
Datawarehouse ‘Wegvervoercontroles’
(Federaal Wetenschapsbeleid – Agora/01/082)

eindrapport en toelichting ontwerp van model

Christophe Mues
Promotor: Jan Vanthienen

K.U.Leuven

Dept. Toegepaste Economische Wetenschappen

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
jan.vanthienen@econ.kuleuven.ac.be

Inhoudsopgave

1.	SITUERING EN DEELNEMENDE INSTANTIES	3
2.	DOELSTELLINGEN VAN HET DATAWAREHOUSE-PROJECT	3
3.	DATAWAREHOUSING: METHODOLOGISCHE ASPECTEN	5
3.1.	Definitie en kenmerken van een datawarehouse.....	5
3.2.	Het ontwerpen van een datawarehouse	6
3.3.	Het datawarehouse-bevoorradingsproces.....	10
3.4.	Gebruik van het datawarehouse	11
4.	INVENTARIS HUIDIGE SITUATIE.....	12
4.1.	Vervoer te land	12
4.2.	Douane & Accijnzen.....	14
4.3.	Rijksdienst voor Sociale Zekerheid (R.S.Z.).....	14
4.4.	Inspectie van de Sociale Wetten (I.S.W.).....	16
4.5.	Sociale Inspectie (S.I.)	17
4.6.	Federale en Lokale Politie	18
5.	ONTWERP VAN MODEL EN TOELICHTING	18
6.	PROOF-OF-CONCEPT OEFENING: VERWERKING DOUANEGEGEVENS	23
6.1.	ETL-implementatie	23
6.2.	Raadpleging doelgegevens – demonstratieomgeving	28
7.	CONCLUSIE: HUIDIGE SITUATIE & AANBEVELINGEN VERVOLGTRAJECT	33
7.1.	Op wetgevend en juridisch vlak.....	33
7.2.	Op inhoudelijk en technisch vlak.....	34
7.3.	Op organisatorisch en budgettair vlak.....	34
7.4.	Overzicht mogelijke vervolgstappen.....	35
	BIBLIOGRAFIE	36
	AFKORTINGEN EN ACRONIEMEN	36
	BIJLAGEN	36

1. Situering en deelnemende instanties

Het datawarehouse ‘Wegvervoercontroles’ wordt opgezet door de Directie Controle en Transportorganisatoren van het Directoraat-generaal Vervoer te Land van de FOD Mobiliteit en Vervoer, in overeenstemming met het actieplan van 20 november 2001 betreffende de samenwerking tussen de verschillende controlediensten voor het personen- en goederenvervoer over de weg (zoals verschenen in het Belgisch staatsblad van 19/02/2002). Voor de wetenschappelijke ondersteuning van het project wordt een beroep gedaan op het departement Toegepaste Economische Wetenschappen van de K.U.Leuven, in het kader van het Agora-project AG/01/082 van de Programmatorische Overheidsdienst Wetenschapsbeleid. De als partners in het actieplan en projectpartners betrokken, en aldus als datatoeleveranciers en/of -gebruikers fungerende, diensten zijn:

- Directoraat-generaal Vervoer te Land;
- Douane & Accijnzen;
- Inspectie van de Sociale Wetten (FOD Werkgelegenheid, Arbeid en Sociaal Overleg);
- Sociale Inspectie (FOD Sociale Zekerheid);
- de inspectiedienst van de Rijksdienst voor Sociale Zekerheid;
- Federale en lokale politie.

Dit actieplan beoogt door een verbeterd toezicht op de transportsector bij te dragen tot:

- de verhoging van de verkeersveiligheid;
- de bevordering van billijke mededinging tussen de vervoerders;
- de strijd tegen illegale arbeid en sociale fraude;
- de verbetering van de werkomstandigheden van de bestuurders;
- de bescherming van het leefmilieu,

en voorziet hiertoe in de oprichting van een aantal nieuwe organen (onder meer van negen provinciale cellen “vervoer”), de invoering van gecoördineerde controleacties (waaraan controleurs uit de hoger vermelde diensten gezamenlijk deelnemen), en de onderlinge uitwisseling van informatie tussen de controle- en politiediensten. Daarbij stelt zij de uitwerking van een systeem van informatie-uitwisseling voorop en de oprichting van een databank die nuttig is voor de werking van het plan; het is in deze context dat het datawarehouse “Wegvervoercontroles” zich situeert.

2. Doelstellingen van het datawarehouse-project

Rekening houdende met de algemene doelstellingen van het hoger vermelde actieplan, de huidige werking en de informatiebehoefte van de diverse vervoercontrolediensten, zoals die door ons vastgesteld werden, menen wij dat het voorliggende datawarehouse-project volgende concrete doelstellingen dient voor ogen te houden.

Europese rapportering. In de eerste plaats moet het datawarehouse, op relatief eenvoudige wijze, aan het Directoraat-generaal Vervoer te Land toelaten de meest pertinente informatie inzake vastgestelde (zware) vervoersovertredingen door te geven aan andere instanties op hun eigen vraag, conform reglementaire bepalingen of afspraken, of nog ter ondersteuning van het beleid. Deze rapportering gebeurt in het bijzonder met andere Europese transportcontrolediensten in het raam van de administratieve regeling ‘Euro-Controle-Wegen’. De aanlevering van de daartoe vereiste data door de andere partners en hun invoer in het datawarehouse moeten het dan ook mogelijk maken om, op een meer systematische basis, informatie over de in

België vastgestelde overtredingen door wegvervoerders uit een van de andere deelnemende landen over te maken aan de bevoegde Europese instantie.

Ondernemingsprofiel. Ten tweede moet het datawarehouse ertoe bijdragen dat een nauwkeuriger profiel geschetst kan worden van het gedrag van een individuele transportonderneming. Waar de diverse partners zich voorheen enkel konden baseren op de zelf vergaarde controledata, zou het datawarehouse hun voortaan ook toegang moeten geven tot de resultaten aangeleverd door de andere partners of door andere 'Euro-Controle-Wegen'-landen, en zou deze verzamelde informatie geaggregeerd voorgesteld kunnen worden op het niveau van een individuele onderneming. Dit maakt het datawarehouse een geschikt instrument om het handhavingsbeleid op het vlak van de gecontroleerde reglementeringen te ondersteunen: zij kan immers gerichte informatie verschaffen op basis waarvan het Directoraat-generaal Vervoer te Land kan besluiten de vergunning van een bepaalde transportonderneming in te trekken. Bovendien kan dergelijke informatie ook een indicatie vormen voor de wenselijkheid van complementair onderzoek. Indien bijvoorbeeld bij wegcontroles door het Directoraat-generaal Vervoer te Land of door de Douane & Accijnzen geregeld inbreuken worden vastgesteld bij de voertuigen van een bepaald bedrijf, kunnen I.S.W. of S.I. het initiatief nemen om ook een grondig bedrijfsonderzoek te verrichten naar de naleving van de arbeids- en sociale reglementering.

Onderlinge coördinatie. Zo speelt het datawarehouse ten derde ook een rol in het verbeteren van de coördinatie tussen de diverse controlediensten, doordat zij toelaat dat controleurs kennis kunnen nemen van de controleresultaten van andere diensten. Let wel, de periodiciteit van de data-aanvoer houdt uiteraard zekere beperkingen in voor de beschikbaarheid van recent uitgevoerde controles.

Sectorstatistieken en -onderzoek. Hoewel de aanmaak van statistieken, of de uitvoering van allerhande verificatie- of ontdekkingsgebaseerde vormen van analyses (cf. infra: sectie 3.4), geen initiële gebruiksdoelinden zijn voor de bouw van het datawarehouse, kan dit laatste wel de gepaste basis leveren voor een vervolgproject hierover. Het ontwerp van de databank (cf. infra) is immers zo opgevat dat zij efficiënte analyses moet mogelijk maken. Wij denken hierbij bijvoorbeeld aan de berekening van statistieken over aantallen controles, overtredingen en hun onderlinge verhouding, en dit ingedeeld volgens een aantal mogelijke aggregatiecriteria (zoals vracht- versus personenvervoer, soort inbreuk, herkomst van het bedrijf, en zo meer), of aan het opsporen van potentiële indicatoren van frauduleus gedrag. Het feit dat dergelijke analyses weer over de verzamelde gegevens van de verschillende controlediensten heen kunnen reiken, moet een adequate beeldvorming mogelijk maken rond het functioneren van de transportsector in zijn geheel en zou zo in de toekomst waardevolle informatie kunnen verstrekken voor de beleidsvorming.

Toegang tot externe data. Los van de in de diverse operationele bronnen van het datawarehouse gevatte data, bestaan er nog een aantal additionele databanken over de diverse actoren in de vervoersector (werkgevers en -nemers, voertuiggegevens, enz.), die momenteel slechts beperkt toegankelijk zijn, maar waarvan in het kader van dit project nagegaan wordt in hoeverre zij opengesteld kunnen worden voor alle deelnemende partners. Wij denken hierbij in de eerste plaats aan:

- de Transis-vergunningendatabank;
- de nummerplaatgegevens van de Dienst voor Inschrijving van de Voertuigen (D.I.V.);
- de DIMONA-bank (onmiddellijke aangifte van tewerkstelling).

Zij vormen ook een mogelijke verrijking voor het datawarehouse, onder meer omdat zij zouden toelaten de aangeleverde controledata te valideren.

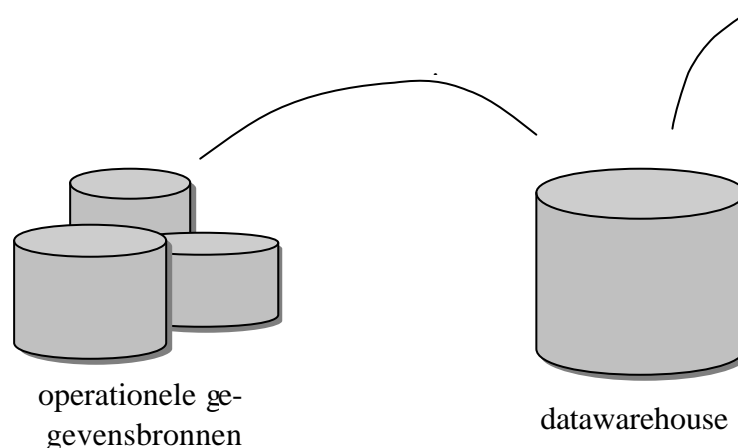
3. Datawarehousing: methodologische aspecten

In deze sectie schetsen wij kort wat bedoeld wordt met een datawarehouse, en welke de in de onderzoeksliteratuur voorgestelde ontwerpprincipes en -methodes zijn die daarvoor gelden.

3.1. Definitie en kenmerken van een datawarehouse

Ons baserend op de definitie van Inmon (1996)¹, zien wij een datawarehouse als een geïntegreerde, subjectgeoriënteerde, tijdsafhankelijke en onveranderlijke verzameling van gegevens die het beslissingsproces in een bedrijf of organisatie kan ondersteunen.

Geïntegreerd. Wellicht het voornaamste kenmerk van een datawarehouse is dat het, op een zo geautomatiseerd mogelijke wijze, gegevens verzamelt en integreert uit een heterogene verzameling bronsystemen. Vermits deze bronsystemen doorgaans elk op hun eigen manier ontworpen werden en afzonderlijk van elkaar verder evolueerden, is van enige consistentie in de gebruikte coderingswijzen, benaming en meting van attributen, sleutelstructuur, en zo meer, in regel geen sprake. Om toch betrouwbare globale bevestigingen of analyses over de verschillende systemen heen mogelijk te maken, is het daarom nodig om de structuur en inhoud van de diverse data-toevoerstromen zodanig te bewerken dat zij consistent gemaakt worden met die van het globale datawarehouse.



Figuur 1: gegevensintegratie in het datawarehouse

Subjectgeoriënteerd. Een tweede eigenschap van een datawarehouse is dat het typisch gestructureerd is rond enkele welomlijnde subjecten uit de organisatierealiteit, waaromtrent men informatie en kennis wenst te vergaren (bv. klanten, producten, verkopen, en dergelijke, of, in onze context, transportondernemingen, uitgevoerde controles, enz.), in plaats van een afspiegeling te vormen van de interne organisatiestructuur en haar verschillende applicatiedomeinen.

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

Tijdsafhankelijk. Daar waar een operationele databank in eerste instantie ontworpen is om een momentopname van de huidige toestand te bevatten, die dus ook continu – en dan nog liefst in real time – up-to-date moet gehouden worden (bv. welke onderzoeken zijn lopende? Voor welk bedrijf werkt een bepaalde vrachtwagenbestuurder?), is het de primaire doelstelling van een datawarehouse om de feitenhistoriek bij te houden (bv. welke bedrijven werden het afgelopen jaar gecontroleerd?). Hoewel dit onderscheid niet altijd even strikt is (zo worden vaak ook wel historische gegevens bewaard in een operationele databank, zij het over een beperktere tijdspanne), laat dit verschillend uitgangspunt uiteraard zijn invloed gelden op het databankontwerp. Zo zullen de meeste tabellen in het datawarehouse ook een tijdelement bevatten; ook is het minder prangend om altijd over de meest recente informatie te beschikken en ligt de data-aanvoerfrequentie dus typisch veel lager.

Onveranderlijk. Hiermee samenhangend bestaat een vierde onderscheidend kenmerk van een datawarehouse erin dat de opgeslagen gegevens in principe niet meer gewijzigd worden; er worden enkel nieuwe gegevens toegevoegd. Stel, bijvoorbeeld, dat bij een nieuwe controle blijkt dat de nationaliteit van een vrachtwagenbestuurder gewijzigd is, en wij deze data rechtstreeks zouden actualiseren in het eerder opgeslagen record voor deze persoon. In dat geval zouden alle voormalige controles die betrekking hadden op diezelfde persoon eveneens indirect naar dit gewijzigd gegeven verwijzen, met het vanzelfsprekende verlies van historische informatie (en de vertekening van eventuele statistische analyses) als gevolg. Liever dan een dergelijk record te muteren, wordt er in een datawarehouse meestal de voorkeur aan gegeven om de identificatiesleutel ook een tijdelement te laten bevatten (bv. geldig van ... tot ...). Zo zou, in dit voorbeeld, de tabel met bestuurders twee records kunnen bevatten voor dezelfde persoon, elk met een eigen geldigheidsperiode.

Beslissingsondersteunend. Tenslotte heeft een datawarehouse een specifiek doel en gebruikersgroep: het moet de basis vormen voor een beslissingsondersteunend systeem. Niet alle data hoeven daarom in het datawarehouse opgenomen te worden, enkel die die potentieel bijdragen tot het verbeteren van de genomen beslissingen. Vaak gaat het daarbij om beslissingen van tactische en/of strategische aard; in ons geval gaat het (zeker in een eerste fase waarin nog geen sectorstatistieken gegenereerd worden) evenzeer om operationele aspecten, bv. het inschatten van het gedrag van een individuele transportonderneming ter ondersteuning van het handhavingsbeleid.

Op basis van het voorgaande menen wij dat de functionaliteit en kenmerken van de te ontwikkelen databank voor vervoercontroles in voldoende mate beantwoordt aan het profiel van een datawarehouse, om ons, bij het ontwerp ervan, mede te baseren op de datawarehouse-onderzoeksliteratuur. In de volgende subsectie gaan wij dan ook kort in op enkele relevante ontwerpaspecten.

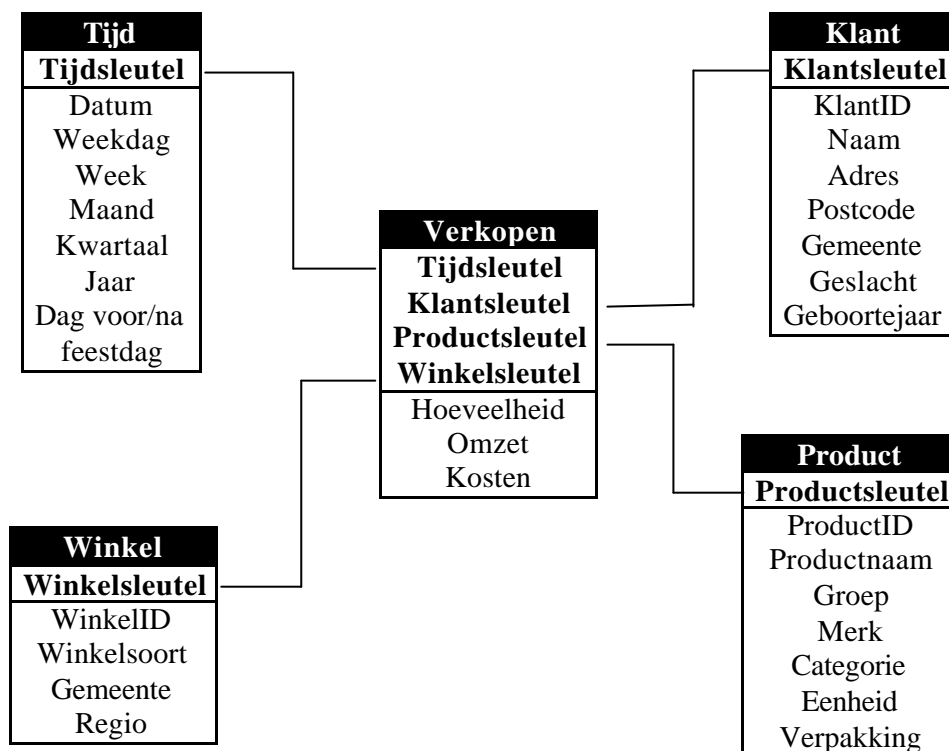
3.2. Het ontwerpen van een datawarehouse

Een interessant gevolg van het feit dat er in principe geen mutaties hoeven te gebeuren in een datawarehouse, is dat gegevensduplicatie en -denormalisatie niet dezelfde inconsistentie- en performantieproblemen veroorzaken als bij de operationele databanken van waaruit zij gevoed worden. Bovendien worden nieuw toegevoegde data typisch in batch ingeladen, wat een heel ander gebruikspatroon oplevert dan dat van de talrijke kleine transacties typisch voor een operationeel systeem. Bovendien ligt de nadruk bij de leestoegang in regel niet op het frequent ophalen van kleine hoeveelheden detaildata, maar op minder talrijke maar grootschaligere verwerkingen (bv. de aggregatie van cijfergegevens per kwartaal, of per regio, enz.). Samengevat, de update-performantie van datawarehouse-systemen is minder kritiek; wél

moeten zij goed kunnen omgaan met grote hoeveelheden gegevens en met complexe queries. Het is dan ook niet verwonderlijk dat het ontwerp van een datawarehouse vaak anders dan dat van een operationele databank gebeurt.

Sterschema. Een populaire datawarehouse-ontwerpmethode is gebaseerd op het definiëren van een zogeheten “sterschema”. Zo'n sterschema is niets anders dan een relationeel ontwerp waarbij vanuit één centrale feitentabel verwezen wordt naar een of meer dimensietabellen.

Figuur 2 hieronder geeft een fictief voorbeeld van een sterschema voor de verkopen van een winkelketen. Figuur 3 toont de ingevulde rijen voor een voorbeeldverkoop (van een krat Spa bruis aan de heer Lens uit Heverlee, op 22 juli 2003, in de superette van Leuven).



Figuur 2: voorbeeld van een sterschema voor winkelketenverkopen²

De feitentabel is een centrale referentietabel waarlangs toegang tot de gearchiveerde gebeurtenissen of activiteiten die inherent zijn aan een bepaald proces, wordt verkregen. In het voorbeeld hierboven gaat het om een winkelketen die, uitgaande van een verkoopproces, feiten registreert die betrekking hebben op wie, wat, waar en wanneer heeft gekocht (cf. de tabel ‘verkopen’). Een feitentabel bevat in hoofdzaak numerieke informatie: een meervoudige sleutel, waarlangs wordt verwezen naar rijen van de dimensietabellen (cf. de tabellen ‘tijd’, ‘klant’, ‘product’ en ‘winkel’), en een aantal aggregerbare meetwaarden die toewijsbaar zijn aan een bepaald feit (hier: hoeveelheid, omzet en kosten). Een belangrijke ontwerpbeslissing

² Vandenbulcke, J.A. en Lemahieu, W., Databasesystemen voor de praktijk, zevende druk, Ten Hagen en Stam, Den Haag, 560 pp.

hier is de bepaling van de fijnmazigheid van de feitentabel: op welk detail- of aggregatieniveau definiëren wij een atomair feit? Zo lang de benodigde opslagcapaciteit geen onoverkomelijk probleem vormt, verdient het wellicht aanbeveling om een zo groot mogelijke fijnmazigheid aan te houden. Immers, aggregatie kan achteraf nog steeds toegepast worden (vaak worden een reeks aggregaattabellen, waarin een of meer dimensies geheel of gedeeltelijk verdwijnen, afgeleid uit de feitentabel).

Een dimensietabel bestaat typisch uit een betekenisloze sleutel die het verband legt met rijen uit de feitentabel (bv., ‘productsleutel’), een betekenisvolle sleutel die is overgenomen uit een operationele of externe gegevensbron (bv., ‘productID’) en een aantal attributen waarmee de dimensie kan worden gekarakteriseerd (bv., productnaam, groep, merk, ...). Deze laatste verschaffen de selectiecriteria waarover de meetwaarden in de feitentabel kunnen worden geaggregeerd. Zo zouden wij, in het gegeven voorbeeld, een rapport kunnen opvragen over de verkochte hoeveelheden en omzet per productgroep en (daarbinnen) per winkelregio.

Verkopen						
<i>Tijdsleutel</i>	<i>Klantsleutel</i>	<i>Productsleutel</i>	<i>Winkelsleutel</i>	<i>Hoeveelheid</i>	<i>Omzet</i>	<i>Kosten</i>
203	55	68	07	1	7,00	6,25

Tijd							
<i>Tijdsleutel</i>	<i>Datum</i>	<i>Weekdag</i>	<i>Week</i>	<i>Maand</i>	<i>Kwartaal</i>	<i>Jaar</i>	<i>Dag voor/na feestdag</i>
203	22/07/03	2	30	juli	3	2003	Ja

Klant							
<i>Klantsleutel</i>	<i>KlantID</i>	<i>Naam</i>	<i>Adres</i>	<i>Postcode</i>	<i>Gemeente</i>	<i>Geslacht</i>	<i>Geboortjaar</i>
55	K100567	Lens	Burgstraat 45	B3001	Heverlee	M	1975

Product							
<i>Product-sleutel</i>	<i>ProductID</i>	<i>Product-naam</i>	<i>Groep</i>	<i>Merk</i>	<i>Categorie</i>	<i>Eenheid</i>	<i>Verpakking</i>
68	P658942	Water	Bruis	Spa	Levensmiddel	6	Glas

Winkel				
<i>Winkelsleutel</i>	<i>WinkelID</i>	<i>Winkelsoort</i>	<i>Gemeente</i>	<i>Regio</i>
07	W894	Superette	Leuven	Vlaams Brabant

Figuur 3: winkelketenverkopen; registratie van een voorbeeldverkooptransactie

Bij het voorgaande kan worden opgemerkt dat dimensietabellen meestal zijn gedenormaliseerd.³ Indien men de hiërarchische structuren binnen dimensies toch rechtstreeks expliciteert, dan spreekt men van een “sneeuwvlokschema”. Zo zou de (transitieve) functionele afhankelijkheid tussen de attributen ‘Gemeente’ en ‘Regio’ in een afzonderlijke tabel kunnen opgenomen worden, waarnaar dan verwezen wordt vanuit de winkeltabel. Een dergelijke

³ Zoals eerder aangehaald: omdat er over het algemeen geen mutatie-opdrachten op de databank uitgevoerd worden, kunnen er dus ook geen update-anomalieën optreden, terwijl de performantie hier mogelijk mee kan verbeterd worden (door het vermijden van dure join-operaties).

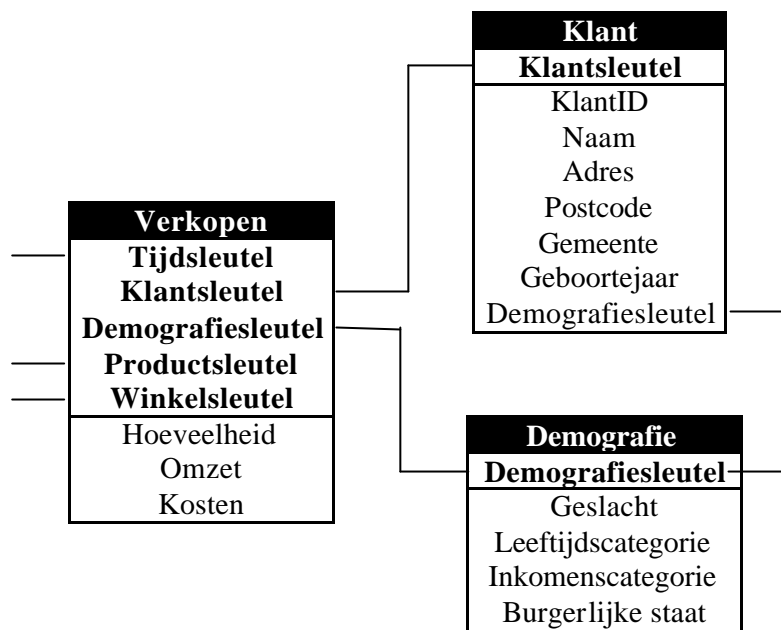
strategie kan weliswaar ruimtebesparingen opleveren maar zal ook vaak een negatieve invloed hebben op de query-performantie.

Omgaan met historiek. Het hoger beschreven voorbeeldschema laat toe de verkoophistoriek rechtstreeks in te bedden en maakt het eenvoudig om allerlei analyses uit te voeren waarin de tijdsdimensie een rol speelt. In de loop van de tijd kan echter de situatie weergegeven door de attributen in de dimensietabellen wijzigen. Zo kan een klant op zeker ogenblik van adres veranderen. Willen wij ervoor zorgen dat met dergelijke wijzigingen op correcte wijze wordt omgegaan in de analyses, dan moet hiermee in het model rekening gehouden worden. Een eerste mogelijke werkwijze houdt in dat wij het betreffende record in de dimensietabel zouden overschrijven. Dit zou echter verlies van historische informatie met zich mee brengen: willen we onderzoeken hoe het verkoops patroon afhangt van de woonplaats van de klant, dan kan deze oplossing leiden tot foutieve conclusies. Immers, de vroegere aankooptransacties voor de pas verhuisde klant zouden dan ook ongewenst in verband gebracht worden met diens nieuwe adresgegevens. Een betere oplossing bestaat erin wijzigingen te “stapelen”. Dit komt er in het hoger gegeven voorbeeld op neer dat wij een tweede record voor de klant toevoegen, met diens nieuwe adresgegevens, waarvoor wij tegelijk een nieuwe klantsleutel genereren (zie Figuur 4). Ook kunnen er additionele velden voorzien worden voor de begin- en einddatum van de geldigheidsperiode. De (betekenisloze) klantsleutel identificeert bijgevolg niet langer een bepaalde klant maar een bepaalde klant in een bepaalde periode; dezelfde (betekenisvolle) klantID komt nu dus meermaals voor.

Klant								
Klant-sleutel	KlantID	Begin-datum	Eind-datum	Naam	Adres	Post-code	Gemeente	...
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	Leuven	...

Figuur 4: winkelketenverkoppen; omgaan met veranderingen in klantdimensie

Een andere mogelijke oplossing houdt verband met het onderscheiden van zogeheten “*mini-dimensies*”. Deze laatste ontstaan wanneer men uit een omvangrijke dimensietabel een subset van met elkaar gecorreleerde attributen gaat afscheuren van de oorspronkelijke dimensie. Stel dat men in het gegeven voorbeeld een afzonderlijke tabel zou definiëren met demografische gegevens over klanten (geslacht, leeftijds- en inkomenscategorie, burgerlijke staat, enz.), iedere keer voorzien van een eigen sleutel; vanuit de records in de klanttabel wordt iedere keer verwezen naar deze laatste (zie Figuur 5). Op die manier fungeert demografie als een mini-dimensie van klant. Indien de demografiegegevens van een klant wijzigen, wordt een nieuw record ingevoerd in de demografietabel en wordt de verwijzing in de klanttabel in overeenstemming gebracht met de recentste data. Door nu echter de demografie-sleutelverwijzing niet enkel in de klanttabel maar ook in de centrale verkoptabel onder te brengen (zoals in Figuur 5), kunnen we er bovendien eenvoudig voor zorgen dat de verkoopfeiten uit het verleden gekoppeld blijven aan de toepasselijke oude demografische gegevens.



Figuur 5: winkelketenverkoop; miniaamensie mei rustoriek

3.3. Het datawarehouse-bevoorradingproces

Het bevoorraden van een datawarehouse is de activiteit die ervoor zorgt dat correcte en relevante gegevens volgens afgesproken procedures in een datawarehouse worden binnengebracht en naderhand worden onderhouden. Hierin kunnen de volgende vier deelstappen onderscheiden worden:

- gegevensextractie;
- gegevenstransformatie;
- (bij)laden van het datawarehouse;
- nabewerking.

Gegevensextractie. Onder gegevensextractie wordt het onttrekken van de benodigde gegevens uit een operationele bron verstaan (d.m.v. SELECT-instructies of UNLOAD-utilities). Vaak verloopt het extractieproces asynchroon periodiek, en wordt op gegeven tijdstippen (bv. dagelijks of maandelijks) een databestand aangemaakt dat enkel de incrementele wijzigingen in de brongegevens (i.e., sinds de vorige extractie) omvat. Indien het bronsysteem een tijdsaanduiding voorziet voor de laatste wijziging van een record (*time stamping*), dan kan de extractie hier uiteraard nuttig gebruik van maken. Zo niet, dan zijn andere, complexere methoden nodig, zoals het scannen van logbestanden of het vergelijken van before- and after-images.

Gegevenstransformatie. Het daaropvolgend gegevenstransformatieproces behelst het bewerken van de geëxtraheerde gegevens ter voorbereiding van het laden. Dit veronderstelt volgende activiteiten:

- dataformattering: omvorming naar het vereiste formaat (bv. comma-delimited textfiles versus productspecifieke import-/export-formaten);
- data-opschoning: wegwerken of corrigeren van vervuilde gegevens (o.a. het consistent maken van de toegepaste codes; geslacht zou bv. door elkaar als 'M'/'V', 'man'/'vrouw' of met behulp van een cijfercode weergegeven kunnen zijn);

-
- data-aggregatie en -samenvoeging: indien de fijnmazigheid van de feiten in het datawarehouse minder is dan die in het bronsysteem (bv., winkelverkoop zouden geaccumuleerd kunnen worden op dagbasis i.p.v. per transactie);
 - dataverrijking: aanvullen van de eigen gegevens met externe data.

Wij merken op dat deze activiteiten (deels) voor of na het fysieke gegevenstransport kunnen gebeuren, dus zijnde op het bronplatform of het doelplatform.

(Bij-)laden van het datawarehouse. Vervolgens dienen de in de vorige stappen verkregen gegevensbestanden (in bulk) ingeladen te worden naar de feiten- en dimensietabellen van het datawarehouse. Tijdens het laadproces worden ook de betekenisloze sleutels gegenereerd.

Nabewerking. Na het laden gebeurt het nabewerken van de gegevens. Tijdens dit proces worden er afgeleide gegevens gegenereerd, aggregaattabellen gevuld (bv. tabellen met opgetelde kwartaalgegevens, die direct consulteerbaar zijn), en zo meer.

Het spreekt vanzelf dat de organisatie van dit hele proces en de implementatie en documentatie van de vereiste procedures een zeer omvangrijke onderneming is, die zeker niet mag onderschat worden. Vaak gebeurt de ontwikkeling dan ook volgens een iteratief proces, en wordt gestart met slechts enkele belangrijke subjecten; geleidelijk aan kan de inhoud van het datawarehouse dan verder uitgebreid worden.

3.4. Gebruik van het datawarehouse

Uiteindelijk is een datawarehouse bedoeld om geraadpleegd te worden, en de gebruiker aldus te ondersteunen bij de uitvoering van zijn/haar activiteiten. In heel wat gevallen fungeert het datawarehouse als voornaamste gegevensleverancier voor een beslissingsondersteunend systeem. Hierbij maakt men soms het onderscheid tussen *verificatiegebaseerde* analysemethoden, waarbij men de gebruiker toelaat een hypothese te toetsen op basis van gegevens uit het datawarehouse, en *ontdekkingsgebaseerde* methoden, die tot doel hebben nieuwe, onvermoede kennispatronen aan het licht te brengen.

Verificatiegebaseerde analysemethoden. In het eerste geval moet de gebruiker in de mogelijkheid gesteld worden om de gegevens naar wens te analyseren langs de beschikbare dimensies (bv., door toe te laten de verkoopgegevens te aggregeren over productcategorieën, winkellocaties, enz.). Moderne visuele query- en rapporteringstools ondersteunen dergelijke benadering. Krachtige verwerkingsvormen worden bovendien aangeboden door diverse OLAP-tools (On-Line Analytical Processing).

Ontdekkingsgebaseerde analysemethoden. In het tweede geval is het de bedoeling met behulp van zogeheten *data mining*-technieken nieuwe kennis te ontdekken, bv. onvermoede verbanden tussen demografische gegevens en koopgedrag, in functie waarvan vervolgens specifieke acties kunnen ondernomen worden. Vanzelfsprekend zullen de verkregen resultaten sterk afhangen van de kwaliteit en volledigheid van de in het datawarehouse verzamelde data.

4. Inventaris huidige situatie

Elk van de hoger genoemde partners voert zekere controles of onderzoeken uit binnen het domein van het personen- en goederenvervoer over de weg. De aard en het volume van de daaruit verzamelde data verschilt vanzelfsprekend naargelang de (deels overlappende, deels gemeenschappelijke) bevoegdheden⁴ en de middelen van de specifieke controledienst. Een ander verschilpunt is de technologie waarlangs en de uiteindelijke vorm waarin deze data bijgehouden en toegankelijk gemaakt worden. In wat volgt schetsen wij een overzicht van de huidige situatie.⁵

Een onderscheid dat wij daarbij zullen maken betreft de aard van de controle, meer bepaald dat tussen een *wegcontrole* en een *bedrijfsonderzoek*. Waar het eerste:

- betrekking heeft op een voertuig dat door een (team van) controleur(s) wordt tegengehouden langs de openbare weg;
- beperkt is in duur;
- herleid kan worden tot één (al dan niet geleed) voertuig en tot een zeker bestuurder,

gaat het in het tweede geval om een controle:

- van (en doorgaans op) het bedrijf;
- waarvan de (administratieve) afhandeling zekere tijd in beslag kan nemen;
- die indirect betrekking heeft op meerdere voertuigen en/of werknemers.

4.1. Vervoer te land

Uitgevoerde controletaken. Deze controledienst, die deel uitmaakt van de FOD Mobiliteit en Vervoer, verricht controles inzake wegvervoer van zaken – hierbij in het bijzonder ook het transport van gevaarlijke goederen (cf. ADR) – en van personen, en dat zowel voor het beroepsvervoer als dat uitgevoerd voor eigen rekening. Hij ziet daarbij toe op de naleving van diverse internationale verdragen, EG-verordeningen (onder meer deze op de rij- en rusttijden en het gebruik van de tachograaf, cf. EG 3820/85 en 3821/85) en nationale wetgeving en doet dit zowel aan de hand van wegcontroles als door inspecties in de bedrijven welke eventueel aanleiding kunnen geven tot de intrekking van vervoervergunningen. Hieromtrent beheert hij een databank van vergunde Belgische transportbedrijven, genaamd *Transis* (recent ook raadpleegbaar langs de webstek van de FOD: <http://mobilit.fgov.be>). Tevens is hij de instantie die instaat voor de uitwisseling van data in het kader van de administratieve regeling ‘Euro-Controle-Wegen’ (cf. supra: sectie 2). Verder verzamelt en rapporteert hij ook, op maand-, kwartaal- en jaarbasis, statistieken over de door de eigen dienst verrichte controles en hun resultaten. Daarin aggregereert hij de aantallen controles, gecontroleerde tachograafschijven, vastgestelde inbreuken, PV's en onmiddellijke inningen, volgens een reeks criteria waaronder herkomst van de bestuurder (Belg vs. EU-bewoner vs. niet-EU), het type van vervoer (goederenvervoer vs. personenvervoer, beroepsvervoer vs. eigen vervoer), de aard van de inbreuk, enz. Ten slotte fungeert hij als de initiatiefnemer binnen het Actieplan en het daaruit voortvloeiende Agora-project voor de oprichting van een databank voor het wegvervoer.

⁴ Voor wat betreft de diverse bevoegdhedenpakketten verwijzen wij naar het vademecum dat, eveneens in het kader van het Actieplan van 20 november 2001, hieromtrent opgesteld is.

⁵ Voor wat betreft de partners van Federale en Lokale Politie is vooreerst nader onderzoek naar het wettelijke kader aangaande de mogelijke uitwisseling van (andere dan anoniem gemaakte) data vereist; daarom zijn de (voor een deel nog in ontwikkeling zijnde) toevoersystemen van beide nog niet in dit overzicht opgenomen.

Operationele gegevensbronnen. Indien de controleur een inbreuk vaststelt, maakt deze een formulier op met daarin details aangaande:

- plaats, datum, uur, controleur;
- *werknemer(s)/bestuurder*: naam, nationaliteit, geboortedatum en -plaats, adres, identiteitskaartnummer, rijbewijsnummer;
- werkgever/transporteur: naam en adres;
- *voertuig*: (o.a.) merk, chassis-, plaat- en schouwingsbewijsnummer (iedere keer van trekker en oplegger), vertrekplaats en bestemming, eigenaar;
- *vaststellingen*: bepalingen in overtreding (aangekruist in keuzelijst – aparte formulieren naargelang personen- of goederenvervoer), aantal PV's, eventueel bedrag onmiddellijke in-ning, andere genomen sancties.

Dit formulier wordt echter niet systematisch in een elektronische databank verwerkt. Het overeenkomstig onderdeel binnen de in eigen FOD ontwikkelde *Contris*-software, waarin de controleurs momenteel de verzamelde wetgeving voor hun werkzaamheden kunnen terugvinden, is voorsnog niet in algemeen gebruik, maar kan in de toekomst mogelijk een aanvoerb-ron voor het datawarehouse worden. Echter, zonder voorafgaande fase waarin een dergelij-ke operationele databank ontwikkeld en in gebruik genomen wordt (waarbinnen de controle-data van de Dienst Controle van het Wegvervoer van het Directoraat-generaal Vervoer te Land zelf eerst systematisch gevat kunnen worden), is de implementatie van een gezamenlijk datawarehouse volgens ons moeilijk uitvoerbaar.

De hoger vermelde statistische informatie wordt periodiek aangemaakt in MS Excel. Een tweede, beknopter formulier dient als basis voor de verzameling van de ruwe data. Omdat het daarbij om geaggregeerde cijferinformatie gaat, die dus abstractie maakt van individuele on-dernehmelingen, is dit proces niet van rechtstreeks nut voor het voeden van het datawarehouse. Echter, in de toekomst zou het wel mogelijk moeten worden om, op automatische wijze, de-zelfde statistieken te genereren vanuit het datawarehouse, wat een administratieve verlichting zou kunnen inhouden. Dit zal zeker geen onderdeel uitmaken van een huidige projectfase, maar wij zullen in het databankontwerp wel al rekening trachten te houden met deze mogelijk-ke toekomstige toepassing.

Additionele/externe bronnen. Hogerop haalden we al de vergunningendatabank, Transis, aan. Deze is een waardevolle bron van informatie voor het datawarehouse, omdat zij een vol-ledige lijst verschaft van de Belgische transportbedrijven. Door de opname van BTW-nummers is het bovendien mogelijk de link te leggen met andere gegevensbronnen, waarlangs de data verder verrijkt kunnen worden.⁶ Verder laat zij toe de uit andere diensten afkomstige controlegegevens te valideren; bij het laden van de data zou immers een probleemlijst kunnen geretourneerd worden van gecontroleerde werkgevers waarvoor geen record in de vergunde-transporteurslijst werd teruggevonden. Ten slotte verschaft het informaticasysteem ook toe-gang tot de voor een transporteur ingeschreven voertuigen. De opname van deze nummer-plaatgegevens in het datawarehouse maakt een bijkomende validatie op de invoergegevens mogelijk. Deze data worden verzameld binnen dezelfde FOD maar dan door de Dienst voor Inschrijving van de Voertuigen (D.I.V.).

⁶ Uiteraard denken wij hierbij ook aan de *KruispuntBank van Ondernemingen* (KBO), welke zeker ook haar invloed zal hebben op het Transis-systeem (cf. FOD Economie, KMO, Middenstand en Energie, webstek: <http://mineco.fgov.be>).

4.2. Douane & Accijnzen

Uitgevoerde controletaken. De Administratie der Douane & Accijnzen, die ondergebracht is binnen de FOD Financiën, oefent diverse controletaken uit inzake wegvervoer, zowel door particulieren (ook personenwagens worden gecontroleerd) als door ondernemingen. Het brigadepersoneel verricht daartoe controles langs de weg en van de garagehouders, waarbij volgende bevoegdheidencategorieën centraal staan:

- controle van de goederenbeweging inzake douane, accijnzen en/of BTW;
- controle van de gasolie door monsternamen;
- controle verkeersbelasting en Eurovignet.

Daarbij komen nog een aantal bijkomende (ook niet-fiscale) taken (bv. met betrekking tot drugs, immigratie, ...).

Operationele gegevensbronnen. Aan de hand van een controleverslag worden op het terrein gegevens ingewonnen met betrekking tot:

- plaats, datum, uur, uitvoerende brigade en ploeg;
- *aard van de controle*: statische/dynamische controle, routine/gericht/operatie;
- *werkgever/transporteur*: BTW-nummer (eventueel ook van opdrachtgever);
- *voertuig*: type (personenwagen, lichte vrachtwagen, vrachtwagen, autobus, mobilhome, trekker en/of oplegger), nummerplaat en nationaliteit (apart voor trekker/opleggercombinatie);
- *vaststellingen/gevolg*: bepalingen waarop gecontroleerd werd (aangekruist in een keuzelijst); telkens al dan niet overtreding vastgesteld en (indien toepasselijk) PV opgesteld.

Achteraf worden deze controleverslagen, op de brigade zelf, elektronisch verwerkt en vervolgens maandelijks doorgestuurd naar de C.D.I.B (Centrale Dienst Informatiebeheer), waar de data gebruikt worden voor de opvolging van de brigades. Deze databank is geïmplementeerd in MS Access. Wij merken op dat de identiteit van de bestuurder niet geregistreerd wordt.

Additionele/externe bronnen. De dienst beschikt over externe ondernemingsinformatie afkomstig van de centrale administratie van de BTW, en over deze, verzameld in de *Bel-first* toepassing van Bureau van Dijk. Hij heeft ook (beperkt) toegang tot de nummerplaatgegevens van de D.I.V.

4.3. Rijksdienst voor Sociale Zekerheid (R.S.Z.)

Uitgevoerde controletaken. De R.S.Z. is de beheerder van een databank van alle Belgische werkgevers, het zogeheten werkgeversrepertorium. De inspectiediensten van de R.S.Z. hebben dan ook in hoofdzaak een administratieve controletaak ten aanzien van de correcte aangifte van werknemers en lonen door werkgevers (inclusief dus deze in de transportsector); doorgaans gebeurt een dergelijk onderzoek op verzoek van de eigen binnendiensten, soms op verzoek van een externe instantie (o.a. ook I.S.W. en S.I.). Momenteel verricht de R.S.Z. zelf (nog) geen wegcontroles. Enkele voor het datawarehouse potentieel relevante categorieën van uitgevoerde onderzoeken zijn:

-
- onderwerpingonderzoek (werknemers): echte of schijnzelfstandige;
 - onderwerpingonderzoek (lonen): kostenvergoedingen als verdoken lonen;
 - niet-indiening van aangifte (“stiltwijgers”).

Operationele gegevensbronnen. In de centrale administratie te Brussel bevindt zich een Oracle-/Powerbuilder-databank en -toepassing, genaamd *DBEO*, waarmee de onderzoeksaanvragen en hun (in meerderheid administratieve) afhandeling opgevolgd worden. Volgende gegevens worden daarin bijgehouden:

- opstart- en afsluitdatum, uitvoerend controleur;
- *aard van de controle*: oorsprong (bv. instantie van wie vraag uitging);
- *werkgever/transporteur* (gekoppeld aan het werkgeversrepertorium): RSZ-nummer, naam, adres en sociaal secretariaat;
- *vaststellingen*: inhoud van het onderzoek (voorbeelden: zie hoger) en gevolg (o.a. aangifte, wijzigend bericht – eventueel met vermelding van het bedrag –, pro-justitia, waarschuwing, aanvullend onderzoek).

Op het niveau van de individuele controleur is er ten slotte een Lotus Notes-applicatie ter ondersteuning van diens controleactiviteiten, de opmaak van verslagen en dossierbeheer; dit laatste heeft geen verder belang voor het datawarehouse.

Additionele/externe bronnen. Het *werkgeversrepertorium van de R.S.Z.* is een belangrijke databank waarin de basisidentificatiegegevens wat betreft de werkgever worden bewaard, alsook de aanduiding van de werkgeverscategorie waartoe deze behoort. Deze informatie wordt door de werkgever meegedeeld bij zijn inschrijving bij de RSZ; ook de historiek van wijzigingen wordt bijgehouden. Dit werkgeversrepertorium vormt een belangrijke gegevensbron voor het netwerk beheerd door de Kruispuntbank van de Sociale Zekerheid (KSZ). Zowel de I.S.W. als de S.I. maken er gebruik van bij de uitoefening van hun onderzoeksactiviteiten. Wij merken op dat ook het uniek ondernemingsnummer opgenomen werd in het repertorium, zodat de link met de Kruispuntbank van Ondernemingen (KBO) kan gelegd worden.

Een tweede voornamelijk gegevensbron betreft de DIMONA-aangifte en het daaruit resulterend elektronisch personeelsregister. De DIMONA-aangifte heeft als doel het begin en het einde van een arbeidsrelatie onmiddellijk op elektronische wijze aan te geven aan de R.S.Z. DIMONA staat voor Déclaration Immédiate/Onmiddellijke Aangifte en maakt deel uit van de portaal-site van de Sociale Zekerheid, www.sociale-zekerheid.be. De aangifte bevat gegevens over de werkgever zelf, over de werknemer (waar mogelijk geïdentificeerd aan de hand van zijn/haar INSZ – het unieke identificatienummer voor de sociale zekerheid), de plaats van tewerkstelling van een student, het nummer van het paritair comité en de datum waarop de werknemer in en/of uit dienst treedt. Dankzij DIMONA worden bepaalde verplichtingen in hoofde van de werkgevers inzake sociale documenten vereenvoudigd of afgeschaft (o.a. afschaffing van het papieren personeelsregister). De verzamelde DIMONA-databank is uiteraard ook consulteerbaar door de controleurs van de R.S.Z., I.S.W. en S.I. Systematisch nagaan of de bij een wegcontroleactie tegengehouden bestuurders wel als werknemer gemeld zijn in het DIMONA-systeem is een van de typische taken die achteraf uitgevoerd worden door de mensen van de I.S.W. en S.I. (cf. infra).

Een interessante additionele externe bron waarover de R.S.Z., net als de I.S.W. en S.I., ook toegang krijgt is de zogeheten *Kruispuntbank voor Sociaal Handhavingsrecht*. Hierlangs kan men gegevens over strafvervolgingen inzake arbeidsrecht raadplegen. Dit systeem moet aan de controleurs toelaten op te volgen hoe de overgemaakte dossiers verder afgehandeld worden

door de arbeidsauditeurs. Wij willen in dat verband benadrukken dat dergelijke informatie in geen van de in deze en volgende secties beschouwde operationele gegevensbronnen voorkomt: men beperkt zich iedere keer tot de vermelding “pro-justitia”.

4.4. Inspectie van de Sociale Wetten (I.S.W.)

Uitgevoerde controletaken. De dienst Inspectie van de Sociale Wetten (I.S.W.) valt onder de FOD Werkgelegenheid, Arbeid en Sociaal Overleg (het vroegere Federaal Ministerie van Tewerkstelling & Arbeid), waar zij een van de drie arbeidsinspectiediensten is. Zij verricht controles in het domein van het arbeidsrecht. De I.S.W. omvat een hoofdbestuur en 24 externe directies (13 in het Vlaamse Gewest, 10 in het Waalse Gewest en 1 in het Brussels Hoofdstedelijk Gewest). Twee directies daarvan hebben specifiek opdracht gekregen zich op de transportsector te richten, namelijk die van Mechelen en Huy. Enkele typische inbreuken die door de dienst vastgesteld kunnen worden, hebben betrekking op: zwartwerk (niet-melding aan DIMONA (zie hoger), of in het zwart uitbetaalde prestaties), gebrekkige uitbetaling van loon, feestdagen of verplaatsingsvergoedingen, het onjuist bijhouden van de prestatiebladen, het onjuist opmaken van de individuele rekeningen, tewerkstelling van werknemers van vreemde nationaliteit zonder verblijfsvergunning of arbeidskaart/-vergunning, en – een bevoegdheid die zij deelt met de meeste andere diensten – de niet-naleving van de Europese verordeningen inzake de rij- en rusttijden of het incorrect gebruik van de tachograaf.

Operationele gegevensbronnen. De volgende gegevens worden verzameld:

- opstart- en afsluitdatum, inspecteur;
- *aard van de controle*: motief, oorsprong, protocol;
- *werknemer(s)/bestuurder*: niet direct beschikbaar (cf. infra); wél de aantallen bij het onderzoek betrokken mannen/vrouwen, werklieden/bedienden, deeltijdse werknemers, enz.;
- *werkgever/transporteur*: RSZ-nummer, NACE-code, naam, adres, rechtsvorm, enz.;
- *voertuig*: niet direct beschikbaar (cf. infra); wél het aantal gecontroleerde voertuigen (1 voor een wegcontrole) en het aantal nageziene schijven;
- *vaststellingen*: voor elk dossier worden de afzonderlijke bezoeken en hun resultaten geregistreerd, met vermelding van de plaats (onderneming, sociaal secretariaat of andere – onder deze laatste categorie vallen de wegcontroles), code van de wettelijke bepaling(en) waarop gecontroleerd werd, het gevolg (waarschuwing, regularisatie, pro-justitia, geen onregelmatigheden of doorzending naar andere directie) en, waar toepasselijk, het regularisatiebedrag of het PV-nummer.

Wij merken daarbij op dat, in de gegevensbank, momenteel geen afzonderlijke velden voorzien zijn die zouden toelaten een langs de weg gecontroleerd bestuurder en diens voertuig automatisch te identificeren: naam, INSZ (of rijksregisternummer), nummerplaat en alerhande extra gegevens worden typisch wel genoteerd door de controleur ter plekke, maar kunnen later hoogstens in een commentaartekstveld ingebracht worden en zijn dus niet direct extra-heerbaar uit de uiteindelijke databank (deze laatste is immers niet uitsluitend op het transportdomein toegespitst).

Het huidige systeem is geïmplementeerd in Omnis (vroeger van het Engelse Blythe Software, nu bij Raining Data; zie <http://www.omnis.net>), een geïntegreerde database- en applicatie-ontwikkelingsomgeving, die ook de nodige extractiefaciliteiten voorziet. Een moeilijkheid is dat de gegevens momenteel niet gecentraliseerd worden in één grote databank; er worden wel halfjaarlijks bestanden op tape ingezameld vanuit de diverse lokale directies. Daarom wordt voorgesteld om in een eerste datawarehousing-fase enkel de gegevens van Mechelen en Huy

op te nemen. In een tweede fase zou dan een systeem opgezet kunnen worden waarbij de voor het datawarehouse relevante databestanden periodiek onttrokken worden uit alle directies, om vervolgens elektronisch doorgezonden te worden naar het hoofdbestuur, waar zij dan samengevoegd zouden worden voor doorverzending naar het Directoraat-generaal Vervoer te Land.

Additionele/externe bronnen. De I.S.W. maakt gebruik van het werkgeversrepertorium van de R.S.Z., de DIMONA-databank en heeft ook toegang tot Kruispuntbank voor Sociaal Handhavingsrecht; deze systemen werden hogerop besproken in sectie 4.3.

4.5. Sociale Inspectie (S.I.)

Uitgevoerde controletaken. De Sociale Inspectie maakt deel uit van de FOD Sociale Zekerheid en heeft tot taak toe te zien op de naleving van de wetgeving inzake de sociale zekerheid voor werknemers. Met het oog daarop doet zij onderzoeken bij werkgevers, werknemers en instellingen van de sociale zekerheid. Specifiek voor wat betreft de transportsector, voert zij ook wegcontroles uit. Haar controlebevoegdheden omvatten volgende categorieën:

- sociale zekerheidsbijdrage: correcte aangifte werknemers (DIMONA-melding) en bnen; onderzoek schijnzelfstandigen;
- toekenning jaarlijkse vakantie en uitbetaling vakantiegeld;
- niet-afgifte documenten inzake ziekte- en invaliditeitsverzekering;
- verplichtingen inzake kinderbijslag;
- bijhouden en bewaren van de sociale documenten (o.a. personeelsregister, aanwezigheidsregister, individuele rekeningen);
- arbeidsongevallen: verzekering en aangifte arbeidsongeval;
- deeltijds werk: openbaarmaking en navolging van het werkrooster;
- illegale tewerkstelling/verblijf van werknemers van vreemde nationaliteit;
- verhinderen van toezicht.

Wij merken op dat tachograafcontroles hier niet toe behoren.

Operationele gegevensbronnen. Hoewel op het terrein, en in het bijzonder bij wegcontroles, additionele gegevens kunnen worden ingewonnen, onder meer over de bestuurder en het voertuig, is de huidige databank (SIS) in hoofdzaak afgestemd op de opvolging van onderzoeksopdrachten op werkgeverniveau. Van een individuele betrokkene worden eventueel wel naam, rijksregisternummer en adres bijgehouden, maar voertuigdata vinden wij bv. niet terug. De werkgever die het voorwerp uitmaakt van een onderzoek is zowel identificeerbaar op basis van RSZ- als BTW-nummer. De werkgevergegevens worden voorts aangevuld met officiële, uitbating- en contactgegevens; bovendien wordt opgenomen bij welke fondsen, verzekeringsmaatschappij, sociaal secretariaat, enz. de werkgever aangesloten is. Wat betreft de aard van het onderzoek wordt onder meer aangegeven of deze op verzoek van een andere instantie (zo ja, welke), als reactie op een klacht of op eigen initiatief ondernomen werd, door welke cel zij uitgevoerd werd/wordt (bv., de transportcel van een bepaald bureau), en, waar toepasselijk, binnen welk protocol (bv., “transport”, nu voor wat specifiek de wegcontroles betreft).

De bij het onderzoek gedane vaststellingen en hun gevolgen worden als volgt gecodeerd. Een onderscheid wordt gemaakt tussen een aantal hoofdcategorieën (deze in voorgaande subsectie: uitgevoerde controletaken), en daarbinnen iedere keer tussen een aantal subcategorieën. Voor elk van de hoofdcategorieën wordt iedere keer aangegeven: 1 (nagezien en in orde); 2 (nagezien en regularisatie doorgevoerd of verwittiging); 3 (nagezien en PV opgesteld). Voor

vaststellingen van type 2 en 3 wordt dan, naargelang de precieze detailcategorie, het volgende geregistreerd: ja/nee; het regularisatiebedrag; het aantal betrokken personen; een bepaalde combinatie van voorgaande elementen. Ten slotte worden nog de diensten opgesomd die op de hoogte gesteld zijn.

Wij merken op dat momenteel een nieuwe systeemversie in ontwikkeling is, die geïmplementeerd wordt in Oracle/Business Objects. Hierover moet echter nog bijkomende informatie ingewonnen worden.

Additionele/externe bronnen. De externe gegevensbronnen waartoe de S.I. toegang heeft, zijn soortgelijk aan die van de I.S.W. (zie hoger).

4.6. Federale en Lokale Politie

Voor wat betreft de partners van federale en lokale politie is in de eerste plaats nader onderzoek naar het wettelijke kader aangaande de mogelijke uitwisseling van (andere dan anoniem gemaakte) data vereist. De beoogde doelstellingen van het datawarehouse veronderstellen immers dat zekere identificatiegegevens (bv. nummerplaat) doorgegeven zouden worden. Daarom worden de (voor een deel nog in ontwikkeling zijnde) systemen van federale en lokale politie hier nog niet in detail opgenomen. Het spreekt echter vanzelf dat, voor beide, het PV-register een nuttige gegevensbron zou kunnen vormen voor bepaalde categorieën van vaststellingen aangaande het goederen- en personenvervoer over de weg. Wij denken dan bijvoorbeeld aan verbalisatiegegevens over tachograaf, technische controle, lading, ADR en rijbewijs.

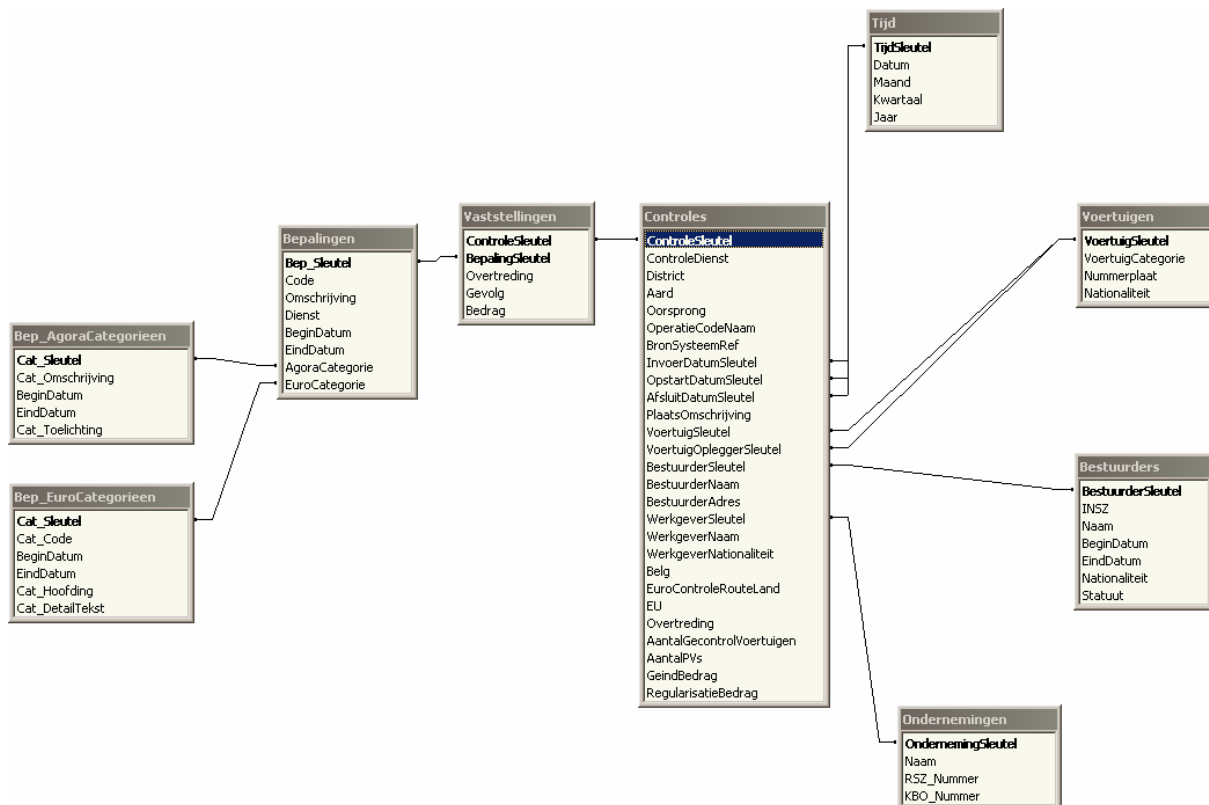
5. Ontwerp van model en toelichting

Ondanks het feit dat sommige bronsystemen niet specifiek gericht zijn op het transportdomein en dan ook weinig voertuig- of bestuurderinformatie kunnen verschaffen, zijn wij niettemin van mening dat de administratieve overlast die het manueel transfereren van dergelijke gegevens met zich meebrengt, niet lijkt op te wegen tegen de vollediger gegevens die aldus verkregen zouden kunnen worden. Wij zouden daarom willen kiezen voor een systeem waarbij het datawarehouse in hoofdzaak gevoed wordt door automatische extractie uit een aantal van de in sectie 4 besproken bronsystemen (enkel voor wat betreft de controlegegevens van het Directoraat-generaal Vervoer te Land en deze aangeleverd in het kader van de administratieve regeling 'Euro-Controle-Wegen'-regeling, zal manuele invoer – of in het eerste geval, de verdere uitbouw van een databanksysteem – wellicht (voorlopig) vereist zijn). Wij trachten in ons modelvoorstel echter maximaal rekening te houden met de mogelijke toekomstige uitbreiding van de beschikbare data in de operationele bronsystemen.

Een andere zorg is het vermijden van onnodige redundantie ten opzichte van de bestaande systemen. Zo beschikken de afzonderlijke diensten elk al over een uitgebreide verzameling ondernemingsdata, die relevant zijn voor hun werkzaamheden. Al deze gegevens dupliceren in het datawarehouse, lijkt ons niet aangewezen. Liever stellen wij voor de nodige gegevensvelden te voorzien die toelaten een entiteit (onderneming, voertuig, bestuurder) eenduidig te identificeren in de huidige systemen, samen met die die gewenst zijn als selectiecriteria voor de opvragingen en analyses, en zo onnodige redundantie te vermijden.

Het voorgestelde ontwerpmodel is in grote lijnen gebaseerd op de in sectie 3 besproken methoden. Het onderscheidt twee centrale feitentabellen, respectievelijk voor de uitgevoerde *controles* en hun resultaten (*vaststellingen*). Daarrond situeren zich een aantal tabellen

voor de dimensies *tijd*, *voertuig*, *bestuurder*, *onderneming* en (wettelijke) *bepaling*, en – hiërarchisch gelinkt aan deze laatste dimensietabel (cf. sneeuwvlokschema, pag. 8) – de tabellen *Bep_AgoraCategorieen* en *Bep_EuroCategorieen*. De hieruit volgende relatie-structuur is afgebeeld in Figuur 6 (deze toont een schermafdruk voor een eenvoudig prototype opgesteld in MS Access, zie bestand in bijlage: ‘AgoraDB.mdb’).



Figuur 6: ontwerpvoorstel datawarehouse; relatiestructuur

Laten wij enkele van de gemaakte keuzes verder toelichten, en dan in eerste instantie die die enigszins afwijken van het in sectie 3 geschetste methodologisch kader. Een vollediger specificatie van de tabellen en hun attributen is opgenomen in bijlage (zie Appendix A: bestand ‘Agora-01-082_Appendix A_DW-Model.xls’).

Controles vs. vaststellingen. De registratie van de uitgevoerde controles en onderzoeken wordt niet gebundeld langs één centrale feitentabel, maar is uitgesplitst over twee tabellen, *controles* en (de daarbij gedane) *vaststellingen*. Voorbeeld van een controle zou kunnen zijn: bestuurder x in voertuig y en werknemer van de onderneming z werd door een controleur van dienst d tegengehouden op datum t . Voor elk van de wettelijke bepalingen waarop daarbij gecontroleerd werd (en/of waarop een inbreuk werd vastgesteld, cf. infra), komt er een rij in de vaststellingentabel, met vermelding van de afhandeling. Zonder te kiezen voor een dergelijke genormaliseerde oplossing en de feitentabel dus uit te splitsen, zouden voor elke vaststelling die tijdens een dergelijke controle gedaan wordt, iedere keer de tijdsleutels, de bestuurder-, voertuig- en werkgeversleutel, net als een aantal andere niet-sleutelgegevens, moeten herhaald worden. Bovendien zijn heel wat analyses denkbaar die niet tot op het niveau van de individuele vaststellingen hoeven te gaan, maar slechts tot op het niveau van de uitgevoerde controles. Om vervolgens het 1-op- n verband tussen beide resulterende tabellen te modelle-

ren, definiëren wij in de controletabel een betekenisloze primaire sleutel, “ControleSleutel”, waarnaar wij in de vaststellingentabel verwijzen.

Bedrijfsonderzoeken vs. wegcontroles; verband met werknemers/voertuigen. Om de structuur zo eenvoudig mogelijk te houden, stellen wij voor om (zoals in de meeste bronsystemen eveneens het geval is) wegcontroles en bedrijfsonderzoeken in dezelfde feitentabel onder te brengen. Deze bevat daarom een veld, ‘Aard’, dat aangeeft om welk van beide typen het gaat. Voor wegcontroles wordt de mogelijkheid voorzien om de betrokken bestuurder en diens voertuig te identificeren. Echter, bedrijfsonderzoeken hebben in regel betrekking op meerdere werknemers en/of voertuigen. In principe zijn beide toewijsbaar aan specifieke vaststellingen (i.e., men zou voor elke vastgestelde overtreding kunnen aanduiden om welke individuele werknemer en/of voertuig het gaat). In de praktijk echter is deze informatie niet beschikbaar in de bronsystemen. Enkel over de aantallen betrokken werknemers en voertuigen vinden we zekere informatie terug. De I.S.W. en S.I. registreren zowel het totaal aantal betrokken werknemers (verder ingedeeld volgens statuut), als de deelaantallen per vaststelling. Het totaal aantal gecontroleerde voertuigen (dus niet per vaststelling) vinden we dan weer enkel terug in de databanken van de I.S.W. Dat maakt dat we ten hoogste zouden kiezen voor volgende oplossing, i.e. – waar beschikbaar –: het totaal aantal gecontroleerde voertuigen en werknemers ter hoogte van de “controles”-tabel, en, meer specifiek, telkens ook het aantal betrokken werknemers bij de individuele vaststellingen. Indien in de toekomst meer gedetailleerde data beschikbaar zouden worden, kan het ontwerp later alsnog bijgestuurd worden om hier gebruik van te maken.

Dimensie ‘onderneming’. Zoals besproken in sectie 3.2, wordt tekstuele informatie in principe ondergebracht in de dimensietabellen. In bepaalde gevallen wordt duplicatie in de feitentabel toegestaan, bv. omwille van efficiëntieredenen of als oplossing voor de historische modellering van dimensiewijzigingen. In ons geval stelt zich echter een bijkomend probleem voor wat betreft de afbakening en de unieke identificatie van de dimensietabelpopulaties. Neem bijvoorbeeld de populatie van transportondernemingen. Belgische ondernemingen kunnen we eenduidig identificeren aan de hand van BTW- (of KBO-) nummer, of eventueel ook aan de hand van het RSZ-inschrijvingsnummer. De ondernemingsnaam leent zich hier veel minder toe omwille van het risico van inconsistente noteerwijzen. Voor buitenlandse ondernemingen echter zijn deze identificatiemiddelen niet voorhanden; sommige diensten hoeven zich ook uitsluitend te richten op Belgische werkgevers. Niettemin veronderstelt de ‘Euro-Controle-Wegen’-rapportering dat vastgestelde (zware) overtredingen van buitenlandse wegvervoerders worden doorgegeven met vermelding van de naam van de onderneming. Liever dan alle ondernemingen als rijen in de dimensietabel trachten onder te brengen (met het gevaar op ongewenste dubbele voorkomens en bijgevolg foutieve analyses – de meerderheid hiervan kan bovendien enkel nuttig gebruikt worden voor de Belgische ondernemingen), stellen wij voor om enkel een welgedefinieerde subset van controleerbare ondernemingen onder te brengen in de ondernemingen-dimensietabel (namelijk de vergunde transportondernemingen zoals deze opgenomen zijn in de transis-databank), en voor de andere enkel naam en nationaliteit rechtstreeks in de Controles-feitentabel op te nemen, dit laatste vooral voor rapporterings-, niet voor analysedoeleinden.⁷ In de toekomst kan, bijvoorbeeld voor onderzoeksdoeleinden, verder de link gelegd worden via het ondernemingsnummer naar diverse externe bronnen van bedrijfsinformatie.

⁷ Nog een mogelijke oplossing zou zijn een aanvullende dimensietabel te voorzien voor deze restcategorie van ondernemingen.

Dimensie ‘bestuurder’ (m.b.t. wegcontroles). Een analoog probleem stelt zich voor de gecontroleerde bestuurders. Voor werknemers die onder het Belgische sociale zekerheidsstelsel vallen, kan het INSZ (wat voor de in België wonende personen overeenstemt met het rijksregisternummer) in de toekomst fungeren als identificatiemiddel. Voor werknemers van buitenlandse transportbedrijven gelden uiteraard andere identificatiewijzen. Daar komt nog bij dat niet alle diensten het INSZ of het rijksregisternummer (mogen) noteren, meer nog, dat momenteel geen van de diensten dit gegeven automatisch kan aanbieden vanuit de beschikbare operationele databank. Ook de bescherming van de privacy is een bijkomend aspect dat nog onderzocht dient te worden in dit verband. Met ook de gebruiksdoeleinden van het datawarehouse in het achterhoofd (waar de analyse van ondernemingsgedrag belangrijker lijkt dan dat van individuele werknemers), stellen wij voor om, in een eerste fase, te starten zonder gebruik te maken van de bestuurders-dimensietabel, en dit onderdeel pas later verder uit te bouwen zodra de praktische en wettelijke bezwaren verholpen zijn. Om aan de eisen van de ‘Euro-Controle-Wegen’-rapportering te voldoen, voorzien wij wel al twee velden voor de naam en het adres van de gecontroleerde bestuurder in de feitentabel zelf. Zij hoeven niet later aangepast te worden, maar geven de toestand weer op het moment van de controle.

Dimensie ‘voertuig’ (m.b.t. wegcontroles). Voor wegcontroles zijn twee velden voorzien in de controles-feitentabel die verwijzen naar het gecontroleerde voertuig (twee dergelijke velden zijn nodig voor trekker/opleggercombinaties). De unieke identificatie hiervan kan gebeuren op basis van nummerplaat (en nationaliteit). Voor wat betreft de in België ingeschreven voertuigen kan deze dimensietabel gevuld worden op basis van de nummerplattendatabank van de D.I.V., en vervolgens periodiek verder aangevuld (geen verwijderingen; cf. bewaren van de controlehistoriek). Dit laat toe de aangeleverde controledata, waar deze nummerplaat-informatie bevat, te valideren. Voor elk niet eerder gecontroleerd buitenlands voertuig kan iedere keer een rij toegevoegd worden. Voor wat betreft de data van I.S.W. en S.I. moet bekeken worden of een systeem zou kunnen opgezet worden om ook voertuigidentificatiegegevens systematisch door te geven.⁸

Dimensie ‘tijd’. Voor bedrijfsonderzoeken wordt de begin- en einddatum opgenomen; voor wegcontroles zijn beide identiek. Om de bevraging te vergemakkelijken worden in de tijdtabel enkele redundante velden opgenomen voor maand, kwartaal en jaar. Merk op dat langdurige bedrijfsonderzoeken een additionele extractiemoeilijkheid met zich mee brengen: onderzoeken die nog niet afgesloten werden, kunnen in de volgende periode wijzigingen ondergaan hebben (bv. nieuwe vaststellingen door bijkomende bezoeken). De extractieprocedures moeten dit in rekening nemen, en kunnen dus niet louter de in de nieuwe periode toegevoegde/gewijzigde records binnenhalen zonder bepaalde wijzigingen aan eerder ingelezen records. Tenzij men zou besluiten de data pas in te laden na beëindiging van het onderzoek, wordt hier dus noodgedwongen afgeweken van het principe dat bij voorkeur enkel nieuwe feiten toegevoegd worden aan een datawarehouse, en geen bestaande gewijzigd.

Dimensie ‘Bepalingen’. Elke dienst gebruikt een eigen codeerschema voor de wettelijke bepalingen waarop gecontroleerd wordt. Om de resultaten begrijpelijk te maken en om analy-

⁸ Meer algemeen willen wij erop wijzen dat de niet- of ongestructureerde opname van een aantal belangrijke identificatiegegevens in de bronsystemen maakt dat het datawarehouse, in de huidige context, onvermijdelijk slechts onvolledige data zal bevatten. Het zou daarom aanbeveling verdienen dat alle partners in de mate van het mogelijke, en binnen de grenzen van hun specifieke bevoegdheden, ernaar zouden streven sommige van deze additionele identificatiedata (bijvoorbeeld het INSZ van een gecontroleerd bestuurder en de nummerplaatgegevens) in een systematische vorm op te nemen in hun toekomstige systeemversies.

ses mogelijk te maken inzake inbreuken waarover de diensten gedeelde bevoegdheden hebben, is het aangewezen om deze schema's te mappen naar een gezamenlijk classificatieschema. Niettemin is het wellicht nuttig ook de dienstspecifieke coderingen te behouden, vooral uit flexibiliteitoverwegingen: de indeling kan zo eenvoudig aangepast worden zonder breuk met de oudere data, of additionele classificaties kunnen voorzien worden. In de tabel 'bepalingen' worden daarom de dienstspecifieke wettelijke bepalingen bijgehouden, waarnaar verwezen wordt in de vaststellingentabel.⁹ Twee velden in de bepalingentabel verwijzen dan door naar twee soorten classificaties die wij momenteel nuttig zien voor het datawarehouse, namelijk de codering gebruikt in het kader van de administratieve regeling 'Euro-Controle-Wegen'-regeling (zie tabel 'Bep_Eurocategorieen'), en een nieuwe classificatie voor het datawarehouse-project zelf (zie tabel 'Bep_Agoracategorieen').

'Bep_Eurocategorieen'. Deze eerste tabel bevat de codes voorgeschreven voor de uitwisseling van informatie tussen 'Euro-Controle-Wegen'-landen.

'Bep_Agoracategorieen'. Bovenstaande classificatie bestrijkt zeker niet het hele spectrum van controles waaraan Belgische transportondernemingen onderworpen worden door de partners in het datawarehouse-project, vandaar de behoefte aan een bijkomende classificatie, op basis waarvan een ruimer beeld zou kunnen gegeven worden van het soort overtredingen dat vastgesteld wordt en de frequentie ervan. Op basis van de informatie die wij hebben ingewonnen bij de diverse diensten, zouden we hiertoe volgende initiële "Agora"-classificatie willen voorstellen:

1. Ontbreken vervoervergunning;
2. Vrachtbrief/ vervoerdocument;
3. (Over-)lading;
4. A.D.R.-reglementering;
5. Uitzonderlijk vervoer: vergunning;
6. Personenvervoer: attest eigen vervoer;
7. Personenvervoer: communautaire vergunning/ kwaliteitscontrole/ exploitatietoelating/ ritblad;
8. Technische controle/ schouwingsbewijs;
9. Tachograaf: rij-/rusttijden & onderbrekingen;
10. Tachograaf: gebruik toestel;
11. Wegcode;
12. Rijbewijs/ geneeskundige schifting;
13. Bestuurdersattest onderdanen derde landen;
14. Inschrijving voertuig/ nummerplaat;
15. Verzekering;
16. Verkeerstaks;
17. Eurovignet;
18. Controle van de gasolie;
19. Douanedocumenten: TIR-carnet/ T-documenten/ begeleidingsdocument;
20. Accijnzen/ B.T.W.;
21. DIMONA-aangifte begin/einde arbeidsrelatie;
22. RSZ-aangifte lonen (DMFA);
23. Onderwerping: schijnzelfstandigen/ verdoken lonen;

⁹ Omtrent het onderhoud hiervan (i.e., het doorgeven van nieuw ingevoerde codes) moeten dan wel duidelijke afspraken gemaakt worden met elk van de partners.

-
24. Arbeidsreglement: opmaken/ bijhouden;
 25. Deeltijds werk: uurrooster;
 26. Bijhouden sociale documenten: personeelsregister/ aanwezigheidsregister/ individuele rekeningen/ ...;
 27. Betaling van loon en div. vergoedingen;
 28. Feestdagen: betaling/ inhaalrust;
 29. Jaarlijkse vakantie/ vakantiegeld;
 30. Arbeidsongevallen: verzekering/ aangifte;
 31. Illegale tewerkstelling/verblijf vreemde arbeidskrachten;
 32. Ziekte- en invaliditeitsverzekering;
 33. Kinderbijslag;
 34. Verhinderings toezicht;
 35. Taaldecreet;
 36. Andere.

Bij hun respectieve aansluiting op het systeem, zouden de deelnemende diensten deze lijst moeten valideren en indien nodig verder aanpassen naar gelang van de geëvolueerde behoeften. Een belangrijke oefening die elke partner daarbij ook moet maken is voor elke individuele bepaling (zoals die gebruikt wordt in de eigen databank) aan te geven bij welke categorie deze behoort. Deze mapping-informatie kan dan vervolgens ingebracht worden in de Bepalingentabel van het datawarehouse (zoals geïllustreerd in volgende sectie). Een ander element hier betreft de eventuele inbouw van filtermechanismen die enkel gedetailleerde informatievoorziening toelaat voor de materies waartoe elke afzonderlijke dienst bevoegd is.

6. Proof-of-concept oefening: verwerking douanegegevens

Ter illustratie van het voorliggende datawarehousing-voorstel werd een (eenmalig) ETL-proces (extractie, transformatie, laden) geïmplementeerd voor een van de deeltrajecten, meer bepaald voor een deelset van de door Douane & Accijnzen vergaarde controlegegevens (cf. sectie 4.2). Dergelijke oefening kan immers ook bijkomende inzichten opleveren over de moeilijkheidsgraad van het hele opzet of onvoorziene problemen met de brondatabanken aan de oppervlakte brengen. Bovendien is het ook nuttig om een eenvoudige gebruikersinterface te ontwikkelen voor de consultatie van de resultaatgegevens. Een dergelijk demonstratiehulpmiddel stelt de eindgebruiker immers in staat zich een beter idee te kunnen vormen van de query- en analysemogelijkheden en beperkingen van het finale datawarehouse.

6.1. ETL-implementatie

Voor deze opdracht gebruikten wij als testset de in het eerste kwartaal van 2003 verzamelde wegcontrolegegevens. In Tabel 1 tonen wij de lijst van beschikbare gegevensvelden in de brontabel. Tabel 2 geeft de gebruikte codering (cf. velden 'CONTROLESOORT' en 'VASTSTELLING', en mapt deze tevens naar de in sectie 5 voorgestelde Agora-classificatie (cf. volgnummers) en naar de 'Euro-Controle-Wegen'-codering (met dat laatste bestaan maar weinig raakpunten: enkel D03 – "Absence of technical inspection" kon gelinkt worden). Aangezien het vooralsnog niet de bedoeling was om een in de tijd te herhalen procedure te implementeren, werd geen mechanisme voorzien om bij een eventuele latere extractie slechts de recentere/gewijzigde gegevens binnen te halen. Deze brongegevens werden aangeleverd in MS Access-formaat, en worden na uitgebreide bewerking in het eerder vermelde prototype (eveneens in MS Access) ingeladen. De ETL-code zelf werd ontwikkeld met behulp van Borland Delphi 7 (zie: <http://www.borland.com/delphi/>), vermits wij daartoe geen special-purpose tool ter beschikking hadden.

Attribuut	Type	Omschrijving
(Weg-)controles douanebrigaden		
VOLGNUMMER	Getal	Controlevolgnummer in de maand (per forscodel)
FORSODE	Tekst	Forscode van de brigade
INPUTDATUM	Datum	Datum waarop het controleverslag ingevoerd werd
CONTROLEDATUM	Datum	Datum waarop de controle uitgevoerd werd
UUR	Tekst	Uurindicatie uitgevoerde controle
PLOEG	Tekst	Nummerplaat van het dienstvoertuig of roepnaam van het vast team
CONTROLEPUNT	Tekst	Omschrijving locatie v/d controle (b.v., "Gent-Brugge E17")
IMPORTEUR	Tekst	BTW-nummer importeur/bestemming*
EXPORTEUR	Tekst	BTW-nummer exporteur/verzender*
TRANSPORTEUR	Tekst	BTW-nummer transporteur*
VOERTUIG	Tekst	Gestructureerde informatie m.b.t. het gecontroleerde voertuig; ;type1/nummerplaat1/nationaliteit1;type2/nummerplaat2/nationaliteit2 (types: 1: personenwagen; 2: lichte vrachtwagen; 3: vrachtwagen; 4: autobus; 5: mobilhome; 6: trekker; 7: oplegger; 8: andere)
NATIONALITEIT	Tekst	Omschrijving nationaliteit*
CONTROLETYPE	Getal	Controletype: 1: dynamische controle (interceptie v/h voertuig); 2: statische controle 1 (langs de weg; selectie); 3: statische controle 2 (langs de weg: alle)
CONTROLEAARD	Getal	Controleaard: 1: routinecontrole; 2: gerichte controle (op een specifieke bepaling, cf. infra); 3: (evt. gecoördineerde) operatie
CODENAAM	Tekst	Detailinfo controleaard: bepaling gerichte controle of operatiecodenaam (b.v. 210 of "ZW VERV LOKPOL ZOTTEGEM")
CONTROLESOORT	Tekst	Codelijst bepalingen waarop gecontroleerd werd (elke deelcode wordt vooafgegaan door een puntkomma, b.v.: ";210;320;340"); zie elders voor een overzicht van de gehanteerde codes
VASTSTELLING	Tekst	Codelijst bepalingen waarop een overtreding vastgesteld werd (moet een deellijst zijn v/d bovenstaande lijst)
A115	Tekst	Detailinfo bepaling 115 (transport accijnsgoederen - andere: ...)
A122	Tekst	Detailinfo bepaling 122 (invoerrechten - geviseerde goederen: ...)
A132	Tekst	Detailinfo bepaling 132 (BTW-controle - geviseerde goederen: ...)
A145	Tekst	Detailinfo bepaling 145 (niet-fiscale taken - andere: ...)
A380	Tekst	Detailinfo bepaling 380 (controle transportmiddel - andere: ...)
DOSSIERNUMMER	Tekst	(referentienummer voor intern gebruik)
BATCH	Ja/nee	(intern gebruik)

*: niet of slechts onsystematisch beschikbaar in de testset

Tabel 1: specificatie brontabel controlegegevens douane (zie ook Appendix B)

Code	Omschrijving	Agoracateg.	Eurocateg.
111	Transport accijnsgoederen – Alle goederen	20	–
112	Transport accijnsgoederen – Alcohol	20	–
113	Transport accijnsgoederen – Minerale oliën	20	–
114	Transport accijnsgoederen – Tabak	20	–
115	Transport accijnsgoederen – Andere	20	–
121	Invoerrechten – Alle goederen	19	–
122	Invoerrechten – Geviseerde goederen (...)	19	–
131	BTW-controle – Alle goederen	20	–
132	BTW-controle – Geviseerde goederen (...)	20	–
133	BTW-controle – PV 108 opgesteld	–	–
134	BTW-controle – PV 109 opgesteld	–	–
141	Niet-fiscale taken – Drugs	36	–
142	Niet-fiscale taken – Immigratie	36	–
143	Niet-fiscale taken – CITES	36	–
144	Niet-fiscale taken – Cultuurpatrim.	36	–
145	Niet-fiscale taken – Andere	36	–
210	Brandstofcontrole – Op de openbare weg	18	–
220	Brandstofcontrole – In tankstations	18	–
230	Brandstofcontrole – Klein gasoliemonster voor labo	18	–
240	Brandstofcontrole – Groot gasoliemonster voor labo	18	–
310	Controle transportmiddel – Verkeersbelasting	16	–
311	Controle transportmiddel – Aanv. verk.bel.	16	–
320	Controle transportmiddel – Bezoldigd verkeer	36	–
330	Controle transportmiddel – Verzekering	15	–
340	Controle transportmiddel – Eurovignet	17	–
350	Controle transportmiddel – Inschrijvingsbewijs	14	–
360	Controle transportmiddel – Technische controle	8	D03
370	Controle transportmiddel – Rijbewijs	12	–
380	Controle transportmiddel – Andere	36	–
410	PV inzake douane en accijnzen – PV opgemaakt	–	–
510	Controle stopgezet – Leeg transportmiddel (vrachtwagen, koffer, ...)	–	–
520	Controle stopgezet – Andere goederen dan gerichte controle	–	–
530	Controle stopgezet – Dienstverhindering, vluchtmisdrijf, enz.	34	–

Tabel 2: codeerschema douane; mapping naar Agora- en ‘Euro-Controle-Wegen’-classificatie

Gezien het opzet van de oefening nam het uitwerken van de gegevenstransformatie en de data-cleaning het merendeel van het werk in beslag.

Met name het parsen en opschonen van de voertuiginformatie (Figuur 7 toont een eenvoudig voorbeeld van de omzetting), die dient als basis voor het opvullen van de dimensietabel ‘Voertuigen’ vergt heel wat code om een aantal redenen. Ten eerste werd de voertuigtype-codering (1: personenwagen, enz.) niet altijd strikt gevolgd (ook tekstomschrijvingen werden teruggevonden). Daarom onderscheidt onze code iedere keer een aantal mogelijke synoniemen voor elk voertuigtype. Ten tweede dienen de nummerplaatkentekens eerst bewerkt te worden om dubbels te vermijden, onder meer door het verwijderen van spaties, koppeltkens, en dergelijke meer. Ten derde wordt de voertuignationaliteit niet consequent geregistreerd.

Zo vinden we bijvoorbeeld zowel ‘Belg’, ‘België’, ‘Belgique’, ‘B’, enz. terug voor de Belgische nationaliteit. Om hieraan te verhelpen werd een hulptabel verzameld (‘NationMapping.mdb’) die elke in de testset voorkomende omschrijving (een 130-tal) linkt aan een van de gestandaardiseerde landcodes. In tweede instantie wordt een passende nationaliteit opgezocht in de gestandaardiseerde landenlijst zelf (drie-letter-ISO-afkorting of de volledige landnaam). Indien de nationaliteitsinformatie ontbreekt in het ‘VOERTUIG’-veld, wordt tevens getracht de inhoud van het ‘NATIONALITEIT’-veld ter identificatie te gebruiken. Tenslotte wordt de Belgische nationaliteit aangenomen, indien de nationaliteit ontbreekt en het kenteken aan de vormvereisten van een Belgische nummerplaat voldoet (drie letters gevolgd door drie cijfers). Wij merken op dat dergelijke databewerkingen veeleer regel dan uitzondering zijn bij de uitvoering van datawarehousing-projecten, en dat bijgevolg zelden 100% nauwkeurige informatie kan gegarandeerd worden. De kwaliteit van de vergaarde informatie kan op termijn verder verbeterd worden door deze cleansing-aspecten zoveel mogelijk op te vangen bij de bron zelf, maar de datawarehousing-literatuur leert ons dat hier doorgaans niet kan en mag op gewacht worden.

VOERTUIG: ";6/HIX220/België;7/UJJ317/België" (broninformatie douane)

⇒

Controles			
...	VoertuigSleutel	VoertuigOpleggerSleutel	...
...	82	83	...

Voertuigen			
VoertuigSleutel	VoertuigCategorie	Nummerplaat	Nationaliteit
82	1	HIX220	150
83	1	UJJ317	150

Figuur 7: voorbeeld omzetting voertuigdata

Een ander probleem deed zich voor bij de identificatie van de werkgever (m.a.w. voor het opvullen van de dimensietabel ‘ondernemingen’). Hoewel in de brongegevenstabel een veld wordt voorzien voor de registratie van het BTW-nummer, werd deze informatie in de testperiode (het eerste kwartaal van 2003) vooralsnog niet op het terrein vergaard. Deze beperking maakt dat het in de context van deze proefstudie onmogelijk was controles en vastgestelde overtredingen te linken aan individuele bedrijven. Indien een tweede bron aangeboord zou worden, namelijk de DIV-nummerplattendatabank, dan zou de vergaarde voertuiginformatie (zie hoger) wél toelaten Belgische transporteurs – zij het indirect – te identificeren. Voor buitenlandse ondernemingen (en de Europese rapportering hierover) is dit uiteraard geen oplossing. In hoeverre dit probleem in de toekomst kan aangepakt worden door aanpassing van de procedures op het terrein en uitbreiding van de douanedatabank moet verder bestudeerd worden door de betrokken partners.

Eerder merkten wij al op dat bij de douane, net zoals bij de meeste andere diensten, geen data geregistreerd worden in verband met de identiteit van de gecontroleerde bestuurder. De corresponderende velden in het datawarehouse blijven dan ook leeg.

In bijlage geven we zowel een overzicht van de datawarehousegegevens die uit de douaneset geëxtraheerd worden, en hun toekenning of herkomst (cf. Appendix C), als de daartoe vereiste Delphi-datamodule-code (cf. Appendix D).¹⁰ Merk op dat, gezien het doel van het datawarehouse, enkel die gegevens weerhouden worden die betrekking hebben op het personen- en goederenverkeer. Deze selectie gebeurt op basis van de beschikbare voertuiginformatie: de categorieën ‘lichte vrachtwagen’, ‘vrachtwagen’, ‘trekker’, ‘oplegger’ en ‘autobus’ worden weerhouden, personenwagens, mobilhomes en andere niet.

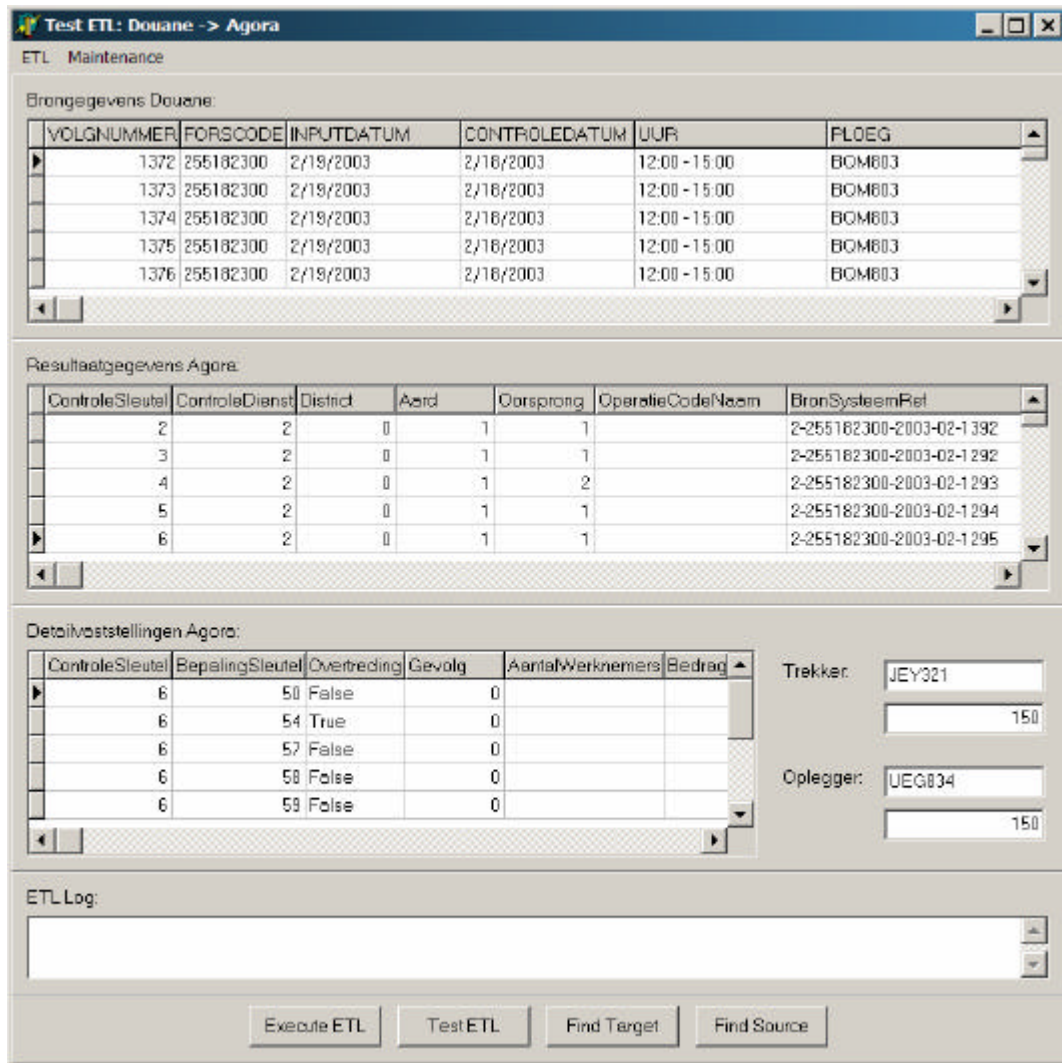
De hoger beschreven ETL-proefimplementatie werd geïntegreerd in een eenvoudige grafische interface (de applicatie ‘AgoraETLDouane’), die toelaat de functionaliteit ervan te testen (cf. Figuur 9). Deze Delphi-applicatie maakt de verbinding tussen bron- en doeldatabank (respectievelijk ‘TestDataDouane.mdb’ en ‘AgoraDB.mdb’, beide in de applicatie-directory te plaatsen). Om de gebruiker toe te laten een nieuw ETL-proces uit te voeren, naar een doeldatabank zonder controledata, bevat de menu-hoofding ‘Maintenance’ een reeks items waarlangs de inhoud van ‘AgoraDB.mdb’ opnieuw kan gewist worden (‘Delete Agora Table’: ‘Controles’, ‘Vaststellingen’, ‘Tijd’, ‘Voertuigen’) en de tijdtabel kan opnieuw gevuld worden (‘Auto-fill tijd’). De ETL-logica kan of getest worden op het huidig geselecteerde record in de douanedatabank (menu-ingang: ETL – Test), of in haar geheel (menu-ingang: ETL – Execute). In de loop van het laatstgenoemde proces worden bovendien een reeks waarschuwingen gelogd naar een tekstbestand (‘Log.txt’), dat nadien bijvoorbeeld kan gebruikt worden voor validatiedoeleinden (zie Figuur 8). De tussen haakjes opgenomen bronsysteemreferentienummers laten toe het originele record te identificeren (de samenstelling van dit veld wordt nader beschreven in Appendix 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"  
(...)
```

Figuur 8: fragment logbestand

Uiteraard neemt de ETL-uitvoering enige tijd in beslag en wordt daarom een vooruitgangvenster getoond. Ter indicatie, de brondatabank bevat 44657 controlerecords, waarvan, na selectie (zie hoger), 14602 controles weerhouden worden in de resultaatbank, gelinkt aan 56194 individuele vaststellingen (positief of negatief). Het formulier laat toe door bron- en resultaatgegevens te browsen. Voor een gebruiksvriendelijkere raadplegingwijze verwijzen wij echter naar de volgende sectie.

¹⁰ Het leek ons weinig zinvol om ook de gebruikersinterface-code, of de code van de in de volgende sectie beschreven toepassing, in tekstbijlage op te nemen. Deze bevat dus slechts een deel van de implementatie.



Figuur 9: schermvoorbeeld gebruikersinterface ETL-implementatie

6.2. Raadpleging doelgegevens – demonstratieomgeving

Naast het hoger beschreven ETL-proces, werd bovendien een grafische gebruikersinterface ontwikkeld om de resultaatgegevens te doorzoeken en te analyseren. Deze beoogt zeker niet om volledig te zijn maar is vooral bedoeld om een beter idee te geven van de mogelijkheden en beperkingen van het datawarehouse en om de resultaatgegevens te kunnen valideren.

Het hoofdscherm (zie Figuur 10) is opgevat als zoekformulier: aan de hand van een aantal door de gebruiker te specificeren zoekcriteria (bijvoorbeeld datum, zaken- of personenvervoer, voertuignationaliteit, nummerplaat¹¹, soort bepaling/inbreuk) wordt een selectie uitgevoerd over de verzamelde controlevaststellingen.¹² Hiertoe wordt at run-time een join-query

¹¹ Hier kan ook een onvolledige zoekterm opgegeven worden met behulp van wildcards, namelijk '%' voor een willekeurig aantal tekens of '_' voor één teken.

¹² Wij hebben ons hierbij enigszins beperkt op basis van de beschikbare gegevens, en bijvoorbeeld geen zoekcriterium voor het betrokken transportbedrijf (b.v., BTW-nummer of naam) voorzien om de eenvoudige reden dat deze vooralsnog niet beschikbaar waren in de brongegevens.

opgesteld en uitgevoerd over de feiten- en dimensietabellen van het datawarehouse. Het resultaat van deze zoekopdracht wordt in spreadsheet-vorm afgebeeld.

Resultaat zoekopdracht:

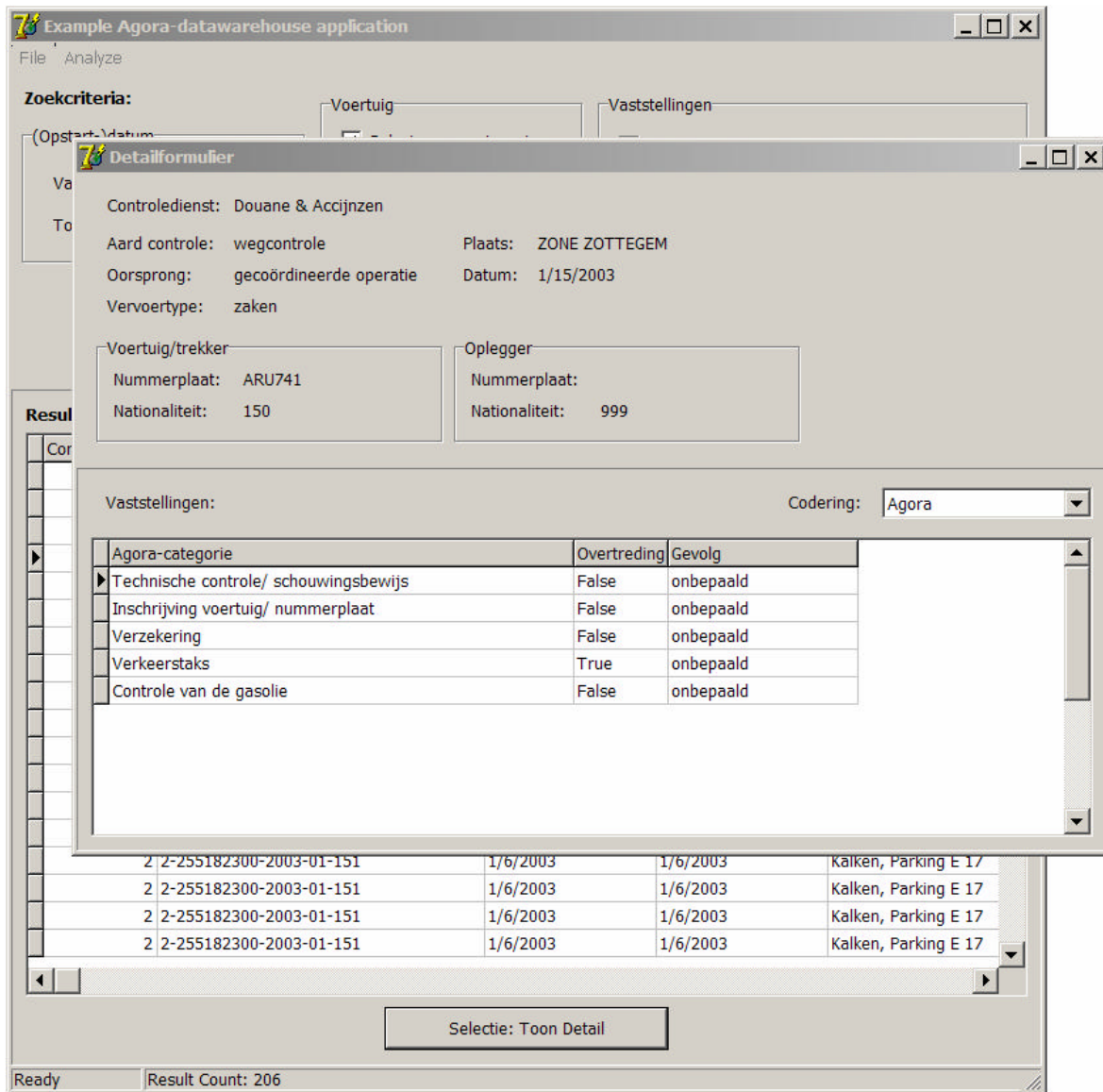
Controledienst	Bronstysteemref.	Opstartdatum	Afsluitdatum	Plaatsomschrijving
2	2-255184400-2003-01-301	1/8/2003	1/8/2003	ZONE DENDERMONDE
2	2-255184400-2003-01-302	1/8/2003	1/8/2003	ZONE DENDERMONDE
2	2-255184400-2003-01-303	1/8/2003	1/8/2003	ZONE DENDERMONDE
2	2-255184400-2003-01-754	1/15/2003	1/15/2003	ZONE ZOTTEGEM
2	2-255184400-2003-01-754	1/15/2003	1/15/2003	ZONE ZOTTEGEM
2	2-255184400-2003-01-754	1/15/2003	1/15/2003	ZONE ZOTTEGEM
2	2-255184400-2003-01-754	1/15/2003	1/15/2003	ZONE ZOTTEGEM
2	2-255182300-2003-01-150	1/6/2003	1/6/2003	Kalken, Parking E 17
2	2-255182300-2003-01-150	1/6/2003	1/6/2003	Kalken, Parking E 17
2	2-255182300-2003-01-150	1/6/2003	1/6/2003	Kalken, Parking E 17
2	2-255182300-2003-01-150	1/6/2003	1/6/2003	Kalken, Parking E 17
2	2-255182300-2003-01-150	1/6/2003	1/6/2003	Kalken, Parking E 17
2	2-255182300-2003-01-150	1/6/2003	1/6/2003	Kalken, Parking E 17
2	2-255182300-2003-01-151	1/6/2003	1/6/2003	Kalken, Parking E 17
2	2-255182300-2003-01-151	1/6/2003	1/6/2003	Kalken, Parking E 17
2	2-255182300-2003-01-151	1/6/2003	1/6/2003	Kalken, Parking E 17
2	2-255182300-2003-01-151	1/6/2003	1/6/2003	Kalken, Parking E 17

Ready | Result Count: 206

Figuur 10: schermvoorbeeld; opzoekscherf

Een detailbeeld van de in dit overzicht geselecteerde controle en de daarbij gedane vaststellingen (zie Figuur 11) kan opgeroepen worden door een druk op de knop onder aan het venster. In het vaststellingen-paneel onder aan dit detailvenster kan de gebruiker vervolgens kiezen uit drie mogelijke voorstellingswijzen: de dienstspecifieke codering, de Agora-classificatie (cf. het afgebeelde voorbeeld) of die van de administratieve regeling 'Euro-Controle-Wegen'.

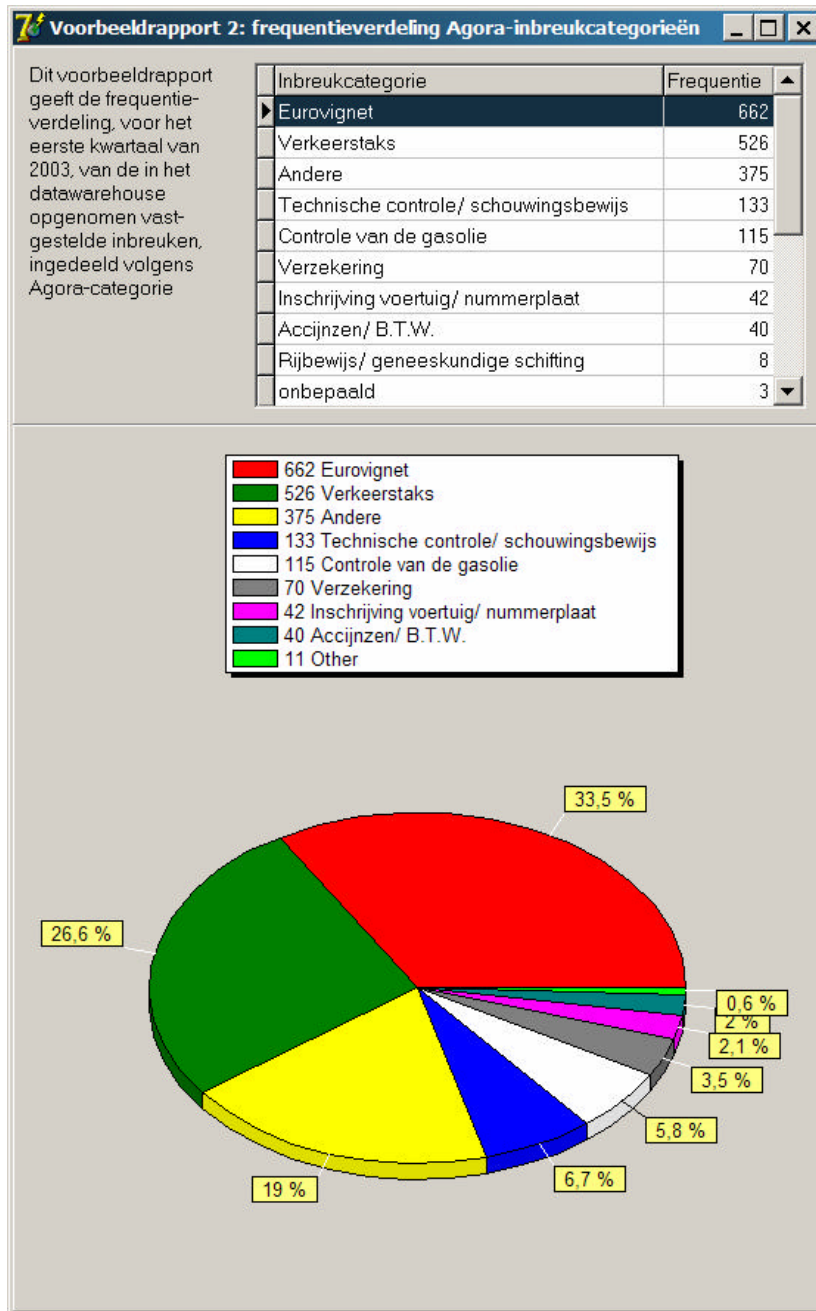
In de uiteindelijke implementatie zou een vergelijkbaar opzoeksysteem, mits de nodige beveiliging en filtermechanismen, via het web aangeboden kunnen worden aan de diverse diensten.



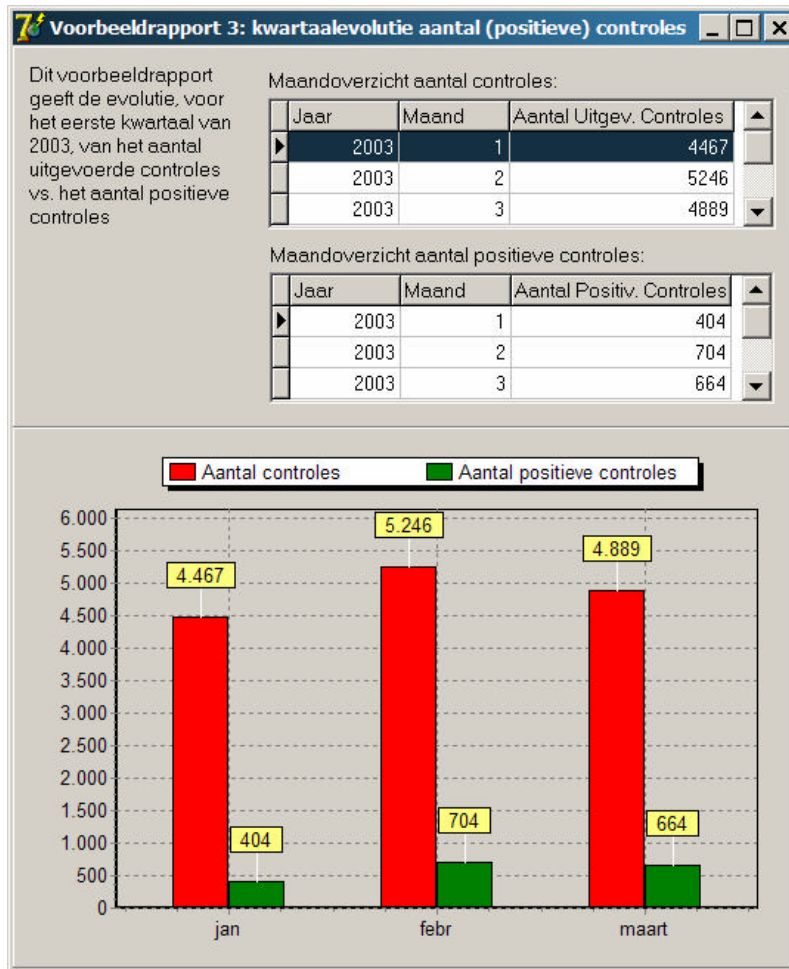
Figuur 11: schermvoorbeeld; detailscherm

Naast zoekmogelijkheden biedt de ontwikkelde omgeving ook een aantal voorbeeldrapporten aan die aan de hand van de verzamelde data aangemaakt worden. Deze kunnen geselecteerd worden met het menu boven aan het hoofdvenster. Een eerste voorbeeldrapport geeft de frequentieverdeling van de nationaliteiten van alle in het datawarehouse opgenomen voertuigen (waar beschikbaar), zowel in de vorm van een rangschikking als in taartdiagram. Een tweede voorbeeldrapport geeft de frequenties, voor het eerste kwartaal van 2003, van de in het datawarehouse opgenomen vastgestelde inbreuken, ingedeeld volgens Agora-categorie. Figuur 12 toont een afdruk van dit rapport. Een derde voorbeeldrapport geeft de evolutie, voor het eerste kwartaal van 2003, van het aantal uitgevoerde controles versus het aantal positieve controles (zie Figuur 13).

Indien bij de uiteindelijke implementatie gebruik gemaakt wordt van een krachtigere datawarehouse-tool, zouden dergelijke rapporten eventueel ook relatief eenvoudig door de gebruiker zelf kunnen samengesteld worden langs een point-and-click interface. In deze oefening hebben wij ons uiteraard enigszins moeten beperken tot het manueel opstellen van de nodige (niet verder door de gebruiker customizeerbare) queries.



Figuur 12: schermvoorbeeld; frequentieverdeling Agora-inbreukcategorieën



Figuur 13: schermvoorbeeld; kwartaalevolutie aantal (positieve) controles

Ten slotte maken de verzamelde data meer dimensionele analyses mogelijk, bijvoorbeeld aan de hand van kruistabellen waarin (sub)totalen en gemiddelden uitgezet worden tegen een aantal selecteerbare dimensies. Een dergelijke voorstelling laat de gebruiker toe mogelijke trends en relaties actief te verkennen. Voor dit onderdeel maakten wij gebruik van de TDecisionCube-componentklassen die deel uitmaken van de met de Delphi-omgeving meegeleverde beslissingsondersteuningscomponenten. Figuur 14 en Figuur 15 tonen twee schermvoorbeelden. In Figuur 14 wordt het aantal uitgevoerde controles geaggregeerd per maand (kolommen) en per vervoercategorie (rijen); in de laatste rij (kolom) wordt iedere keer de deelsom gemaakt voor alle categorieën (maanden). Door op het kruistekentje te klikken kunnen deze totalen verder uitgesplitst worden naargelang een overtreding werd vastgesteld of niet. De resulterende tabel vinden we terug in Figuur 15. Nog een mogelijke dimensie die in deze voorbeeldtoepassing aangekruist kan worden is nationaliteit. Voorts kunnen al deze dimensies via drag-and-drop onderling herschikt worden, zowel van volgorde als tussen rijen en kolommen, om zo verschillende mogelijke verbanden op te sporen. Ten slotte kan de gebruiker ook een andere meetwaarde kiezen, meer bepaald het aantal processen verbaal (door rechts te klikken en het desbetreffende item in het popupmenu te kiezen). Dit is uiteraard slechts ter illustratie; meer algemeen zou het in een volledig uitgewerkte beslissingsondersteuningsomgeving mogelijk moeten zijn om volledig à la carte te kiezen uit alle beschikbare dimensietabelattributen (voor de rij- en kolomindelingen) en meetwaarden (voor de celwaarden).

Meerdimensionale 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

Figuur 14: schermvoorbeeld; aantal controles per maand/vervoercategorie

Meerdimensionale 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

Figuur 15: schermvoorbeeld; aantal controles per maand/vervoercategorie

7. Conclusie: huidige situatie & aanbevelingen vervolgtraject

7.1. Op wetgevend en juridisch vlak

De wet op de bescherming van de persoonlijke levenssfeer is van toepassing op de persoonsgebonden gegevens die verzameld zullen worden. Er zal daarom advies gevraagd moeten worden aan de privacycommissie om deze databank te starten en verder te ontwikkelen.

Er zal verder nagegaan moeten worden, in overleg met parket en auditoraten, in hoeverre de opname van gegevens wat betreft inbreuken in overeenstemming is of kan gebracht worden met het geheim van onderzoek in strafzaken.

Om de koppeling te maken met de operationele databanken beheerd door de I.S.W. (Inspectie Sociale Wetten) en de S.I. (Sociale Inspectie), en, in een latere fase, met additionele gegevensbronnen binnen het netwerk beheerd door de Kruispuntbank van de Sociale Zekerheid (KSZ) (zoals het door de DIMONA-aangifte gevoede elektronische personeelsregister), zal het in de toekomst aangewezen worden om het INSZ (het unieke identificatienummer voor de sociale zekerheid) van een gecontroleerde werknemer in de databank op te nemen. Mogelijk zal in een eerste fase ook het RSZ-nummer van de werkgever opgenomen worden. Voor bei-

de zal een verzoek ingediend moeten worden bij het Beheerscomité van de KSZ. Deze aanvraag kan bijvoorbeeld uitgewerkt worden in samenwerking met de partners van de R.S.Z..

Met het oog op de opname van het KBO-nummer (Kruispuntbank Ondernemingen) dient verder geïnformeerd te worden bij de systeembeheerder over eventuele procedure-eisen.

Om het actieplan van 20 november 2001 verder te operationaliseren, met name voor wat betreft de uitwisseling van data tussen de vermelde partners, zullen mogelijk bijkomende uitvoeringsbesluiten vereist zijn. Met name voor de uitwisseling van andere dan anoniem gemaakte data door federale en lokale politie bestaat vooralsnog geen precedent.

7.2. Op inhoudelijk en technisch vlak

Bij de inventarisering van de beschikbare operationele databronnen in de diverse diensten stelden wij enkele belangrijke tekortkomingen vast die het opzetten van een datawarehouse-systeem voor de wegvervoercontroles in de weg staan. In de eerste plaats is momenteel nog geen operationele databank voorhanden bij het Directoraat-generaal Vervoer te Land, wat een noodzakelijke voorwaarde is om, in een daaropvolgende fase, een datawarehouse op te zetten en te voeden. Om deze voorafgaande fase versneld te kunnen doorvoeren, zou additionele ondersteuning aangewezen zijn; deze zou daarna ook nuttig zijn bij de ontwikkeling van het datawarehouse.

Ook bij de systemen van de andere partners werden bepaalde problemen vastgesteld, die de automatische data-aanvoer zullen compliceren. Zo bestaat bij de I.S.W. de moeilijkheid dat de gegevens momenteel niet gecentraliseerd worden in één grote databank. Daarom werd voorgesteld om in een eerste datawarehousing-fase enkel de gegevens van Mechelen en Huy op te nemen, alvorens een uitgebreider toevoersysteem op te zetten.

Wij stelden bovendien vast dat de beschikbare bronsystemen slechts in zeer beperkte mate voertuig- of bestuurder-identificatiedata kunnen aanleveren. Dit maakt dat het datawarehouse, in de huidige context, onvermijdelijk slechts onvolledige data zal bevatten. Het verdient daarom aanbeveling dat alle partners in de mate van het mogelijke, en binnen de grenzen van hun specifieke bevoegdheden, ernaar zouden streven sommige van deze additionele identificatiedata in een systematische vorm op te nemen in hun toekomstige systeemversies.

Bij de federale en lokale politie zijn de nieuwe systemen die een toevoerbron kunnen vormen voor het geplande datawarehouse nog deels in ontwikkeling. Nauwer overleg is hier aangewezen, onder meer om onze behoeften kenbaar te maken, en beter zicht te krijgen op het wettelijk kader.

7.3. Op organisatorisch en budgettair vlak

Een dergelijk omvangrijk datawarehouse-project kan enkel slagen indien er de nodige middelen voor gealloceerd worden. Uit deze studie bleek dat het implementeren van de data-extractieroutines en de organisatie van het bevoorradingproces een omvangrijke opdracht zal zijn, zeker gezien de heterogeniteit van de beschikbare bronsystemen. Bovendien dient een dergelijk datawarehouse, ook ná de bouw van een eerste versie, verder te evolueren op basis van de behoeften en de beschikbare aanvoerbronnen, en is een continue bestraffing nodig voor de periodieke bevoorrading, foutbehandeling, analyse/rapportering, en de archivering van oude data.

Wij willen er op wijzen dat niet alleen bij de initiatiefnemende dienst, maar ook bij de andere partners, een inspanning nodig zal zijn om een dergelijk systeem mee te helpen opzetten en (vooral) te onderhouden. Duidelijke afspraken moeten worden gemaakt omtrent de organisatie van de data-aanlevering, de terugzending en verdere verwerking van data indien fouten worden vastgesteld bij de invoer ervan, het doorgeven van nieuwe dienstspecifieke inbreukcodes en zo meer. De verdere uitbouw van het datawarehouse zou bovendien een belangrijke inputfactor moeten vormen bij de besluitvorming omtrent de eigen databankinfrastructuur, aangezien het huidige aanbod van aanleverbare data nog vrij beperkt is.

7.4. Overzicht mogelijke vervolgstappen

De inventarisatie van de beschikbare bronnen, een initieel modelontwerp voor de structuur van het datawarehouse en een 'proof-of-concept' oefening zijn gebeurd. In parallel zouden nu het wetgevende/juridische luik en het ontwikkelingsluik verder uitgewerkt kunnen worden. Voor wat betreft dit tweede zou voorafgaandelijk binnen Vervoer te Land een operationele databank ontwikkeld en in gebruik genomen moeten worden. De ontwikkeling van het datawarehouse zelf gebeurt typisch volgens een iteratief proces; er zou gestart kunnen worden met de volledige uitwerking van het traject voor een van de partners, op basis waarvan het initiële ontwerp eventueel kan bijgestuurd worden. In een eerste fase denken wij dan vooral aan I.S.W. (met wie al een nauwe samenwerking bestaat) en/of Douane, waarvoor de hoger beschreven proefimplementatie uitgewerkt werd. Sociale Inspectie en Federale Politie kunnen volgen in een tweede fase. De gegevens van Lokale Politie en de controlediensten van de R.S.Z. zijn voorlopig minder relevant voor de wegvervoercontroles. Elk van deze deeltrajecten veronderstelt het uitwerken van volgende deelstappen: (1) gegevensextractie; (2) gegevenstransformatie; (3) laden van het datawarehouse; (4) nabewerking. Bovendien dienen de nodige bevraging- en rapporteringfaciliteiten voorzien te worden.

Ten slotte merken wij nog op dat zich, los van het voorliggende datawarehousing-project voor de controles en overtredingen, eveneens opportuniteiten situeren voor een verbeterde informatie-uitwisseling door het toegankelijker maken van bestaande gegevensbanken. Dit kan bijvoorbeeld door de toegang tot de nummerplaat- en vergunningendatabanken enerzijds en de DIMONA-bank anderzijds uit te breiden.

Bibliografie

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.

Afkortingen en acroniemen

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

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

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

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

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

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

Bijlagen

Bijlagen kunnen opgevraagd worden bij de onderzoeksploeg op het volgende adres:

KU Leuven,
Departement Toegepaste Wetenschappen
Naamsestraat 69
3000 Leuven
e-mail: christophe.mues@econ.kuleuven.ac.be of jan.vanthienen@econ.kuleuven.ac.be

TABEL/ATTRIBUUT	TYPE	OMSCHRIJVING
Controles		centrale feitentabel met de door de diverse diensten uitgevoerde wegcontroles en/of bedrijfsonderzoeken
ControleSleutel	Automatisch toegekend getal	Automatisch toegekend controle-identificatienummer
ControleDienst	Getal	Dienst die de controle uitgevoerd heeft; zie ControleDienstCodes-tabel (0: onbepaald; 1: Vervoer te Land; 2: Douane & Accijnzen; 3: I.S.W.; 4: S.I.; 5: R.S.Z.; 6: Federale Politie; 7: Lokale Politie; 8: Euro Contrôle Route)
District	Getal	District binnen de controledienst (0 indien niet toepasselijk); zie DistrictCodes-tabel (b.v., 1: I.S.W.-Mechelen)
Aard	Getal	Aard van de controle; zie ControleAardCodes-tabel (1: wegcontrole; 2: bedrijfsonderzoek)
Oorsprong	Getal	Oorsprong van de controle; zie OorsprongCodes-tabel (0: onbepaald; 1: routinecontrole; 2: gerichte controle; 3: gecoördineerde operatie; 4: aanvraag externe instantie; 5: klacht; 6: eigen onderzoek)
OperatieCodeNaam	Tekst	Codenaam van de operatie waarbinnen de controle kaderde (veld blijft leeg indien niet beschikbaar/toepasselijk)
BronSysteemRef	Tekst	Karakterreeks die toelaat het oorspronkelijk controle-record in het operationeel bronsysteem te identificeren
InvoerDatumSleutel	Getal	Getalsleutelverwijzing naar de datum waarop het controle-record ingevoerd werd in het datawarehouse; zie Tijd-dimensietabel
OpstartDatumSleutel	Getal	Getalsleutelverwijzing naar de datum waarop de/het wegcontrole/bedrijfsonderzoek uitgevoerd/opgestart werd; zie Tijd-dimensietabel
AfsluitDatumSleutel	Getal	Getalsleutelverwijzing naar de datum waarop de/het wegcontrole/bedrijfsonderzoek uitgevoerd/afgesloten werd (m.a.w., identiek aan de opstartdatumsleutel bij een wegcontrole); zie Tijd-dimensietabel
PlaatsOmschrijving	Tekst	Locatie waar de controle plaatsvond (b.v., "Gent-Brugge E17"; enkel toepasselijk bij een wegcontrole; niet gestandaardiseerd; leeg indien niet beschikbaar/toepasselijk)
VoertuigSleutel	Getal	Getalsleutelverwijzing naar het gecontroleerde voertuig, of, voor een geleed voertuig, naar de trekker (enkel toepasselijk bij een wegcontrole; 0 indien niet van toepassing of onbeschikbaar); zie Voertuigen-dimensietabel
VoertuigOpleggerSleutel	Getal	Getalsleutelverwijzing naar de oplegger (enkel toepasselijk bij een wegcontrole van een geleed voertuig; 0 indien niet van toepassing of onbeschikbaar); zie Voertuigen-dimensietabel
BestuurderSleutel	Getal	Getalsleutelverwijzing naar de bestuurder, indien deze uniek identificeerbaar is op basis van zijn/haar INSZ (enkel toepasselijk bij een wegcontrole; 0 indien niet van toepassing of onbeschikbaar); zie Bestuurders-dimensietabel
BestuurderNaam	Tekst	Naam en voornaam van de bestuurder (veld blijft leeg indien niet beschikbaar/toepasselijk; geen controle op unieke identificatie)
BestuurderAdres	Tekst	Adres van de bestuurder (veld blijft leeg indien niet beschikbaar/toepasselijk)
WerkgeverSleutel	Getal	Getalsleutelverwijzing naar de gecontroleerde (Belgische) transportonderneming; zie Ondernemingen-dimensietabel; 0 indien niet identificeerbaar in deze laatste
WerkgeverNaam	Tekst	Naam van de gecontroleerde onderneming; laat b.v. rapportering toe m.b.t. wegcontroles van buitenlandse ondernemingen (die niet in de Ondernemingen-dimensietabel opgenomen zijn); leeg indien onbeschikbaar

TABEL/ATTRIBUUT	TYPE	OMSCHRIJVING
WerkgeverNationaliteit	Getal	Getalcode voor de nationaliteit van de gecontroleerde onderneming; zie Landcodes -tabel (b.v., 150=België; 999=onbepaald)
Belg	Ja/nee	Ja/nee-variabele: ja: Belgische onderneming; nee: niet-Belgisch; leeg indien informatie onbeschikbaar
EuroControleRouteLand	Ja/nee	Ja/nee-variabele: ja: onderneming uit Euro Contrôle Route land; nee: ander; leeg indien informatie onbeschikbaar
EU	Ja/nee	Ja/nee-variabele: ja: EU-onderneming; nee: niet-EU; leeg indien informatie onbeschikbaar
Overtreding	Ja/nee	Ja/nee-variabele die aangeeft of een overtreding werd vastgesteld bij de controle; leeg indien informatie onbeschikbaar
AantalGecontrolVoertuigen	Getal	Getalmeetwaarde: aantal gecontroleerde voertuigen; altijd 1 bij een wegcontrole; 0, 1 of hoger bij een bedrijfsonderzoek; leeg indien informatie onbeschikbaar
AantalGecontrolWerknemers	Getal	Getalmeetwaarde: aantal gecontroleerde werknemers; leeg indien informatie onbeschikbaar
AantalPVs	Getal	Getalmeetwaarde: aantal PV's opgesteld in het kader van de controle; leeg indien informatie onbeschikbaar
GeindBedrag	Getal	Getalmeetwaarde: (totaal) geïnd bedrag in Euro; 0 indien geen onmiddellijke inning; leeg indien informatie onbeschikbaar
RegularisatieBedrag	Getal	Getalmeetwaarde: (totaal) geregulariseerd bedrag in Euro; 0 indien geen regularisatie; leeg indien informatie onbeschikbaar
Tijd		dimensietabel met tijdsinformatie omtrent de uitgevoerde controles
TijdSleutel	Automatisch toegelend. getal	Automatisch toegekend recordidentificatienummer
Datum	Datum	Datum (b.v., 31/5/2003)
Maand	Getal	Maand - als cijfer (b.v., 5)
Kwartaal	Getal	Kwartaal (b.v., 2)
Jaar	Getal	Jaar (b.v., 2003)
Voertuigen		dimensietabel met detailinformatie inzake de gecontroleerde voertuigen
VoertuigSleutel	(Automatisch toegelend.) getal	Automatisch toe te kennen voertuigrecord-identificatienummer (0 wordt gereserveerd voor nulverwijzingen in de Controles-feitentabel)
VoertuigCategorie	Getal	Voertuigcategorie: zaken/personen; zie VoertuigCategorieCodes -tabel (0: onbepaald; 1: zaken; 2: personen)
Nummerplaat	Tekst	Nummerplaat (max. 10 karakters)
Nationaliteit	Getal	Nationaliteit-getalcode; zie Landcodes -tabel (b.v., 150=België; 999=onbepaald)
Bestuurders		dimensietabel met detailinformatie inzake de op basis van hun INSZ identificeerbare gecontroleerde bestuurders
BestuurderSleutel	(Automatisch toegelend.) getal	Automatisch toe te kennen bestuurderrecord-identificatienummer (0 wordt gereserveerd voor nulverwijzingen in de Controles-feitentabel)
INSZ	Tekst	Identificatienummer v/d sociale zekerheid (INSZ); leeg indien niet toepasselijk/beschikbaar
Naam	Tekst	Naam en voornaam van de bestuurder
BeginDatum	Datum	Begindatum van de geldigheidsperiode van de onderstaande bestuurdergegevens
EindDatum	Datum	Einddatum van de geldigheidsperiode van de onderstaande bestuurdergegevens; 12/31/2999 zolang ongewijzigd

TABEL/ATTRIBUUT	TYPE	OMSCHRIJVING
Nationaliteit	Getal	Getalcode voor de nationaliteit van de bestuurder; zie Landcodes-tabel (b.v., 150=België; 999=onbepaald)
Statuut	Getal	Sociaal statuut van de bestuurder; zie BestuurderStatuutCodes-tabel (0: onbepaald; 1: werknemer; 2: zelfstandige; 3: interim)
Ondernemingen dimensietabel met detailinformatie inzake de Belgische transportondernemingen		
OndernemingSleutel	(Automatisch toegekend) getal	Automatisch toe te kennen record-identificatienummer van de (Belgische) transportonderneming (0 wordt gereserveerd voor nulverwijzingen in de Controles-feitentabel)
Naam	Tekst	Officiële naam van de onderneming
RSZ_Nummer	Tekst	RSZ-nummer
KBO_Nummer	Tekst	KBO (Kruispuntbank van Ondernemingen)-nummer; vervangt BTW-nummer
Vaststellingen feitentabel met de bij de controles gedane vaststellingen		
ControleSleutel	Getal	Getalsleutelverwijzing naar het automatisch toegekend controle-identificatienummer uit de Controles-feitentabel
BepalingSleutel	Getal	Getalsleutelverwijzing naar de respectieve (dienstspecifieke) bepalingen waarop, tijdens die controle, gecontroleerd werd; zie Bepalingen-tabel
Overtreding	Ja/nee	Ja/nee-variabele die aangeeft of bij de controle al dan niet een overtreding werd vastgesteld op hogervermelde bepaling (opm.: bij niet-overtreding zal niet altijd een record hieromtrent beschikbaar zijn in de operationele bron)
Gevolg	Getal	Wijze waarop de vaststelling verder afgehandeld werd; zie GevolgCodes-tabel (0: onbepaald; 1: geen; 2: waarschuwing; 3: regularisatie; 4: onmiddellijke inning; 5: pro justitia; 6: doorverzending; 7: ander)
AantalWerknemers	Getal	Aantal betrokken werknemers; leeg indien informatie onbeschikbaar
Bedrag	Getal	Geïnd/geregulariseerd (afhankelijk van "Gevolg"-waarde) deelbedrag in Euro voor de betreffende vaststelling; 0 indien geen onmiddellijke inning of regularisatie; leeg indien informatie onbeschikbaar
Bepalingen tabel met de door de diensten gehanteerde coderingen inzake de controleerbare wettelijke bepalingen		
Bep_Sleutel	Automatisch toegekend getal	Automatisch toegekend record-identificatienummer voor de (dienstspecifieke) wettelijke bepaling
Code	Tekst	Dienstspecifieke codering van de bepaling (b.v., "D20811")
Omschrijving	Tekst	Tekstuele omschrijving van de bepaling (b.v., "onvoldoende dagelijkse rusttijd - minder dan 11 uren - art. 8.1 1e alinea = 62 euro per periode van 30 minuten")
Dienst	Getal	Controledienst die de bepaling/code hanteert; zie ControleDienstCodes-tabel (1: Vervoer te Land; 2: Douane & Accijnzen; 3: I.S.W.; 4: S.I.; 5: R.S.Z.; 6: Federale Politie; 7: Lokale Politie; 8: Euro Contrôle Route)
BeginDatum	Datum	Datum van invoering (defaultwaarde: 1/1/1800, indien reeds geldig vóór de ingebruikname v/h datawarehouse)
EindDatum	Datum	Datum van schrapping indien de code niet langer gehanteerd wordt; 12/31/2999 zolang in gebruik
AgoraCategorie	Getal	Getalsleutelverwijzing naar een van de nog door de partners overeen te komen gemeenschappelijke controlecategorieën; zie tabel Bep_AgoraCategorieën

TABEL/ATTRIBUUT	TYPE	OMSCHRIJVING
EuroCategorie	Getal	Getalsleutelverwijzing naar de overeenkomstige Euro Contrôle Route code; zie tabel Bep_EuroCategorieën (0 indien de bepaling niet thuishoort in een van deze categorieën)
Bep_AgoraCategorieën		tabel met de door de diensten te bepalen gemeenschappelijke categorisatie van bepalingen
Cat_Sleutel	(Automatisch toegek.) getal	Automatisch toe te kennen identificatienummer voor de (gemeenschappelijk af te spreken) verzamelcategorie van bepalingen
Cat_Omschrijving	Tekst	Tekstuele omschrijving (b.v., "Tachograaf: rij-/rusttijden & onderbrekingen")
BeginDatum	Datum	Datum van invoering
EindDatum	Datum	Datum van schrapping indien de code niet langer gehanteerd wordt; 12/31/2999 zolang in gebruik
Cat_Toelichting	Tekst	Optionele toelichting
Bep_EuroCategorieën		tabel met de in het kader van Euro Contrôle Route gehanteerde categorisatie van bepalingen
Cat_Sleutel	(Automatisch toegek.) getal	Automatisch toe te kennen identificatienummer voor de in het kader van Euro Contrôle Route gehanteerde categorie v. bepalingen
Cat_Code	Tekst	Codering van de categorie (b.v., "A04")
BeginDatum	Datum	Datum van invoering
EindDatum	Datum	Datum van schrapping indien de code niet langer gehanteerd wordt; 12/31/2999 zolang in gebruik
Cat_Hoofding	Tekst	Hoofding waaronder de categorie is ondergebracht (b.v., "European Social Regulations")
Cat_DetailTekst	Tekst	Tekstuele omschrijving van de (detail-)categorie ("Daily Rest")
BestuurderStatuutCodes		codetabellen v/h datawarehouse
ControleAardCodes		
ControleDienstCodes		
DistrictCodes		
GevolgCodes, LandCodes		
OorsprongCodes		
VoertuigCategorie Codes		

ATTRIBUUT	TYPE	OMSCHRIJVING
(Weg-)controles douanebrigaden		
VOLGNUMMER	Getal	Controlevolnummer in de maand (per forscodel)
FORSODE	Tekst	Forscode van de brigade
INPUTDATUM	Datum	Datum waarop het controleverslag ingevoerd werd
CONTROLEDATUM	Datum	Datum waarop de controle uitgevoerd werd
UUR	Tekst	Uurindicatie uitgevoerde controle
PLOEG	Tekst	Nummerplaat van het dienstvoertuig of roepnaam van het vast team
CONTROLEPUNT	Tekst	Omschrijving locatie v/d controle (b.v., "Gent-Brugge E17")
IMPORTEUR	Tekst	BTW-nummer importeur/bestemming*
EXPORTEUR	Tekst	BTW-nummer exporteur/verzender*
TRANSPORTEUR	Tekst	BTW-nummer transporteur*
VOERTUIG	Tekst	Gestructureerde informatie m.b.t. het gecontroleerde voertuig; type1/nummerplaat1/nationaliteit1; type2/nummerplaat2/nationaliteit2 (types: 1: personenwagen; 2: lichte vrachtwagen; 3: vrachtwagen; 4: autobus; 5: mobilhome; 6: trekker; 7: oplegger; 8: andere)
NATIONALITEIT	Tekst	Omschrijving nationaliteit*
CONTROLETYPE	Getal	Controletype: 1: dynamische controle (interceptie v/h voertuig); 2: statische controle 1 (langs de weg; selectie); 3: statische controle 2 (langs de weg; alle)
CONTROLEAARD	Getal	Controleaard: 1: routinecontrole; 2: gerichte controle (op een specifieke bepaling, cf. infra); 3: (evt. gecoördineerde) operatie
CODENAAM	Tekst	Detailinfo controleaard: bepaling gerichte controle of operatiecodenaam (b.v. 210 of "ZW VERV LOKPOL ZOTTEGEM")
CONTROLESOORT	Tekst	Codelijst bepalingen waarop gecontroleerd werd (elke deelcode wordt vooafgegaan door een puntkomma, b.v.: ";210;320;340"); zie elders voor een overzicht van de gehanteerde codes
VASTSTELLING	Tekst	Codelijst bepalingen waarop een overtreding vastgesteld werd (moet een deellijst zijn v/d bovenstaande lijst)
A115	Tekst	Detailinfo bepaling 115 (transport accijnsgoederen - andere: ...)
A122	Tekst	Detailinfo bepaling 122 (invoerrechten - geviseerde goederen: ...)
A132	Tekst	Detailinfo bepaling 132 (BTW-controle - geviseerde goederen: ...)
A145	Tekst	Detailinfo bepaling 145 (niet-fiscale taken - andere: ...)
A380	Tekst	Detailinfo bepaling 380 (controle transportmiddel - andere: ...)
DOSSIERNUMMER	Tekst	(referentinummer voor intern gebruik)
BATCH	Ja/nee	(intern gebruik)

*: niet of slechts onsystematisch beschikbaar in de testset

TABEL/ATTRIBUUT	TYPE	TOEKENNING/HERKOMST
Controles		
ControleSleutel	Getal	Automatisch toegekende sleutelwaarde: vorig nummer + 1 (procedure-implementatie: zie TSleutelGenerator-klasse); enkel controles die betrekking hebben op personen- of goederentransport worden weerhouden, waarbij de selectie gebeurt op basis van voertuigtype (zie "Voertuigen"-tabel)
ControleDienst	Getal	2 (code voor de controledienst Douane & Accijnzen)
District	Getal	0 (niet toepasselijk)
Aard	Getal	1 (wegcontrole)
Oorsprong	Getal	Bronveld: CONTROLEAARD; 1 -> 1 (routinecontrole); 2 -> 2 (gerichte controle); 3 -> 3 (gecoördineerde operatie)
OperatieCodeNaam	Tekst	Bronveld: CODENAAM
BronSysteemRef	Tekst	Bronvelden: FORSCODE; CONTROLEDATUM; VOLGNUMMER Voorbeeld: 2-255182300-2003-02-1391, waarin - 255182300: forscodelnummer motorbrigade - 2003: jaar - 02: februari - 1391: volgorde in maand, voor brigade
InvoerDatumSleutel	Getal	Sleutelwaarde (cf. tabel "tijd") datum ETL-procesuitvoering
OpstartDatumSleutel	Getal	Bronveld: CONTROLEDATUM; omgezet naar overeenkomstige sleutelwaarde in tabel "tijd"
AfsluitDatumSleutel	Getal	Identiek aan OpstartDatumSleutel (immers enkel wegcontroles)
PlaatsOmschrijving	Tekst	Bronveld: CONTROLEPUNT
VoertuigSleutel	Getal	Bronveld: VOERTUIG; cleansing/parsing procedure-implementatie: ParseVoertuigField; procedure-implementatie opzoeken bestaand record in "Voertuigen"-tabel (cf. infra) of aanmaken nieuw record bij niet-identificatie: UpdateVoertuigenTabel (secundaire identificatiesleutel: Nummerplaat; Nationaliteit). Resultaat is getalsleutelverwijzing naar het gecontroleerde voertuig, of, voor een geleased voertuig, naar de trekker (0 indien onbeschikbaar)
VoertuigOpleggerSleutel	Getal	Bronveld: VOERTUIG; omzetting gelijkaardig aan hierboven; getal-sleutelverwijzing naar de oplegger indien VOERTUIG een tweede deelstring hieraangaande bevat; zoniet 0 (merk op dat opleggers een afzonderlijk record binnen de "Voertuigen"-tabel vormen)
BestuurderSleutel	Getal	0 (niet beschikbaar in het bronsysteem)
BestuurderNaam	Tekst	Leeg (id.)
BestuurderAdres	Tekst	Leeg (id.)
WerkgeverSleutel	Getal	0 (niet systematisch beschikbaar in het bronsysteem)
WerkgeverNaam	Tekst	Leeg (id.)
WerkgeverNationaliteit	Getal	999 (id.)
Belg	Ja/nee	Niet geïmplementeerd
EuroControleRouteLand	Ja/nee	Niet geïmplementeerd
EU	Ja/nee	Niet geïmplementeerd
Overtreding	Ja/nee	waarde 'Ja' indien, na verwerking van de bronvelden CONTROLESOORT en VASTSTELLING (cf. "Vaststellingen"-tabel) een overtreding werd geregistreerd bij minstens één van de aan de controle gelinkte vaststellingen; anders 'Nee'
AantalGecontrolVoertuigen	Getal	1 (wegcontrole)

TABEL/ATTRIBUUT	TYPE	TOEKENNING/HERKOMST
Controles		
AantalGecontrolWerknemers	Getal	Leeg (niet beschikbaar in het bronsysteem)
AantalPVs	Getal	Aantal PV's opgesteld in het kader van de controle; ingevuld na het verwerken van de vaststellingen (cf. bronvelden CONTROLE-SOORT en VASTSTELLING); wij zijn er in deze proefimplementatie vanuit gegaan dat indien 133 (PV 108), 134 (PV 109) of 410 (PV douane & accijnzen) voorkomen als deelcodes, een PV geteld moet worden; bijkomende validatie is hier echter nog gewenst
GeindBedrag	Getal	Leeg (niet beschikbaar in het bronsysteem)
RegularisatieBedrag	Getal	Leeg

Tijd		
TijdSleutel	Getal	Automatisch toegekend record-identificatienummer; in onze proefimplementatie: het aantal dagen sinds 1 januari 1995; een aparte procedure (FillTijdTabel) werd voorzien om deze tabel te vullen
Datum	Datum	Datum (b.v., 31/5/2003)
Maand	Getal	Maand - als cijfer (b.v., 5)
Kwartaal	Getal	Kwartaal (b.v., 2)
Jaar	Getal	Jaar (b.v., 2003)

Voertuigen		
VoertuigSleutel	Getal	Automatisch toegekend voertuigrecord-identificatienummer: vorig nummer + 1 (procedure-implementatie: zie TsleutelGenerator-klasse); 0 wordt gereserveerd voor nulverwijzingen
VoertuigCategorie	Getal	Bronveld: VOERTUIG; cleansing/parsing-deelprocedure: ParseCategorieStr - vervangen van codesynoniemen - omzetting: 2 (lichte vrachtw.), 3 (vrachtw.), 6 (trekk.), 7 (oplegg.) -> 1 (zaken); 4 (autobus) -> 2 (personen); 1 (personenwag.), 5 (mobilhome) of andere: niet weerhouden
Nummerplaat	Tekst	Bronveld: VOERTUIG; cleansing/parsing-deelprocedure: ParsePlaatStr - omzetten naar hoofdletters - verwijderen van controletekens (spaties, koppeltekens, ...) - controle op min. en max. lengte
Nationaliteit	Getal	Bronveld: VOERTUIG; cleansing/parsing-deelprocedure: ParseNationStr; beschreven in sectie 6.1 van de rapporttekst; maakt gebruik van het databestand "NationMapping.mdb"

Bestuurders
Niet gevuld (niet beschikbaar in het bronsysteem)

Ondernemingen
Niet gevuld

TABEL/ATTRIBUUT	TYPE	TOEKENNING/HERKOMST
Vaststellingen		
ControleSleutel	Getal	Getalsleutelverwijzing naar het automatisch toegekend controle-identificatienummer uit de Controles-feitentabel
BepalingSleutel	Getal	Bronveld: CONTROLESOORT; de ";"-gescheiden lijst wordt eerst opgedeeld (cf. procedure ParseListString); voor elke resulterende deelcode wordt een record aangemaakt, de overeenkomstige sleutelwaarde opgezocht in de tabel "Bepalingen", en dan hiernaar verwezen; procedure-implementatie: UpdateVaststellingen-Tabel
Overtreding	Ja/nee	Bronvelden: CONTROLESOORT / VASTSTELLING; waarde 'Ja' bij aanwezigheid van de bepaling in beide codelijsten; anders 'Nee'; de verwerking van de tweede lijst is gelijkaardig aan hierboven
Gevolg	Getal	0 (onbepaald): niet beschikbaar op dit detailniveau
AantalWerknemers	Getal	Leeg
Bedrag	Getal	Leeg

Bepalingen
Inhoud aangemaakt voorafgaandelijk aan de ETL-procesuitvoering; beschreven in sectie 6.1 van de rapporttekst

Bep_AgoraCategorieen
Inhoud aangemaakt voorafgaandelijk aan de ETL-procesuitvoering; beschreven in sectie 5 van de rapporttekst

Bep_EuroCategorieen
Inhoud aangemaakt voorafgaandelijk aan de ETL-procesuitvoering

BestuurderStatuutCodes, ControleAardCodes, ControleDienstCodes, DistrictCodes, GevolgCodes, LandCodes, OorsprongCodes, VoertuigCategorieCodes
Inhoud aangemaakt voorafgaandelijk aan de ETL-procesuitvoering

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, NationStr, PlaatStr]));
                    end
                end
            end
        end
    end

```

```

                [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);

```

```

    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;

```