var
    Pos1stSep, Pos2ndSep: Integer;
    ItemStr: string;
begin
    Result := TStringList.Create;
    Pos1stSep := 0;
    repeat
        Pos2ndSep := PosEx(Separator, ListStr, Pos1stSep + 1);
        if Pos2ndSep = 0 then Pos2ndSep := Length(ListStr) + 1;
        ItemStr := Copy(ListStr, Pos1stSep + 1, Pos2ndSep - Pos1stSep - 1);
        ItemStr := Trim(ItemStr);
        Result.Add(ItemStr);
        Pos1stSep := Pos2ndSep;
    until Pos1stSep = Length(ListStr) + 1;
end;

procedure TETLDataModule.UpdateVaststellingenTabel(ControlSleutelValue: Integer;
    ControlSoortStr, VaststellingStr: string; var Overtreding: Boolean;
    var PVCCount: Integer);
var
    ControleSoortLijst, VaststellingenLijst: TStringList;
    ControleSoort: string;
    i, DouaneBepCode: Integer;
begin
    Overtreding := False;
    PVCCount := 0;
    ControleSoortLijst := ParseListString(ControlSoortStr, ';');
    VaststellingenLijst := ParseListString(VaststellingStr, ';');
    try
        for i := 0 to ControleSoortLijst.Count - 1 do
        begin
            ControleSoort := ControleSoortLijst.Strings[i];
            if ControleSoort <> '' then
            if TryStrToInt(ControleSoort, DouaneBepCode) then
            if AgoraBepalingen.Locate('Code;Dienst', VarArrayOf([DouaneBepCode,
                ControleDienstCodes[cdDouane]]), [loCaseInsensitive]) then
            with AgoraVaststellingen do
            begin
                Append;
                FieldByName('ControleSleutel').AsInteger := ControlSleutelValue;
                FieldByName('BepalingSleutel').AsInteger :=
                    AgoraBepalingen.FieldByName('Bep_Sleutel').AsInteger;
                FieldByName('Overtreding').AsBoolean := (VaststellingenLijst.IndexOf(
                    ControleSoort) <> -1); //overtreding yes/no
                if VaststellingenLijst.IndexOf(ControleSoort) <> -1 then
                    Overtreding := True;
                if (DouaneBepCode = 133) {PV 108} or (DouaneBepCode = 134) {PV 109} or

```

```

        (DouaneBepCode = 410) {PV douane & accijnzen}
    then Inc(PVCount);
    //no data available on "Gevolg", "AantalWerknemers" or "Bedrag" at this level
    Post;
end //with
else //Locate failed
begin
    if WritingToLog then
        Writeln(LogFile, Format('Warning (%s): non-identifiable ' +
            'code %d found in CONTROLESOORT/VASTSTELLING',
            [CurrentDouaneBronSysteemRef, DouaneBepCode]));
    end
else //TryStrToInt failed
begin
    if WritingToLog then
        Writeln(LogFile, Format('Warning (%s): non-numeric entry "%s" found in ' +
            'CONTROLESOORT/VASTSTELLING', [CurrentDouaneBronSysteemRef, ControleSoort]));
    end
else //do nothing
end;
finally
    ControleSoortLijst.Free;
    VaststellingenLijst.Free;
end;
end;

procedure TETLDataModule.ExtractCurrentControleRecord;
var
    VoertuigData: TVoertuigData;
    CurrentControleSleutelValue: Integer;
    VoertuigSleutelValue, VoertuigOpleggerSleutelValue: Integer;
    Overtreding: Boolean;
    PVCount: Integer;
begin
    VoertuigData := ParseVoertuigField(DouaneTable.FieldByName('VOERTUIG').AsString);
    try
        if VoertuigData.AgoraCategorie <> vcAndere then
            //we're only interested in transport of goods or persons
            with AgoraControles do
                begin
                    CurrentControleSleutelValue := ControleSleutelGenerator.GetNewSleutelValue;
                    CurrentDouaneBronSysteemRef := GetDouaneBronSysteemRef;
                    Append;
                    FieldByName('ControleSleutel').AsInteger := CurrentControleSleutelValue;
                    FieldByName('ControleDienst').AsInteger := ControleDienstCodes[cdDouane];
                    FieldByName('District').AsInteger := OnbepaaldCode;
                    FieldByName('Aard').AsInteger := ControleAardCodes[caWegControle];
                    case DouaneTable.FieldByName('CONTROLEAARD').AsInteger of
                        1: FieldByName('Oorsprong').AsInteger := OorsprongCodes[osRoutineControle];
                        2: FieldByName('Oorsprong').AsInteger := OorsprongCodes[osGerichteControle];
                        3: begin
                            FieldByName('Oorsprong').AsInteger := OorsprongCodes[osGecoordOperatie];
                            FieldByName('OperatieCodeNaam').AsString := DouaneTable.FieldByName(
                                'CODENAAM').AsString;
                        end;
                    else
                        FieldByName('Oorsprong').AsInteger := OnbepaaldCode;
                    end; //case
                    FieldByName('BronSysteemRef').AsString := CurrentDouaneBronSysteemRef;
                    FieldByName('InvoerDatumSleutel').AsInteger := GetTijdSleutelValue(Today);
                    FieldByName('OpstartDatumSleutel').AsInteger := GetTijdSleutelValue(
                        DouaneTable.FieldByName('CONTROLEDATUM').AsDateTime);
                    FieldByName('AfsluitDatumSleutel').AsInteger := GetTijdSleutelValue(

```

```

        DouaneTable.FieldName('CONTROLEDATUM').AsDateTime);
        //wegcontrole: opstartdatum = afsluitdatum
    FieldByName('PlaatsOmschrijving').AsString := DouaneTable.FieldName(
        'CONTROLEPUNT').AsString;
    UpdateVoertuigenTabel(VoertuigData, VoertuigSleutelValue);
    if VoertuigData.Oplegger <> nil then
        UpdateVoertuigenTabel(VoertuigData.Oplegger, VoertuigOpleggerSleutelValue)
    else
        VoertuigOpleggerSleutelValue := OnbepaaldCode;
        FieldByName('VoertuigSleutel').AsInteger := VoertuigSleutelValue;
        FieldByName('VoertuigOpleggerSleutel').AsInteger :=
            VoertuigOpleggerSleutelValue;
        UpdateVaststellingenTabel(CurrentControleSleutelValue,
            DouaneTable.FieldName('CONTROLESOORT').AsString,
            DouaneTable.FieldName('VASTSTELLING').AsString,
            Overtreding, PVCCount);
        FieldByName('BestuurderSleutel').AsInteger := OnbepaaldCode;
        //data not available in douane-database
        FieldByName('WerkgeverSleutel').AsInteger := OnbepaaldCode;
        //data almost never entered in douane-database
        FieldByName('Overtreding').AsBoolean := Overtreding;
        FieldByName('AantalPVs').AsInteger := PVCCount;
        FieldByName('AantalGecontrolVoertuigen').AsInteger := 1; //wegcontrole
        {GeindBedrag en RegularisatieBedrag not available}
        Post;
    end; //with
finally
    VoertuigData.Free;
end;
end;

procedure TETLDataModule.TestETL;
begin
    ControleSleutelGenerator.Reset;
    VoertuigSleutelGenerator.Reset;
    ExtractCurrentControleRecord;
    CurrentDouaneBronSysteemRef := '';
end;

const
    MAXRECORDS = MaxInt;
    //testing: set to lower value to extract only first MAXRECORDS records

var
    CURRENTRECORDS: Integer;

procedure TETLDataModule.ExecuteETL;
var
    ProgressDlg: TProgressDialog;
begin
    DouaneTable.DisableControls;
    AgoraControles.DisableControls;
    AgoraVoertuigen.DisableControls;
    ProgressDlg := TProgressDialog.Create(nil);
    AssignFile(LogFile, ExtractFilePath(Application.ExeName) + 'Log.txt');
    try
        Rewrite(LogFile);
        WritingToLog := True;
        Writeln(LogFile, Format('Execution started at %s.', [DateTimeToStr(Now)]));
    try
        DouaneTable.First;
        CURRENTRECORDS := 1; //testing
        ControleSleutelGenerator.Reset;

```

```

VoertuigSleutelGenerator.Reset;
with ProgressDlg do
begin
  Title := 'Processing Test Data';
  Options := [prShowProgress, prCancel];
  Open;
end; //with
while not DouaneTable.EOF do
begin
  with ProgressDlg do
  begin
    Text := Format('%d of %d records completed', [DouaneTable.RecNo - 1,
      DouaneTable.RecordCount]);
    PercentageCompleted := Round(100*(DouaneTable.RecNo - 1)/
      (DouaneTable.RecordCount));
  end; //with
  ExtractCurrentControleRecord;
  Application.ProcessMessages;
  if ProgressDlg.Cancelled then
  Abort;
  if CURRENTRECORDS >= MAXRECORDS then
  Abort; //testing
  DouaneTable.Next;
  Inc(CURRENTRECORDS); //testing
end; //while
if WritingToLog then
  Writeln(LogFile, Format('Execution completed succesfully at %s.',
    [DateTimeToStr(Now)]));
except on E:Exception do
begin
  if WritingToLog then
  if E is EAbort then
    Writeln(LogFile, Format('Execution aborted at %s.', [DateTimeToStr(Now)]))
  else
    Writeln(LogFile, Format('Execution failed at record %s. %s.',
      [CurrentDouaneBronSysteemRef, E.Message]));
  raise;
end;
end;
finally
  DouaneTable.EnableControls;
  AgoraControles.EnableControls;
  AgoraVoertuigen.EnableControls;
  ProgressDlg.Free;
  CloseFile(LogFile);
  WritingToLog := False;
  CurrentDouaneBronSysteemRef := '';
end;
end;

function TETLDataModule.LocateAgoraTargetRecord: Boolean;
var
  DouaneRef: string;
begin
  DouaneRef := GetDouaneBronSysteemRef;
  Result := AgoraControles.Locate('BronSysteemRef', DouaneRef, [loCaseInsensitive]);
end;

```