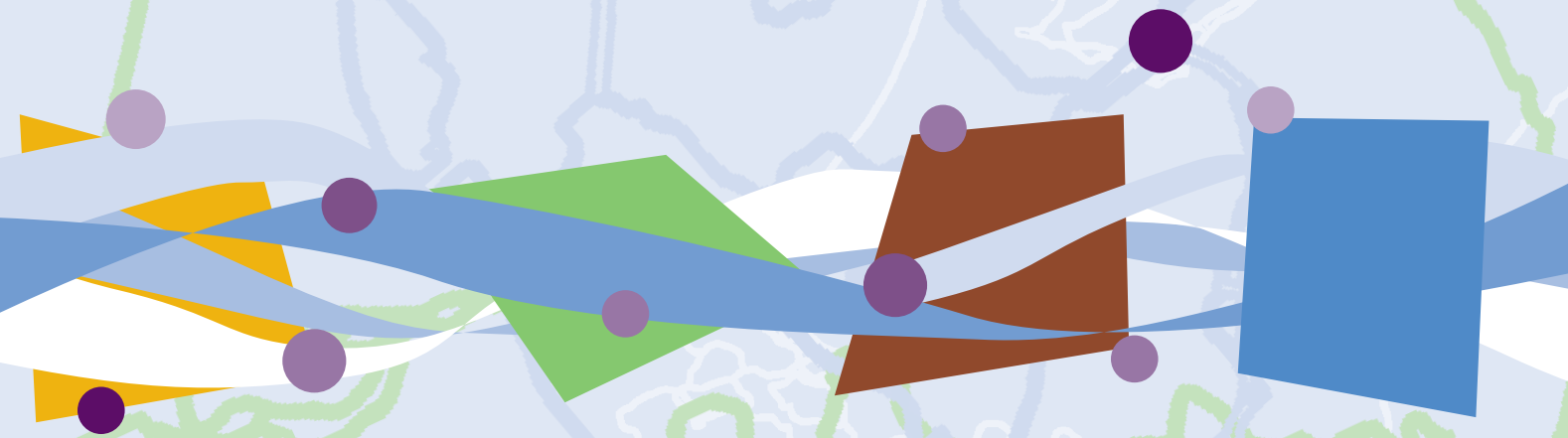
ALTERRA
WAGENINGEN UR

Database ‘Monitoring Stroomgebieden’: Een slimme database voor het beheren van monitoringsgegevens

D.J.J. Walvoort en T.P. van Tol-Leenders



Alterra-rapport 1955, ISSN 1566-7197
Reeks Monitoring Stroomgebieden 17

17



Database ‘Monitoring Stroomgebieden’: Een slimme database voor het beheren van monitoringsgegevens.

In opdracht van de Ministeries van LNV, VROM, en V&W (BO-12.07-005-006)

Database 'Monitoring Stroomgebieden': Een slimme database voor het beheren van monitoringsgegevens.

D. J. J. Walvoort, T. P. van Tol-Leenders

Alterra–Rapport 1955

Monitoring Stroomgebieden

Alterra, Wageningen, 2009

ABSTRACT

Walvoort, D. J. J., T. P. van Tol-Leenders, 2009. *Database 'Monitoring Stroomgebieden': Een slimme database voor het beheren van monitoringsgegevens.*. Wageningen, Alterra–Rapport 1955, Monitoring Stroomgebieden.

In het kader van het project 'Monitoring Stroomgebieden' worden gemeten nutriëntenconcentraties en waterafvoeren opgeslagen in een database. Deze database is voorzien van functionaliteit die de invoer controleert op een groot aantal potentiële fouten. Tevens kan de database op basis van de opgeslagen gegevens afgeleide gegevens berekenen. In dit rapport wordt deze 'slimme database' beschreven.

Trefwoorden: database, monitoring, tijdreeksen, concentratiemetingen, afvoermetingen, stikstof, fosfor

ISSN 1566–7197

Dit rapport is gratis te downloaden van www.alterra.wur.nl (ga naar 'Alterra-rapporten'). Alterra verstrekt geen gedrukte exemplaren van rapporten. Gedrukte exemplaren zijn verkrijgbaar via een externe leverancier. Kijk hiervoor op www.boomblad.nl/rapportenservice.

©2009 Alterra
Postbus 47; 6700 AA Wageningen; Nederland
Tel.: (0317) 474700; fax: (0317) 419000; e-mail: info.alterra@wur.nl

Niets uit deze uitgave mag worden verveelvoudigd en/of openbaar gemaakt door middel van druk, fotokopie, microfilm of op welk ander wijze ook zonder voorafgaande schriftelijke toestemming van Alterra.

Alterra aanvaardt geen aansprakelijkheid voor eventuele schade voortvloeiend uit het gebruik van de resultaten van dit onderzoek of de toepassing van de adviezen.

[Alterra–Rapport 1955/september/2009]

Inhoudsopgave

I De database in vogelvlucht **9**

II Bijlagen:
De database in meer detail **17**

A Motivering van de keuze voor een relationeel database management systeem **19**

B Beschrijving van het datamodel **21**
B.1 Inleiding 21
B.1.1 Tabellengroep `actors` 23
B.1.2 Tabellengroep `geography` 23
B.1.3 Tabellengroep `numerical_data` 24
B.1.4 Tabellengroep `categorical data` 27
B.1.5 Tabellengroep `methodologies` 29
B.1.6 Tabellengroep `verion_control` 29

C Eigenschappen van een slimme database **31**
C.1 Kwaliteitscontrole 31
C.2 Afgeleide waarden 33

D Beschikbare gegevens **41**

E Gebruikershandleiding **43**
E.1 Opslaan van gegevens in de database 43
E.2 Opvragen van gegevens uit de database 46

F Versiebeheer **49**

G Dankbetuiging **51**

Woord vooraf

Deze rapportage vormt een onderdeel van het project 'Monitoring stroomgebieden'. Het primaire doel van dit project is het leveren van een bijdrage aan de evaluatie van het mestbeleid door het kwantificeren van het aandeel van de landbouw in de belasting van het oppervlaktewater en de verandering van dit aandeel van de landbouw als gevolg van (mest)beleid in een aantal representatieve stroomgebieden in karakteristieke landschappelijke regio's. Het secundaire doel is om een methodiek te ontwikkelen die het mogelijk maakt om deze methodiek ook in andere stroomgebieden in te voeren.

Voor dit project zijn vier pilotgebieden geselecteerd: 'de Drentse Aa', 'de Schuitenbeek', 'de Krimpenerwaard' en 'Quarles van Ufford'. De waterbeheerders, Hoogheemraadschap van Schieland en de Krimpenerwaard, Waterschap Veluwe, Waterschap Rivierenland, Waterschap Hunze en Aa's en Waterlaboratorium Noord participeren actief in dit project.

Het project wordt aangestuurd door een stuurgroep. In de stuurgroep hebben de Ministeries LNV, VROM en V&W als opdrachtgevers en een vertegenwoordiger van de Unie van Waterschappen namens de betrokken waterbeheerders zitting. Het project wordt uitgevoerd door Alterra Research Instituut voor de Groene Ruimte onderdeel van Wageningen Universiteit en Researchcentrum en Deltares.

In de reeks rapportages van het project 'Monitoring Stroomgebieden' is per gebied een systeemverkenning verschenen. Op basis van deze uitgevoerde systeemverkenning wordt er aanvullend in de pilotgebieden gemeten en is het gefaseerd opzetten van een modelinstrumentarium per pilotgebied gestart. Deze aanvullende maar ook de historische monitoringsgegevens dienen opgeslagen te worden in een database. Ten behoeve van het project Monitoring Stroomgebieden is hiervoor een nieuwe database ontwikkeld.

Dit rapport beschrijft de motivering achter de ontwikkeling van deze database voor de opslag van monitoringsgegevens, de eigenschappen en het gebruik ervan.

Voor informatie over het project 'Monitoring stroomgebieden' kunt u terecht op `www.monitoringstroomgebieden.nl` of bij:

Dorothée van Tol–Leenders
0317 – 48 42 79
dorothee.vantol-leenders@wur.nl

Frank van der Bolt
0317 – 48 64 44
frank.vanderbolt@wur.nl

Deel I

De database in vogelvlucht

Een slimme database voor het beheer van monitoringsgegevens

In het kader van het project 'Monitoring Stroomgebieden' wordt op een aantal locaties in het oppervlaktewater aanvullend gemeten op het reguliere meetnet. Samen met de reguliere metingen van de waterbeheerders en de historische meetgegevens vergroten deze metingen het inzicht in de werking van de (water)systemen. Alle meetgegevens worden gestructureerd opgeslagen in een database. In tegenstelling tot de meeste gangbare databases is deze database voorzien van een filter waarmee zo veel mogelijk wordt voorkomen dat foutieve gegevens in de database terechtkomen. De database rapporteert aan de gebruiker via een web-browser. Gegevens kunnen eenvoudig aan de database worden toegevoegd. Ook het opvragen van gegevens uit de database is relatief eenvoudig. Bovendien beschikt de database over functionaliteit waarmee op basis van de beschikbare informatie nieuwe informatie kan worden afgeleid. Denk bijvoorbeeld aan het schatten van totaal stikstof op basis van organische en anorganische stikstofcomponenten. De beschikbare informatie in de database wordt op deze wijze optimaal benut. Ook kunnen delen van de database tijdelijk onzichtbaar worden gemaakt voor gebruikers (modelleurs). Deze delen kunnen in een later stadium worden gebruikt voor validatie van het modelsysteem. In onderstaande tekst wordt de database in vogelvlucht beschreven. Voor meer details wordt verwezen naar de bijlagen.

Achtergrond

Bij de start van het project 'Monitoring Stroomgebieden' is in samenwerking met de participerende waterschappen (Waterschap Rivierenland, Waterschap Veluwe, Waterschap Hunze en Aa's en het Hoogheemraadschap van Schieland en de Krimpenerwaard) voor elk pilotgebied ('Quarles van Ufford', 'de Schuitenbeek', 'de Drentse Aa', 'de Krimpenerwaard') een aantal locaties in het oppervlaktewatersysteem geselecteerd waar sinds 2004 aanvullend wordt gemeten op de reguliere meetnetten van de waterbeheerders. Het gaat met name om concentraties van stikstofverbindingen (Kjeldahl-stikstof, nitraat, nitriet, ammonium, en totaal stikstof), fosforverbindingen (orthofosfaat en totaal fosfor), chlorofyl-a, en de waterafvoer. Tevens is vaak ook de pH en de concentraties van een aantal andere kat- en anionen bepaald.

Een meting is meer dan een meetwaarde

Elk waterschap levert vier keer per jaar een bestand aan met daarin de metingen van het afgelopen kwartaal. De meetgegevens zijn zeer heterogeen van aard. Uit figuur 1 blijkt dat een meting niet alleen uit een meetwaarde bestaat, maar ook uit een aantal andere attributen die van groot belang zijn bij de interpretatie van de meting.

Een belangrijk attribuut voor kwantitatieve meetwaarden is de eenheid (bijvoorbeeld mg/l, µg/l, m^3/s). Dit attribuut maakt het mogelijk om tijdens de gegevensverwerking de eenheden op elkaar af te stemmen.

Een ander belangrijk attribuut zijn de coördinaten van de meetlocatie. Hieruit kan de positie worden afgeleid van de meetlocatie ten opzichte van bijvoorbeeld landbouw- en natuurgebieden, inlaten, industriële lozingspunten, of rioolwaterzuiveringsinstallaties (RWZIs). Ook kan worden bepaald of de meetlocatie in een hoofdwaterloop ligt of in een sloot.

Ook de tijdperiode waarop de meting betrekking heeft is relevant. Hieruit kan bijvoorbeeld worden afgeleid of het een momentopname betreft of een dag- of weekgemiddelde. Tevens blijkt hieruit het seizoen waarin de meting is verricht. Dat laatste is relevant voor bijvoorbeeld normtoetsing.

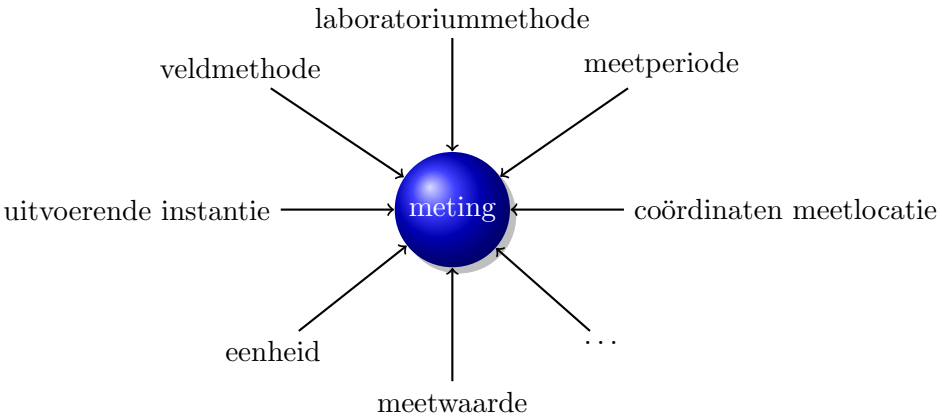
Voorst is de meetwaarde afhankelijk van de de gehanteerde veldmethode (steekmonster, molenmeting, mengmonster, enz.) en de gebruikte laboratoriummethode (colorimetrisch, elektrochemisch, al dan niet gefilterd, enz.). Kennis van deze attributen is noodzakelijk om te voorkomen dat bijvoorbeeld gefilterde en ongefilterde fosformetingen bij de gegevensverwerking op één hoop worden geveegd.

Tot slot kunnen de meetresultaten van laboratorium tot laboratorium verschillen. Daarom zijn tevens de instanties die het veld- en laboratoriumwerk hebben uitgevoerd van belang.

De database als opslagmedium

Het moge duidelijk zijn, niet alleen de meetwaarde zelf is relevant, maar ook de overige attributen in figuur 1. Om al deze informatie op een overzichtelijke en gestructureerde wijze op te slaan, is een database zeer geschikt.

Een database biedt als voordeel dat op eenvoudige wijze selecties kunnen worden uitgevoerd. Zo is het mogelijk om op efficiënte wijze een specifiek deel van de database uit te lezen, zoals alle nitraatmetingen in een bepaalde regio gedurende de laatste 10 jaar, al dan niet geaggregeerd of gesorteerd. Ook maken databases de relaties tussen de gegevens expliciet. Dit schept transparantie. Een ander voordeel is dat databases eenvoudig zijn te koppelen



Figuur 1: Schematische voorstelling van een meting met daaraan gekoppeld een aantal belangrijke attributen.

met externe software zoals geografische informatie systemen (GIS) of statistische pakketten.

Begin 2004 is daarom gestart met het ontwerp en de bouw van een relationeel database management systeem (RDBMS) voor de opslag van de gegevens die in het kader van het project 'Monitoring Stroomgebieden' zijn verzameld. Gedurende de jaren daarna is het RDBMS en de ingebouwde functionaliteit regelmatig uitgebreid op basis van ervaringen met aangeleverde gegevens.

Er is destijds bewust voor gekozen om geen gebruik te maken van bestaande databasesystemen. De functionaliteit van bestaande systemen sloot onvoldoende aan bij het project. Zo moest de database enerzijds in staat zijn om automatisch fouten uit binnenkomende gegevensstromen te filteren en daarover te rapporteren. Anderzijds moest de database op grond van beschikbare gegevens afgeleide gegevens kunnen berekenen (zie ook bijlage A en met name bijlage C voor meer details). De beoogde database gaat dus veel verder dan traditionele databases en moest meer zijn dan alleen een efficiënt opslagmedium. Het idee van een 'slimme database' was geboren.

De *slimme* database: meer dan een database

Een traditionele database voert slechts een zeer beperkte kwaliteitscontrole uit voordat gegevens worden opgeslagen. Veel verder dan de controle op veldtypes (tekst, numerieke waarde, etc.) en niet-unieke gegevens gaan de meeste databasesystemen niet.

De slimme database daarentegen werkt een hele lijst met kwaliteitscontroles af die allemaal succesvol moeten worden doorlopen voordat gegevens worden opgeslagen in de database. Voorbeelden van dergelijke controlestappen zijn cont-

role op interne consistentie van de gegevens (totaalstikstofgehalte moet bijvoorbeeld groter zijn dan of gelijk zijn aan het nitraatstikstofgehalte) en controle op het bereik van systeemeigenschappen (concentraties mogen niet negatief zijn, en de pH moet (veel) kleiner zijn dan veertien). Eventuele onvolkomenheden in de invoergegevens worden gerapporteerd als web-pagina's (HTML-formaat). Een overzicht van de belangrijkste kwaliteitscontroles is gegeven in bijlage C.1. Deze lijst is overigens eenvoudig uit te breiden.

Een andere eigenschap van de slimme database is dat niet alleen opgeslagen gegevens kunnen worden opgevraagd, maar ook gegevens die daarvan zijn afgeleid. Zo kan het totale stikstofgehalte in een watermonster worden geschat op basis van afzonderlijke stikstofcomponenten (bijvoorbeeld als de som van anorganisch stikstof en organisch stikstof). Een overzicht van de belangrijkste rekenregels die de database hiervoor gebruikt, zijn gegeven in bijlage C.2.

Een andere eigenschap van de slimme database is dat ze weet welke gegevens wel en welke niet mogen worden uitgeleverd. Hierdoor wordt het mogelijk om een deel van de database af te schermen ten behoeve van bijvoorbeeld validatie.

Conclusie

Het belang van slimme databases in de waterwereld, maar zeker ook daarbuiten, is evident. Slimme databases maken het voor gebruikers niet alleen eenvoudiger om met databases te werken, ook is de kwaliteit van de opgeslagen informatie beter gewaarborgd. Daarnaast kan een slimme database op basis van opgeslagen gegevens nieuwe gegevens afleiden. Hierdoor wordt informatie ontsloten die bij traditionele databases slechts latent aanwezig is.

Slimme databases zullen steeds slimmer worden naarmate ze intensiever worden gebruikt. Op basis van ervaringen met de gegenereerde uitvoer zullen nieuwe kwaliteitscontroles worden toegevoegd en bestaande controles worden verfijnd. Datzelfde geldt voor de rekenregels op basis waarvan afgeleide informatie wordt berekend.

Meer weten?

Meer informatie over de slimme database is opgenomen in de bijlagen. Bijlage A geeft een overzicht van de belangrijkste criteria op basis waarvan een relationeel database management systeem (RDBMS) is geselecteerd. Een RDBMS kan worden gezien als een raamwerk waarbinnen databases kunnen worden gedefinieerd. Vervolgens wordt in bijlage B het datamodel gepresenteerd. Het datamodel is als het ware de blauwdruk voor de database en geeft de relaties weer tussen de tabellen. Tot zover zijn er nog geen verschillen met traditionele databases. In bijlage C wordt vervolgens aangeven hoe de slimme

database zich onderscheidt van traditionele databases. In deze bijlage worden de geïmplementeerde kwaliteitscontroles en rekenregels besproken. In bijlage E tot slot wordt beschreven hoe met behulp van eenvoudige software gegevens kunnen worden ingelezen en opgevraagd.

Deel II

Bijlagen:
De database in meer detail

A Motivering van de keuze voor een relationeel database management systeem

Bij aanvang van het project is een aantal criteria opgesteld waaraan een relationeel database management systeem (RDBMS) moet voldoen om het project efficient te doen verlopen. Op basis van deze criteria is een geschikt RDBMS geselecteerd. De criteria kunnen als volgt worden samengevat:

- De RDBMS moet een client-server systeem zijn. Zo kunnen alle projectmedewerkers op elk moment (dus ook tegelijkertijd) vanaf een willekeurige computer de centrale database benaderen. Op deze wijze wordt bijvoorbeeld voorkómen dat meerdere versies van de database in omloop zijn;
- De RDBMS moet geschikt zijn voor het opslaan van (zeer) grote bestanden;
- De RDBMS moet eenvoudig zijn te beheren;
- De RDBMS moet gebruik kunnen maken van *stored procedures* en *triggers*. Op basis hiervan kan de functionaliteit van de database worden uitgebreid;
- De RDBMS moet goed zijn gedocumenteerd;
- De RDBMS moet een actieve gebruikersgroep hebben (mailing lists, etc.) ten behoeve van ondersteuning;
- De RDBMS moet bij voorkeur gratis beschikbaar zijn om de kosten beheersbaar te houden;
- De RDBMS moet kunnen worden beveiligd tegen ongeautoriseerd gebruik. De gegevensbestanden die worden opgeslagen zijn namelijk van derden afkomstig en moeten met de nodige zorgvuldigheid worden behandeld;
- De RDBMS moet betrouwbaar zijn (in database terminologie: transacties moeten '[ACID] *complient*' zijn);
- De RDBMS moet koppelbaar zijn met externe software (bijvoorbeeld geografische informatie sytemen ([GIS]), software voor visualisatie en data-analyse);
- De RDBMS moet geschikt zijn voor internetapplicaties.

Op grond van deze criteria is destijds gekozen voor MySQL (met de InnoDB-engine, `http://www.mysql.com`). Deze keuze is deels arbitrair. Andere systemen zoals PostgreSQL (`http://www.postgresql.org/`) voldoen ook aan de (meeste) criteria en hadden ook kunnen worden gekozen.

B Beschrijving van het datamodel

B.1 Inleiding

Met een datamodel (ook wel gegevensmodel genoemd) wordt beschreven hoe de gegevens in een informatiesysteem gestructureerd zijn. Het datamodel vormt als het ware het bouwplan van de database. Bij het opzetten van het datamodel is gepoogd om de structuur eenvoudig te houden[1]. Dit vereenvoudigt het opvragen van gegevens en het vullen van de database. Tevens moet het datamodel generiek zijn en niet project-specifiek. Het datamodel moet daarom in principe ook voor andere projecten kunnen worden gebruikt.

Ter wille van de transparantie, het onderhoud en de overdraagbaarheid zijn de volgende conventies gevolgd wat betreft de naamgeving van tabellen en velden[2].

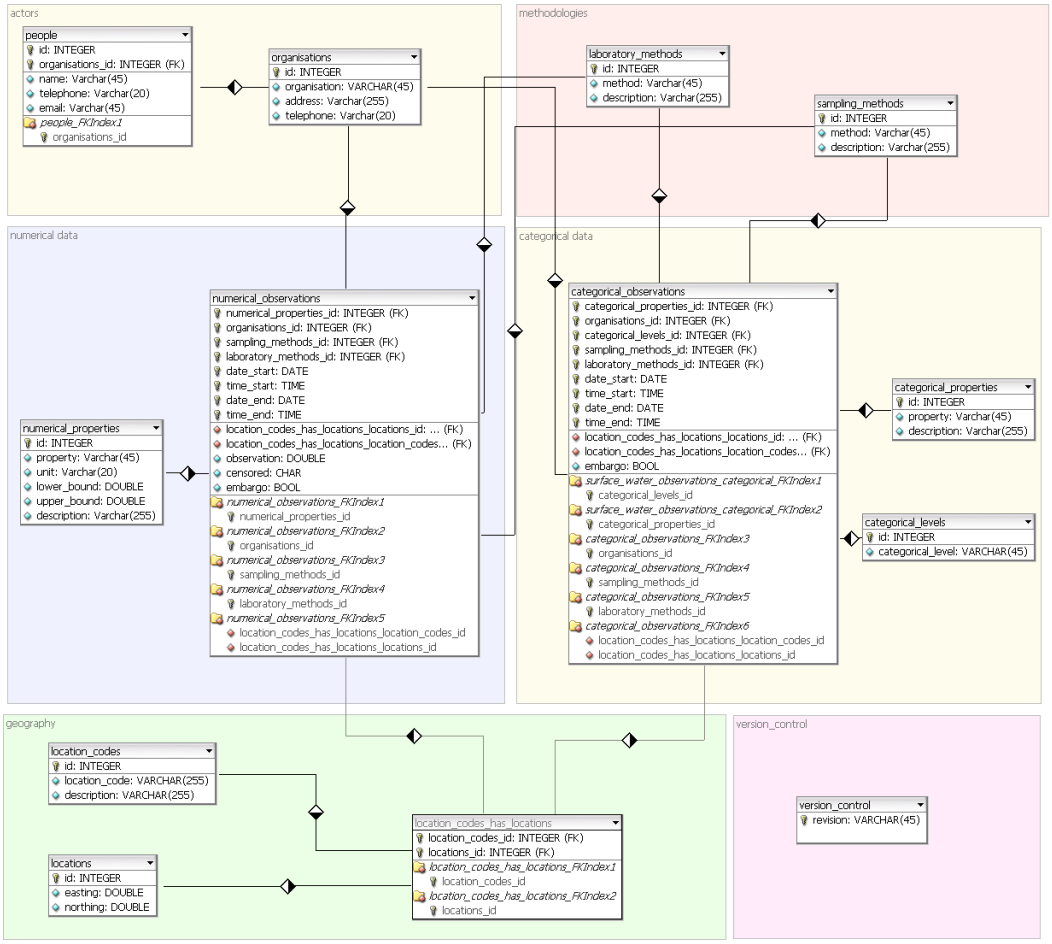
- Alle tabel- en veldnamen bestaan uit kleine letters ('*lower case*');
- Tabelnamen zijn in het meervoud gesteld, veldnamen in het enkelvoud;
- Als scheidingsteken tussen woorden wordt een underscore (_) gebruikt;
- Een *foreign key*[3] bestaat uit de naam van de tabel waarnaar wordt verwezen plus suffix '_id' (bijvoorbeeld '`organisations_id`' verwijst naar veld '`id`' van tabel '`organisations`'.

Het datamodel is schematisch weergegeven in figuur B.1. Een tabel wordt voorgesteld als een witte rechthoek. De tabelnaam is gegeven in een zilvergrijze balk bovenaan de rechthoek. Hieronder zijn de veldnamen gegeven, gevolgd door het gegevenstype. De relaties tussen de tabellen zijn weergegeven met zwarte lijnen. Tabellen zijn gegroepeerd in tabelgroepen. Deze tabelgroepen zijn weergegeven als gekleurde vlakken in figuur B.1. De volgende tabelgroepen kunnen worden onderscheiden:

[1] Er is bijvoorbeeld niet verder genormaliseerd dan de 'derde normaal vorm' (*third normal form (3NF)*)

[2] Een veld is een kolom in een databasetabel.

[3] Een veld dat verwijst naar een andere tabel.



Figuur B.1: Schematische weergave van het ontwikkelde datamodel.

`actors group`:	Deze groep bevat informatie over de herkomst van de gegevens
`geography group`:	Deze groep bevat geografische informatie over de metingen
`numerical data group`:	Deze groep bevat numerieke gegevens zoals concentraties
`categorical data group`:	Deze groep bevat categorische gegevens zoals de kleur van het slootwater
`methodologies group`:	Deze groep bevat informatie over de gebruikte laboratorium- en bemonsteringsmethoden
`version_control group`:	Deze groep bevat informatie t.b.v. het versiebeheer en wordt gewoonlijk onderhouden door het versiebeheerssysteem

In de onderstaande paragrafen worden de tabelgroepen kort beschreven.

B.1.1 Tabellengroep `actors`

Deze tabellengroep bevat informatie over de herkomst van de gegevens. Denk daarbij aan waterschappen, laboratoria, onderzoeksinstituten, enz.. De tabellengroep bestaat uit twee tabellen: één met adresgegevens van de organisatie die de metingen heeft aangeleverd (`organisations`), de ander met gegevens van de contactpersonen bij de betreffende organisaties (`persons`).

Tabel `organisations`

`id`:	Een uniek recordnummer
`organisation`:	Naam van de organisatie (waterschap, laboratorium, enz.)
`address`:	Adresgegevens van de betreffende organisatie
`telephone`:	Algemeen telefoonnummer van de betreffende organisatie

Tabel `people`

`id`:	Een uniek recordnummer
`organisations_id`:	Een foreign key
`name`:	Naam van de contactpersoon
`telephone`:	Telefoonnummer van de contactpersoon
`email`:	E-mailadres van de contactpersoon

B.1.2 Tabellengroep `geography`

Deze tabellengroep bevat geografische informatie. De database moet overweg kunnen met de volgende situaties:

1. Aan één locatie mogen meerdere meetcodes worden toegekend. Een locatie kan bijvoorbeeld door een waterschap met een andere code worden aangeduid dan door de provincie of een onderzoeksinstituut. Hierdoor is het mogelijk om alle meetgegevens op te vragen voor een bepaalde locatie, of slechts het deel dat door een specifieke organisatie is aangeleverd;
2. Op één locatie kunnen meerdere meetinstrumenten zijn geïnstalleerd (bijvoorbeeld meerdere peilbuizen met filters op verschillende dieptes);
3. Op één locatie kunnen zowel operationele als niet-operationele instrumenten (bijvoorbeeld afgedankte peilbuizen) staan;
4. Eén meetcode kan refereren naar verschillende locaties. Hoewel dit een onwenselijke situatie is, komt zij in de praktijk soms voor.

De `geography`-groep bevat de tabellen `location_codes`, `locations` en de koppeltabel `location_codes_has_locations`.

Tabel `location_codes`

`id`:	Een uniek recordnummer
`location_code`:	De codering van de meetlocatie. In het geval van een peilbuis, bijvoorbeeld '12BL0015'.
`description`:	Een beschrijving van de meetlocatie

Tabel `locations`

`id`:	Een uniek recordnummer.
`easting`:	De oostwaarde in het geografische Cartesische coördinatenstelsel
`northing`:	De westwaarde in het geografische Cartesische coördinatenstelsel

Merk op dat de maaiveldhoogte niet is opgenomen in bovenstaande tabel. Dit komt doordat maaiveldhoogte een tijdsafhankelijke variabele is en daarom in de `numerical_data` tabellengroep thuishoort.

Tabel `location_codes_has_locations`

`location_codes_id`:	Een foreign key
`location_id`:	Een foreign key

B.1.3 Tabellengroep `numerical_data`

Deze groep bestaat uit twee gerelateerde tabellen: `numerical_properties` en `numerical_observations` en wordt gebruikt om numerieke gegevens op te slaan.

Tabel `numerical_properties`

`id`	Een uniek recordnummer.
`property`:	De systeemeigenschappen. Voorbeelden zijn, de nitraatconcentratie in het oppervlaktewater, en de grondwaterstand t.o.v. het maaiveld.
`unit`:	De eenheid waarin `property` is uitgedrukt. Dit moet bij voorkeur een SI-eenheid zijn (`http://www.bipm.org/en/si/`). Het gebruik van percentages wordt sterk afgeraden, omdat die niet eenduidig interpreteerbaar zijn (denk aan volumepercentages versus gewichtspercentages). Gebruik als alternatief bijvoorbeeld g/g voor gewichtsfracties en m^3/m^3 voor volumefracties.
`lower_bound`:	De (fysische) ondergrens van `property`. Deze is bijvoorbeeld 0 (mg/l) in het geval van een concentratie. Een leeg veld wordt door de database geïnterpreteerd als $-\infty$.
`upper_bound`:	De (fysische) bovengrens van `property`. Deze is bijvoorbeeld 1 g/g in het geval van het organischestofgehalte en 14 voor de pH. Een leeg veld wordt door de database geïnterpreteerd als $+\infty$.
`description`:	Een korte omschrijving van `property`. Daarnaast biedt dit veld ruimte voor een verwijzing naar de gebruikte laboratoriummethode (NEN/ISO) en/of een schatting van de betrouwbaarheid van de meting.

Tabel `numerical_observations`

`numerical_properties_id`	Een foreign key
`organisations_id`	Een foreign key
`sampling_methods_id`	Een foreign key
`laboratory_methods_id`	Een foreign key
`date_start`	De datum waarop de waarneming is gestart. De datum is conform de ISO8601–standaard, dat wil zeggen `jjjj-mm-dd` (bijvoorbeeld: 2004-12-31).
`time_start`	Het tijdstip waarop de waarneming is gestart. Het tijdstip is conform de ISO8601-standaard, dat wil zeggen `UU:MM:SS` (bijvoorbeeld 21:03:10).
`date_end`	De datum waarop de waarneming is beëindigd (ISO8601–standaard)
`time_end`	Het tijdstip waarop de waarneming is beëindigd (ISO8601-standaard)
`location_codes_has_locations_locations_id`	Een foreign key
`location_codes_has_locations_location_codes_id`	Een foreign key
`observation`	De waarneming. Dit is een numerieke waarde groter dan of gelijk aan `lower_bound` en kleiner dan of gelijk aan `upper_bound`. Missing values kunnen worden opgeslagen als lege velden[a]
`censored`	Dit veld geeft aan of de waarde in veld `observation` al dan niet is gecensureerd. Een waarneming is gecensureerd als de waarde kleiner of groter is dan een bepaalde waarde. Een voorbeeld van een gecensureerde waarneming is een waarde beneden de detectielimiet. In dat geval wordt in het veld `observation` de detectielimiet ingevuld, en in het veld `censored` het teken '<'. Als een waarneming groter is dan een bepaalde waarde wordt het teken '>' gebruikt. Ongecensureerde waarnemingen dienen te worden aangeduid met '='.
`embargo`	Dit veld geeft aan of de betreffende record door een standaard gebruiker mag worden opgevraagd of niet. Dit veld wordt door de database automatisch ingevuld op basis van een 'trigger'-functie.

[a] Coderingen als -999 worden afgeraden omdat die altijd expliciet moeten worden uitgefilterd.

B.1.4 Tabellengroep `categorical data`

Deze groep bestaat uit drie gerelateerde tabellen: `categorical_properties`, `categorical_levels` en `categorical_observations`. `Levels`, ofwel niveaus, zijn voor categorische data wat `observations` zijn voor numerieke data. Een veld met eenheden is niet nodig omdat niveaus altijd dimensieloos zijn. Voorbeelden van niveaus zijn 'geel', 'grijs' 'geel/bruin' 'groen/grijs' voor de `property` 'kleur', en 'helder', 'zwak opalescent', 'opalescent' voor `property` 'helderheid'.

Tabel `categorical_properties`

`id`:	Een uniek recordnummer.
`property`:	De categorische systeemeigenschappen die in de database zijn opgenomen. Voorbeelden zijn de kleur (donker bruin, groen, ...) en de geur (rotte eieren, olieachtig, ...) van het oppervlaktewater.
`description`:	Een korte beschrijving van `property`

Tabel `categorical_levels`

`id`	Een uniek recordnummer.
`categorical_level`	De waarde van de niveaus ('geel', 'grijs', 'bruin', ...)

Tabel categorical_observations

categorical_properties_id	Een foreign key
organisations_id	Een foreign key
categorical_levels_id	Een foreign key
sampling_methods_id	Een foreign key
laboratory_methods_id	Een foreign key
date_start	De datum waarop de waarneming is gestart. De datum is conform de ISO8601–standaard, dat wil zeggen `jjjj-mm-dd` (bijvoorbeeld: 2004-12-31).
time_start	Het tijdstip waarop de waarneming is gestart. Het tijdstip is conform de ISO8601-standaard, dat wil zeggen `UU:MM:SS` (bijvoorbeeld 21:03:10).
date_end	De datum waarop de waarneming is beëindigd (ISO8601–standaard)
time_end	Het tijdstip waarop de waarneming is beëindigd (ISO8601-standaard)
location_codes_has_locations_locations_id	Een foreign key
location_codes_has_locations_location_codes_id	Een foreign key
embargo	Dit veld geeft aan of de betreffende record door een standaard gebruiker mag worden opgevraagd of niet. Dit veld wordt door de database automatisch ingevuld op basis van een 'trigger'-functie.

B.1.5 Tabellengroep `methodologies`

Deze tabellengroep bevat informatie over de gebruikte veldmethode (steekmonster, debietproportionele meting, molenmeting, informatie over steekproefontwerp, etc.) en laboratoriummethode (vlam–AAS, etc.).

Tabel `sampling_methods`

`id`:	Een uniek recordnummer
`method`:	Naam van de methode
`description`:	Omschrijving van de methode

Tabel `laboratory_methods`

`id`:	Een uniek recordnummer
`method`:	Naam van de methode
`description`:	Omschrijving van de methode

B.1.6 Tabellengroep `verion_control`

Deze tabellengroep bevat de versie van de database.

Tabel `verion_control`

`revision`:	databaserevisie volgens het versiebeheerssysteem

C Eigenschappen van een slimme database

De ontwikkelde database onderscheidt zich van andere databases door de toegevoegde functionaliteit. Deze functionaliteit bestaat ruwweg uit twee delen:

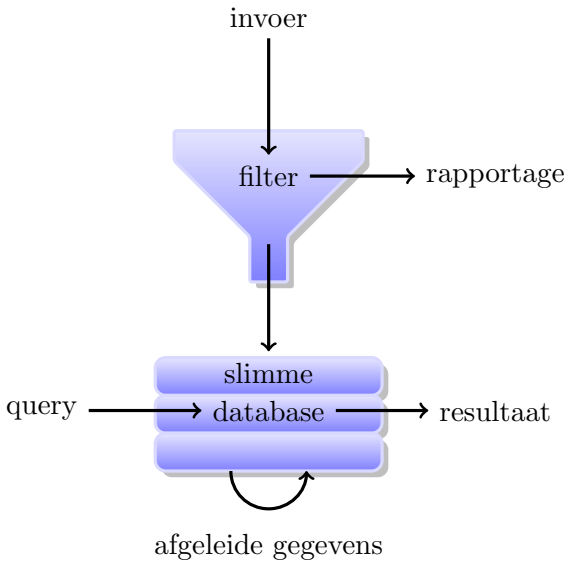
1. Kwaliteitscontrole aan de *invoerzijde*. Alvorens gegevens worden opgeslagen in de database worden ze eerste uitgebreid gecontroleerd op mogelijke fouten (zie bovenste helft van figuur C.1);
2. Toevoeging van afgeleide waarden aan de *uitvoerzijde*. Niet alleen gegevens die in de database zijn opgeslagen kunnen worden opgevraagd, ook gegevens die daarvan zijn afgeleid (zie onderste helft van figuur C.1).

Deze punten worden in de onderstaande paragrafen nader toegelicht.

C.1 Kwaliteitscontrole

Voordat de gegevens in de database worden opgeslagen, worden ze gecontroleerd op mogelijke fouten. De volgende kwaliteitscontroles worden uitgevoerd:

1. Controle op het formaat van het invoerbestand (zie ook figuur E.1 in bijlage E);
2. Controle op unieke waarden. Meetgegevens die abusievelijk meerdere keren voorkomen, worden verwijderd. De database zal hierdoor alleen unieke records bevatten;
3. Controle op overbodige spaties. Overbodige spaties worden uit de invoerbestanden verwijderd. Hetzelfde geldt voor zogenaamde 'control characters'. Dit zijn tekens die soms in ASCII-bestanden voorkomen, maar die niet altijd zichtbaar zijn. 'Control characters' kunnen tot vervelende effecten leiden.
4. Controle op de consistentie van de metingen. Een instantie (zoals een waterschap of laboratorium) mag voor een gegeven locatie en meetperiode slechts één waarde voor een systeemeigenschap (zoals een concentratie of een waterafvoer) aanleveren, in een bepaalde eenheid en bepaald volgens een specifieke veld- en laboratoriummethode. Tevens wordt gecontroleerd of de volgende ongelijkheden opgaan:



Figuur C.1: Opzet van de slimme database. De invoer wordt gecontroleerd op fouten met een filter. De resultaten van deze kwaliteitscontrole worden gerapporteerd als web-document. Correcte gegevens worden in de database geladen, foute gegevens worden teruggekoppeld met de waterschappen. Gegevens kunnen worden opgevraagd met speciaal ontwikkelde software. Op basis van de opgeslagen gegevens kan de database afgeleide gegevens berekenen.

- ortho–fosfaat–P + organisch P < totaal P
- ortho–fosaat–P < totaal P
- organisch P < totaal P
- ammonium–N + nitraat–N + nitriet–N < totaal N
- nitraat–N + nitriet–N < totaal N
- ammonium–N + nitraat–N < totaal N
- ammonium–N + nitriet–N < totaal N
- ammonium–N < totaal N
- nitraat < totaal N
- nitriet < totaal N
- Kjeldahl–N < totaal N
- organisch N < totaal N
- Kjeldahl N + nitraat–N < totaal N
- organisch N < Kjeldahl–N
- ammonium–N < Kjeldahl–N
- organisch–N + ammonium–N < Kjeldahl–N

Door meetfouten hoeven deze ongelijkheden niet altijd op te gaan. Daarom wordt een zekere tolerantie in acht genomen (bijvoorbeeld: 0.05 mg/l als absolute afwijking of 5% als relatieve afwijking);

5. Controle op het datum- en tijdformaat. Deze moeten conform de ISO8601–standaard (`jjjj-mm-dd UU:MM:SS`) zijn aangeleverd. Dataloggers maken veelal geen onderscheid tussen zomer- en wintertijd. De database houdt hier expliciet rekening mee. Tevens wordt gecontroleerd of meetperiodes wel in het verleden liggen (metingen in de toekomstig zijn namelijk niet mogelijk);

6. Controle op de eenheden van numerieke eigenschappen ('`numerical_properties`') en de niveaus van categorische eigenschappen ('`categorical_properties`');

7. Controle op het bereik van de meetwaarden. Hiermee wordt voorkomen dat meetwaarden die buiten fysische grenzen vallen (zoals bijvoorbeeld negatieve concentraties) in de database worden opgenomen. Ook worden zo extreem hoge waarden ('uitbijters') buiten de database gehouden;

8. Controle of de locatiecodes, de coördinaten, de namen van de organisaties, de systeemeigenschappen, de eenheden en de veld- en laboratoriummethodes reeds zijn voorgedefinieerd in de database. Dit verkleint de kans op typefouten. Typefouten hebben tot gevolg dat een deel van de gegevens niet meer opvraagbaar is.

C.2 Afgeleide waarden

Soms is een meetwaarde niet aanwezig in de database maar kan wel worden berekend op basis van beschikbare componenten. Een voorbeeld is het schatten van totaal stikstof op basis van anorganische en organische stikstofcomponenten. Als een gebruiker ervoor kiest om ontbrekende componenten te laten schatten, zullen zowel opgeslagen componenten als geschatte componenten naar het uitvoerbestand worden weggeschreven. Een kolom in dit uitvoerbestand geeft aan of sprake is van een aangeleverde (veelal gemeten) waarde of een geschatte waarde (zie figuur E.3 in bijlage E).

De geïmplementeerde rekenregels zijn in onderstaande '*stored procedures*' gegeven.

Listing C.1: Rekenregels voor het berekenen van nitraatstikstof.

```
   /* $Id: monitoring_stroomgebieden_function_estimate_N_NO3.sql 29 2008-10-15 10:23:45Z walvo001 $ */

DELIMITER //

USE monitoring_stroomgebieden //


/*  Computation of nitrate nitrogen: N_NO3

    N_NO3 = N_NO2NO3 - N_NO2

    approximation(s):
    N_NO3 = N_NO2NO3

*/
DROP FUNCTION IF EXISTS estimate_N_NO3//
CREATE FUNCTION estimate_N_NO3(N_NO2NO3 DOUBLE, N_NO2 DOUBLE) RETURNS DOUBLE

BEGIN

  /* declarations */
  DECLARE N_NO3 DOUBLE DEFAULT NULL;
  DECLARE has_N_NO2NO3, has_N_NO2 TINYINT(1);

  /* set Booleans */
  SET has_N_NO2NO3 = NOT (N_NO2NO3 IS NULL);
  SET has_N_NO2    = NOT (N_NO2 IS NULL);

  /* estimate nitrate nitrogen */
  IF (has_N_NO2NO3) THEN
    IF (has_N_NO2) THEN
      BEGIN
        SET N_NO3 = N_NO2NO3 - N_NO2;
        IF (N_NO3 < 0) THEN
          SET N_NO3 = 0;
        END IF;
      END;
    ELSE
      SET N_NO3 = N_NO2NO3;
    END IF;
  END IF;

  /* return result */
  RETURN N_NO3;

END//

DELIMITER ;
```

Listing C.2: Rekenregels voor het berekenen van anorganisch stikstof.

```
   /* $Id: monitoring_stroomgebieden_function_estimate_N_inorganic.sql 29 2008-10-15 10:23:45Z walvo001 $ */

DELIMITER //

USE monitoring_stroomgebieden //


/*  Computation of inorganic nitrogen content

    N_inorganic = N_NO3 + N_NO2 + N_NH4

    approximations:
    N_inorganic = N_NO3 + N_NO2 + N_NH4
    N_inorganic = N_NO3NO2 + N_NH4
    N_inorganic = N_NO3 + N_NO2
    N_inorganic = N_NO2NO3
    N_inorganic = N_NO3
*/
DROP FUNCTION IF EXISTS estimate_N_inorganic//
CREATE FUNCTION estimate_N_inorganic(N_NH4 DOUBLE, N_NO2 DOUBLE, N_NO3 DOUBLE, N_NO2NO3 DOUBLE) RETURNS DOUBLE

BEGIN

  /* declarations */
  DECLARE N_inorganic DOUBLE DEFAULT NULL;
  DECLARE has_N_NO3, has_N_NO2, has_N_NO2NO3, has_N_NH4 TINYINT(1);

  /* set Booleans */
  SET has_N_NO2    = NOT (N_NO2 IS NULL);
  SET has_N_NO2NO3 = NOT (N_NO2NO3 IS NULL);
  SET has_N_NO3    = NOT (N_NO3 IS NULL);
  SET has_N_NH4    = NOT (N_NH4 IS NULL);

  /* estimate inorganic nitrogen */
  IF (has_N_NO2NO3 OR has_N_NO3) THEN
    IF (has_N_NO2NO3) THEN
      SET N_inorganic = N_NO2NO3;
    ELSE /* has_N_NO3 */
      SET N_inorganic = N_NO3;
      IF (has_N_NO2) THEN
        SET N_inorganic = N_inorganic + N_NO2;
      END IF;
    END IF;
    IF (has_N_NH4) THEN
      SET N_inorganic = N_inorganic + N_NH4;
    END IF;
  END IF;

  /* return result */
  RETURN N_inorganic;

END//

DELIMITER ;
```

Listing C.3: Rekenregels voor het berekenen van organisch stikstof.

```
1   /* $Id: monitoring_stroomgebieden_function_estimate_N_organic.sql 29 2008-10-15 10:23:45Z walvo001 $ */
2
3  DELIMITER //
4
5  USE monitoring_stroomgebieden //
6
7
8  /*  Computation of organic nitrogen content
9
10     N_organic = N_kjeldahl - N_NH4
11
12     approximations:
13     N_organic = N_kjeldahl - N_NH4
14     N_organic = N_kjeldahl
15
16 */
17 DROP FUNCTION IF EXISTS estimate_N_organic//
18 CREATE FUNCTION estimate_N_organic(N_kjeldahl DOUBLE, N_NH4 DOUBLE) RETURNS DOUBLE
19
20 BEGIN
21
22   /* declarations */
23   DECLARE N_organic DOUBLE DEFAULT NULL;
24   DECLARE has_N_kjeldahl, has_N_NH4 TINYINT(1);
25
26   /* set Booleans */
27   SET has_N_kjeldahl = NOT (N_kjeldahl IS NULL);
28   SET has_N_NH4      = NOT (N_NH4 IS NULL);
29
30   /* estimate organic nitrogen */
31   IF (has_N_kjeldahl) THEN
32     SET N_organic = N_kjeldahl;
33     IF (has_N_NH4) THEN
34       SET N_organic = N_organic - N_NH4;
35       IF (N_organic < 0) THEN
36         SET N_organic = 0;
37       END IF;
38     END IF;
39   END IF;
40
41   /* return result */
42   RETURN N_organic;
43
44 END//
45
46 DELIMITER ;
```

Listing C.4: Rekenregels voor het berekenen van totaal stikstof.

```
   /* $Id: monitoring_stroomgebieden_function_estimate_N_total.sql 29 2008-10-15 10:23:45Z walvo001 $ */

DELIMITER //

USE monitoring_stroomgebieden //


/*  Computation of total nitrogen content

    N_total = N_kjeldahl + N_NO3 + N_NO2  = N_organic + N_NH4 + N_NO3 + N_NO2

    approximations:
    N_total = N_kjeldahl + N_NO3
    N_total = N_kjeldahl + N_NO3 + N_NO2
    N_total = N_kjeldahl + N_NO2NO3
    N_total = N_organic  + N_NO3 + N_NH4
    N_total = N_organic  + N_NO3 + N_NO2 + N_NH4
    N_total = N_organic  + N_NO2NO3 + N_NH4
    N_total = N_organic  + N_NO3
    N_total = N_organic  + N_NO3 + N_NO2
    N_total = N_organic  + N_NO2NO3
*/
DROP FUNCTION IF EXISTS estimate_N_total//
CREATE FUNCTION estimate_N_total(N_NH4 DOUBLE, N_NO2 DOUBLE, N_NO3 DOUBLE,
  N_NO2NO3 DOUBLE, N_Kjeldahl DOUBLE, N_organic DOUBLE) RETURNS DOUBLE

BEGIN

  /* declarations */
  DECLARE N_total DOUBLE DEFAULT NULL;
  DECLARE hasOrganic, hasKjeldahl, hasNO3, hasNO2,
    hasNO2NO3, hasNH4, hasTotal TINYINT(1);

  /* set Booleans */
  SET hasOrganic  = NOT (N_organic  IS NULL);
  SET hasKjeldahl = NOT (N_kjeldahl IS NULL);
  SET hasNO3      = NOT (N_NO3 IS NULL);
  SET hasNO2      = NOT (N_NO2 IS NULL);
  SET hasNO2NO3   = NOT (N_NO2NO3 IS NULL);
  SET hasNH4      = NOT (N_NH4 IS NULL);
  SET hasTotal    = (hasOrganic OR hasKjeldahl) AND
                    (hasNO2NO3 OR hasNO3);

  /* estimate total nitrogen */
  IF (hasTotal) THEN
    IF (hasNO2NO3) THEN
      SET N_total = N_NO2NO3;
    ELSE /* hasNO3 */
      SET N_total = N_NO3;
      IF (hasNO2) THEN
        SET N_total = N_total + N_NO2;
      END IF;
    END IF;
    IF (hasKjeldahl) THEN
      SET N_total = N_total + N_kjeldahl;
    ELSE /* has N_organic */
      SET N_total = N_total + N_organic;
      IF (hasNH4) THEN
        SET N_total = N_total + N_NH4;
      END IF;
    END IF;
  END IF;

  /* return result */
  RETURN N_total;

END//

DELIMITER ;
```

Listing C.5: Rekenregels voor het berekenen van organisch fosfor.

```
   /* $Id: monitoring_stroomgebieden_function_estimate_P_organic.sql 29 2008-10-15 10:23:45Z walvo001 $ */

DELIMITER //

USE monitoring_stroomgebieden //


/*  Computation of organic phosphorus content

    P_organic = P_total - P_ortho (assuming that P_total contains no particulate P)

    approximations:
    P_organic = P_total - P_ortho

*/
DROP FUNCTION IF EXISTS estimate_P_organic//
CREATE FUNCTION estimate_P_organic(P_total DOUBLE, P_ortho DOUBLE) RETURNS DOUBLE

BEGIN

  /* declarations */
  DECLARE P_organic DOUBLE DEFAULT NULL;
  DECLARE has_P_total, has_P_ortho TINYINT(1);

  /* set Booleans */
  SET has_P_total = NOT (P_total IS NULL);
  SET has_P_ortho = NOT (P_ortho IS NULL);

  /* estimate organic nitrogen */
  IF ((has_P_total) AND (has_P_ortho)) THEN
    SET P_organic = P_total - P_ortho;
    IF (P_organic < 0) THEN
      SET P_organic = 0;
    END IF;
  END IF;

  /* return result */
  RETURN P_organic;

END//

DELIMITER ;
```

Listing C.6: Rekenregels voor het berekenen van totaal fosfor.

```
1    /* $Id: monitoring_stroomgebieden_function_estimate_P_total.sql 29 2008-10-15 10:23:45Z walvo001 $ */
2
3  DELIMITER //
4
5  USE monitoring_stroomgebieden //
6
7
8  /*  Computation of total phosphorus content
9
10     P_total = P_organic + P_inorganic (assuming that P_total contains no particulate P)
11
12     approximations:
13     P_total = P_organic + P_inorganic
14     P_total = P_organic + P_ortho
15
16 */
17 DROP FUNCTION IF EXISTS estimate_P_total//
18 CREATE FUNCTION estimate_P_total(P_organic DOUBLE, P_inorganic DOUBLE, P_ortho DOUBLE) RETURNS DOUBLE
19
20 BEGIN
21
22   /* declarations */
23   DECLARE P_total DOUBLE DEFAULT NULL;
24   DECLARE has_P_organic, has_P_inorganic, has_P_ortho TINYINT(1);
25
26   /* set Booleans */
27   SET has_P_organic = NOT (P_organic IS NULL);
28   SET has_P_inorganic = NOT (P_inorganic IS NULL);
29   SET has_P_ortho = NOT (P_ortho IS NULL);
30
31   /* estimate total phosphorus */
32   IF ((has_P_organic) AND ((has_P_ortho) OR (has_P_inorganic))) THEN
33     IF (has_P_ortho) THEN
34         SET P_total = P_organic + P_ortho;
35     ELSE
36         SET P_total = P_organic + P_inorganic;
37     END IF;
38   END IF;
39
40   /* return result */
41   RETURN P_total;
42
43 END//
44
45 DELIMITER ;
```

D Beschikbare gegevens

In de database worden een groot aantal oppervlaktewaterkarakteristieken opgeslagen. Een overzicht van het aantal meetlocaties, het aantal meetjaren en het aantal waarnemingen per stroomgebied en oppervlaktewaterkarakteristiek is gegeven in tabel D.1. Deze tabel geeft de status van de database weer voor juni 2009.

Tabel D.1: Het aantal meetlocaties, het aantal meetjaren en het totale aantal waarnemingen uitgesplitst naar stroomgebied (dra: de Drentse Aa, krw: de Krimpenerwaard, qvu: Quarles van Ufford, sbk: de Schuitenbeek) en oppervlaktewaterkarakteristiek (linker kolom).

		aantal locaties				aantal jaren				aantal waarnemingen			
variabele	eenheid	dra	krw	qvu	sbk	dra	krw	qvu	sbk	dra	krw	qvu	sbk
BZV5	mg/l	30	0	23	0	3	0	4	0	11	0	63	0
Ca	mg/l	4	0	23	18	3	0	2	7	16	0	76	101
CHLFa	µg/l	37	92	33	28	29	19	19	12	69	198	110	71
Cl	mg/l	37	116	23	34	29	30	9	33	65	236	357	258
DG	mg/l	0	0	4	0	0	0	1	0	0	0	4	0
discharge	m3/s	7	16	9	21	23	13	4	34	3653	1557	18186	3947
DOC	mg/l	0	0	6	4	0	0	1	3	0	0	36	11
EC	mS/m	37	99	32	38	10	29	19	21	141	134	97	642
GR	%	0	0	19	0	0	0	2	0	0	0	129	0
HCO3	mg/l	30	103	23	21	8	19	5	6	208	245	162	87
Mg	mg/l	4	0	23	18	3	0	2	7	13	0	65	81
N_inorganic	mg/l	42	116	33	37	29	30	19	36	1421	1152	822	1034
N_kjeldahl	mg/l	42	116	32	37	29	30	16	36	100	220	81	476
N_NH3	mg/l	30	0	0	0	3	0	0	0	2	0	0	0
N_NH4	mg/l	42	116	33	37	29	30	19	36	194	189	132	326
N_NO2	mg/l	35	78	33	37	28	7	19	33	78	97	57	70
N_NO2NO3	mg/l	0	100	0	31	0	29	0	15	0	319	0	194
N_NO3	mg/l	42	116	33	37	29	30	19	36	437	913	376	635
N_organic	mg/l	42	116	32	37	29	30	16	36	386	501	356	445
N_total	mg/l	42	116	33	38	29	30	19	36	820	1224	595	1116
Na	mg/l	4	0	30	17	3	0	10	4	10	0	50	62
O2	mg/l	37	113	32	38	6	30	19	29	140	241	214	245
P_inorganic	mg/l	42	116	32	37	29	30	13	34	63	331	45	154
P_organic	mg/l	42	116	32	37	29	30	13	34	169	335	162	204
P_ortho	mg/l	42	116	32	37	29	30	13	34	63	331	45	154
P_total	mg/l	42	116	33	38	29	30	19	36	74	296	114	203
pH	-	37	114	33	38	29	30	19	31	158	71	38	287
SO4	mg/l	30	103	33	34	19	18	19	23	69	238	760	312
TOC	mg/l	0	0	22	1	0	0	1	1	0	0	132	1

E Gebruikershandleiding

Voor het uitwisselen van gegevens tussen de gebruiker en de database is speciale software ontwikkeld. In deze software is de functionaliteit geïmplementeerd die in bijlage C is beschreven. De software is geïmplementeerd in de R-package `msdb`. De belangrijkste functies in deze package zijn `csv2db` voor het opslaan van gegevens in de database volgens de procedures in bijlage C.1, en `db2csv` voor het opvragen van gegevens waarbij ook afgeleide gegevens kunnen worden berekend volgens de rekenregels in bijlage C.2.

E.1 Opslaan van gegevens in de database

De R-functie `csv2db` kan worden gebruikt om gegevens in de database te laden. Hierbij wordt de kwaliteitscontrole die beschreven staat in bijlage C.1 automatisch uitgevoerd. De functie `csv2db` maakt deel uit van de `msdb`-package die (na installatie) kan worden geladen met

```
library(msdb)
```

Een uitgebreide beschrijving van alle argumenten van functie `csv2db` en een werkend voorbeeld kan worden verkregen met

```
?csv2db
```

De functie `csv2db` leest de gegevens uit een csv-bestand[1]. In figuur E.1 is een voorbeeld gegeven van een csv–bestand dat als invoer kan dienen voor `csv2db`. Het bestand bestaat uit dertien door komma's gescheiden kolommen. De betekenis van deze kolommen is gegeven in tabel E.1.

Dergelijke bestanden zijn eenvoudig aan te maken met bijvoorbeeld MS-Excel. De functie `csv2db` zal dit bestand alleen importeren als er geen fouten worden aangetroffen. Fouten en inconsistenties worden gerapporteerd in de vorm van een HTML-document (figuur E.2). Doordat gebruik is gemaakt van hyperlinks kan efficient van fout naar fout worden genavigeerd. Het HTML-document kan worden gelezen met een web-browser (bijvoorbeeld: Microsoft Internet Explorer, Mozilla FireFox, Opera of Safari).

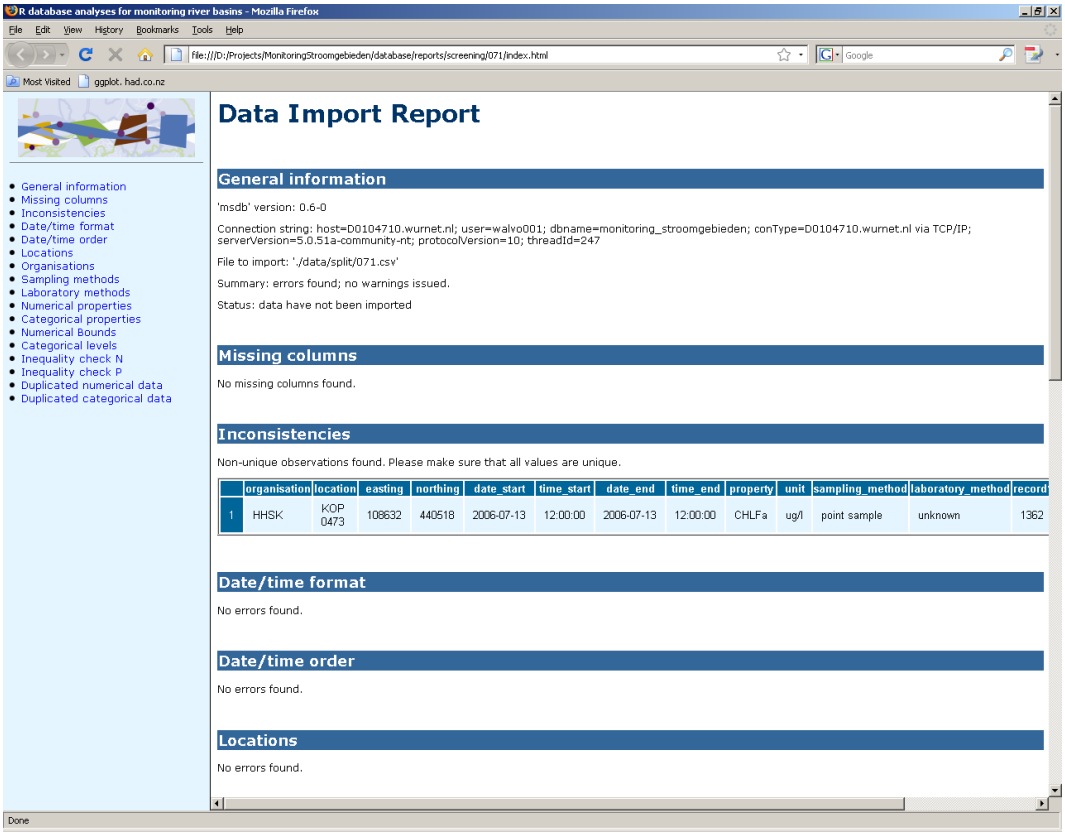
[1] Een ASCII-bestand waarin de data zijn opgeslagen in tabelvorm. De kolommen zijn gescheiden door komma's. Teksten staan tussen dubbele of enkele quotes.

```
"organisation","location","easting","northing","date_start","time_start","date_end","time_end","property","unit","value","sampling_method","laboratory_method"
"Waterschap Hunze en Aas","Schipborg",240432,566009,"1982-02-03","00:00:00","1982-02-03","23:59:59","discharge","m3/s","1.26","stage-discharge","none"
"Waterschap Hunze en Aas","Schipborg",240432,566009,"1982-02-04","00:00:00","1982-02-04","23:59:59","discharge","m3/s","1.19","stage-discharge","none"
"Waterschap Hunze en Aas","Schipborg",240432,566009,"1982-02-05","00:00:00","1982-02-05","23:59:59","discharge","m3/s","1.05","stage-discharge","none"
"Waterschap Hunze en Aas","Schipborg",240432,566009,"1982-02-06","00:00:00","1982-02-06","23:59:59","discharge","m3/s","2.5","stage-discharge","none"
"Waterschap Hunze en Aas","Schipborg",240432,566009,"1982-02-07","00:00:00","1982-02-07","23:59:59","discharge","m3/s","3.27","stage-discharge","none"
"Waterschap Hunze en Aas","Schipborg",240432,566009,"1982-02-08","00:00:00","1982-02-08","23:59:59","discharge","m3/s","3.99","stage-discharge","none"
"ZHEM","KOP 0501",110370,446565,"1982-02-08","10:30:00","1982-02-08","10:30:00","C1","mg/l","38","point sample","unknown"
"ZHEM","KOP 0501",110370,446565,"1982-02-08","10:30:00","1982-02-08","10:30:00","EC","mS/m","59","point sample","unknown"
"ZHEM","KOP 0501",110370,446565,"1982-02-08","10:30:00","1982-02-08","10:30:00","N_kjeldahl","mg/l","6.4","point sample","unknown"
"ZHEM","KOP 0501",110370,446565,"1982-02-08","10:30:00","1982-02-08","10:30:00","N_NH4","mg/l","3.8","point sample","unknown"
"ZHEM","KOP 0501",110370,446565,"1982-02-08","10:30:00","1982-02-08","10:30:00","N_NO2NO3","mg/l","1.35","point sample","unknown"
"ZHEM","KOP 0501",110370,446565,"1982-02-08","10:30:00","1982-02-08","10:30:00","O2","mg/l","9","point sample","unknown"
"ZHEM","KOP 0501",110370,446565,"1982-02-08","10:30:00","1982-02-08","10:30:00","P_ortho","mg/l","0.24","point sample","unknown"
"ZHEM","KOP 0501",110370,446565,"1982-02-08","10:30:00","1982-02-08","10:30:00","P_total","mg/l","0.6","point sample","unknown"
"ZHEM","KOP 0501",110370,446565,"1982-02-08","10:30:00","1982-02-08","10:30:00","pH","-","7.3","point sample","unknown"
"ZHEM","KOP 1005",106693,436399,"1982-02-08","12:30:00","1982-02-08","12:30:00","C1","mg/l","79","point sample","unknown"
"ZHEM","KOP 1005",106693,436399,"1982-02-08","12:30:00","1982-02-08","12:30:00","EC","mS/m","67","point sample","unknown"
"ZHEM","KOP 1005",106693,436399,"1982-02-08","12:30:00","1982-02-08","12:30:00","N_kjeldahl","mg/l","6.9","point sample","unknown"
"ZHEM","KOP 1005",106693,436399,"1982-02-08","12:30:00","1982-02-08","12:30:00","N_NH4","mg/l","4.5","point sample","unknown"
"ZHEM","KOP 1005",106693,436399,"1982-02-08","12:30:00","1982-02-08","12:30:00","N_NO2NO3","mg/l","1.05","point sample","unknown"
"ZHEM","KOP 1005",106693,436399,"1982-02-08","12:30:00","1982-02-08","12:30:00","O2","mg/l","8.3","point sample","unknown"
"ZHEM","KOP 1005",106693,436399,"1982-02-08","12:30:00","1982-02-08","12:30:00","P_ortho","mg/l","0.66","point sample","unknown"
"ZHEM","KOP 1005",106693,436399,"1982-02-08","12:30:00","1982-02-08","12:30:00","P_total","mg/l","0.96","point sample","unknown"
"ZHEM","KOP 1005",106693,436399,"1982-02-08","12:30:00","1982-02-08","12:30:00","pH","-","7.25","point sample","unknown"
"Waterschap Hunze en Aas","Schipborg",240432,566009,"1982-02-09","00:00:00","1982-02-09","23:59:59","discharge","m3/s","1.8","stage-discharge","none"
"Waterschap Veluwe","25201",165188,475775,"1982-02-09","09:30:00","1982-02-09","09:30:00","C1","mg/l","53","point sample","unknown"
"Waterschap Veluwe","25201",165188,475775,"1982-02-09","09:30:00","1982-02-09","09:30:00","N_kjeldahl","mg/l","4.1","point sample","unknown"
"Waterschap Veluwe","25201",165188,475775,"1982-02-09","09:30:00","1982-02-09","09:30:00","N_NH4","mg/l","2.2","point sample","unknown"
"Waterschap Veluwe","25201",165188,475775,"1982-02-09","09:30:00","1982-02-09","09:30:00","N_NO2NO3","mg/l","4.8","point sample","unknown"
"Waterschap Veluwe","25201",165188,475775,"1982-02-09","09:30:00","1982-02-09","09:30:00","O2","mg/l","9.8","point sample","unknown"
"Waterschap Veluwe","25201",165188,475775,"1982-02-09","09:30:00","1982-02-09","09:30:00","P_ortho","mg/l","0.21","point sample","unknown"
"Waterschap Veluwe","25201",165188,475775,"1982-02-09","09:30:00","1982-02-09","09:30:00","P_total","mg/l","0.54","point sample","unknown"
"Waterschap Veluwe","25201",165188,475775,"1982-02-09","09:30:00","1982-02-09","09:30:00","pH","-","7.2","point sample","unknown"
"Waterschap Veluwe","25201",165188,475775,"1982-02-09","09:30:00","1982-02-09","09:30:00","discharge","m3/s","0.564","ott current meter","none"
"Waterschap Hunze en Aas","Schipborg",240432,566009,"1982-02-10","00:00:00","1982-02-10","23:59:59","discharge","m3/s","1.46","stage-discharge","none"
"Waterschap Hunze en Aas","Schipborg",240432,566009,"1982-02-11","00:00:00","1982-02-11","23:59:59","discharge","m3/s","1.33","stage-discharge","none"
"Waterschap Hunze en Aas","Schipborg",240432,566009,"1982-02-12","00:00:00","1982-02-12","23:59:59","discharge","m3/s","1.38","stage-discharge","none"
"Waterschap Hunze en Aas","Schipborg",240432,566009,"1982-02-13","00:00:00","1982-02-13","23:59:59","discharge","m3/s","1.31","stage-discharge","none"
"Waterschap Hunze en Aas","Schipborg",240432,566009,"1982-02-14","00:00:00","1982-02-14","23:59:59","discharge","m3/s","1.21","stage-discharge","none"
"Waterschap Hunze en Aas","Schipborg",240432,566009,"1982-02-15","00:00:00","1982-02-15","23:59:59","discharge","m3/s","1.14","stage-discharge","none"
```


Figuur E.1: Voorbeeld van een csv-bestand met invoergegevens.

Tabel E.1: Betekenis van de kolommen in het csv-bestand met invoergegevens voor de database.

kolomnaam	omschrijving
`organisation`:	naam van de instantie die de betreffende gegevens heeft aangeleverd
`location`:	locatiecode van het meetpunt
`easting`:	oostwaarde in het geografische Cartesische coördinatenstelsel
`northing`:	westwaarde in het geografische Cartesische coördinatenstelsel
`date_start`:	De datum waarop de waarneming is gestart. De datum is conform de ISO8601–standaard, dat wil zeggen `jjjj-mm-dd` (bijvoorbeeld: 2004-01-24).
`time_start`	Het tijdstip waarop de waarneming is gestart. Het tijdstip is conform de ISO8601-standaard, dat wil zeggen `UU:MM:SS` (bijvoorbeeld 21:03:10).
`date_end`:	De datum waarop de waarneming is beëindigd (ISO8601–standaard)
`time_end`:	Het tijdstip waarop de waarneming is beëindigd (ISO8601-standaard)
`property`:	De systeemeigenschappen (bijvoorbeeld N_total)
`unit`:	De eenheid waarin `property` is uitgedrukt
`value`:	De waarneming. Dit is een numerieke waarde groter dan of gelijk aan `lower_bound` en kleiner dan of gelijk aan `upper_bound`. Missing values kunnen worden opgeslagen als lege velden
`sampling_method`:	De gebruikte veldmethode (molenmeting, steekmonster, debietproportionele meting)
`laboratory_method`:	De gebruikte laboratoriummethode (vlam-AAS, etc.)



Figuur E.2: Voorbeeld van een (deel van een) HTML*-rapport met daarin een overzicht van aangetroffen fouten in het invoerbestand. In het voorbeeld wordt Mozilla Firefox gebruikt als browser.*

E.2 Opvragen van gegevens uit de database

De eenvoudigste wijze om gegevens op te vragen is met de R-functie `db2csv`. Deze functie maakt deel uit van de `msdb`-package en kan worden geladen met

```
library(msdb)
```

Een overzicht van de argumenten van de functie `db2csv`, inclusief een werkend voorbeeld, kan worden verkregen met

```
?db2csv
```

De functie `db2csv` maakt het niet alleen mogelijk om gegevens op te vragen die in de database zijn opgeslagen. Het is tevens mogelijk om op basis van de beschikbare gegevens *afgeleide* gegevens te laten uitvoeren (zoals het totale stikstofgehalte, of het totale gehalte aan mineraal stikstof).

De databasebeheerder kan bovendien delen van de database afschermen voor de gebruikers. Zo kan een deel van de gegevens worden achtergehouden voor bijvoorbeeld validatie.

Een voorbeeld van een uitvoerbestand is gegeven in figuur E.3. De meeste kolommen komen overeen met die van het invoerbestand (zie tabel E.1). De laatste twee kolommen zijn nieuw.

De kolom '`embargo`' geeft aan of de betreffende record (regel) al dan niet mag worden uitgevoerd. Voor de gewone gebruiker zal deze kolom altijd de waarde 0 hebben. Dat wil zeggen dat de betreffende records mogen worden uitgeleverd. '`embargo`' kan ook gelijk zijn aan 1. In dat geval staan de records onder embargo en kunnen ze alleen worden opgevraagd door personen die hiertoe gemachtigd zijn.

De kolom '`origin`' geeft aan of de bewuste record in de database is opgeslagen ('retrieved') of dat de record is afgeleid op basis van opgeslagen data ('computed').

```
"organisation","location_code","easting","northing","property","date_start","time_start","date_end","time_end","observation","unit","sampling_method","laboratory_method","embargo","origin"
"ZHEW","RWS 0111",99602,437093,"N_NH4","2004-09-15","12:00:00","2004-09-15","12:00:00",0.05,"mg/l","point sample","unknown",1,"retrieved"
"ZHEW","RWS 0111",99602,437093,"N_NO2","2004-09-15","12:00:00","2004-09-15","12:00:00",0.05,"mg/l","point sample","unknown",1,"retrieved"
"ZHEW","RWS 0111",99602,437093,"N_NO2NO3","2004-09-15","12:00:00","2004-09-15","12:00:00",2,"mg/l","point sample","unknown",1,"retrieved"
"ZHEW","RWS 0111",99602,437093,"O2","2004-09-15","12:00:00","2004-09-15","12:00:00",8.6,"mg/l","point sample","unknown",1,"retrieved"
"ZHEW","RWS 0111",99602,437093,"pH","2004-09-15","12:00:00","2004-09-15","12:00:00",8.1,"-","point sample","unknown",1,"retrieved"
"ZHEW","RWS 0111",99602,437093,"P_ortho","2004-09-15","12:00:00","2004-09-15","12:00:00",0.09,"mg/l","point sample","unknown",1,"retrieved"
"ZHEW","RWS 0111",99602,437093,"P_total","2004-09-15","12:00:00","2004-09-15","12:00:00",0.15,"mg/l","point sample","unknown",1,"retrieved"
"ZHEW","RWS 0111",99602,437093,"SO4","2004-09-15","12:00:00","2004-09-15","12:00:00",63,"mg/l","point sample","unknown",1,"retrieved"
"ZHEW","RWS 0111",99602,437093,"Cl","2004-09-24","12:00:00","2004-09-24","12:00:00",175,"mg/l","point sample","unknown",1,"retrieved"
"ZHEW","RWS 0111",99602,437093,"EC","2004-09-24","12:00:00","2004-09-24","12:00:00",106,"mS/m","point sample","unknown",1,"retrieved"
"ZHEW","RWS 0111",99602,437093,"HCO3","2004-09-24","12:00:00","2004-09-24","12:00:00",174,"mg/l","point sample","unknown",1,"retrieved"
"ZHEW","RWS 0111",99602,437093,"N_kjeldahl","2004-09-24","12:00:00","2004-09-24","12:00:00",0.8,"mg/l","point sample","unknown",1,"retrieved"
"ZHEW","RWS 0111",99602,437093,"N_NH4","2004-09-24","12:00:00","2004-09-24","12:00:00",0.1,"mg/l","point sample","unknown",1,"retrieved"
"ZHEW","RWS 0111",99602,437093,"N_NO2","2004-09-24","12:00:00","2004-09-24","12:00:00",0.05,"mg/l","point sample","unknown",1,"retrieved"
"ZHEW","RWS 0111",99602,437093,"N_NO2NO3","2004-09-24","12:00:00","2004-09-24","12:00:00",2.1,"mg/l","point sample","unknown",1,"retrieved"
"ZHEW","RWS 0111",99602,437093,"O2","2004-09-24","12:00:00","2004-09-24","12:00:00",6.8,"mg/l","point sample","unknown",1,"retrieved"
"ZHEW","RWS 0111",99602,437093,"pH","2004-09-24","12:00:00","2004-09-24","12:00:00",8,"-","point sample","unknown",1,"retrieved"
"ZHEW","RWS 0111",99602,437093,"P_ortho","2004-09-24","12:00:00","2004-09-24","12:00:00",0.11,"mg/l","point sample","unknown",1,"retrieved"
"ZHEW","RWS 0111",99602,437093,"P_total","2004-09-24","12:00:00","2004-09-24","12:00:00",0.16,"mg/l","point sample","unknown",1,"retrieved"
"ZHEW","RWS 0111",99602,437093,"SO4","2004-09-24","12:00:00","2004-09-24","12:00:00",77,"mg/l","point sample","unknown",1,"retrieved"
"ZHEW","RWS 0111",99602,437093,"Cl","2004-09-29","12:00:00","2004-09-29","12:00:00",105,"mg/l","point sample","unknown",1,"retrieved"
"ZHEW","RWS 0111",99602,437093,"EC","2004-09-29","12:00:00","2004-09-29","12:00:00",79,"mS/m","point sample","unknown",1,"retrieved"
"ZHEW","RWS 0111",99602,437093,"HCO3","2004-09-29","12:00:00","2004-09-29","12:00:00",168,"mg/l","point sample","unknown",1,"retrieved"
"ZHEW","RWS 0111",99602,437093,"N_kjeldahl","2004-09-29","12:00:00","2004-09-29","12:00:00",0.7,"mg/l","point sample","unknown",1,"retrieved"
"ZHEW","RWS 0111",99602,437093,"N_NH4","2004-09-29","12:00:00","2004-09-29","12:00:00",0.05,"mg/l","point sample","unknown",1,"retrieved"
"ZHEW","RWS 0111",99602,437093,"N_NO2","2004-09-29","12:00:00","2004-09-29","12:00:00",0.05,"mg/l","point sample","unknown",1,"retrieved"
"ZHEW","RWS 0111",99602,437093,"N_NO2NO3","2004-09-29","12:00:00","2004-09-29","12:00:00",2.3,"mg/l","point sample","unknown",1,"retrieved"
"ZHEW","RWS 0111",99602,437093,"O2","2004-09-29","12:00:00","2004-09-29","12:00:00",7.9,"mg/l","point sample","unknown",1,"retrieved"
"ZHEW","RWS 0111",99602,437093,"pH","2004-09-29","12:00:00","2004-09-29","12:00:00",7.9,"-","point sample","unknown",1,"retrieved"
"ZHEW","RWS 0111",99602,437093,"P_ortho","2004-09-29","12:00:00","2004-09-29","12:00:00",0.12,"mg/l","point sample","unknown",1,"retrieved"
"ZHEW","RWS 0111",99602,437093,"P_total","2004-09-29","12:00:00","2004-09-29","12:00:00",0.18,"mg/l","point sample","unknown",1,"retrieved"
"ZHEW","RWS 0111",99602,437093,"SO4","2004-09-29","12:00:00","2004-09-29","12:00:00",58,"mg/l","point sample","unknown",1,"retrieved"
"HHSK","EO 17",113757,444941,"N_inorganic","2005-01-21","13:30:00","2005-01-21","13:30:00",3.11,"mg/l","point sample","unknown",1,"computed"
"HHSK","EO 17",113757,444941,"N_organic","2005-01-21","13:30:00","2005-01-21","13:30:00",8.49,"mg/l","point sample","unknown",1,"computed"
"HHSK","EO 17",113757,444941,"N_total","2005-01-21","13:30:00","2005-01-21","13:30:00",11.6,"mg/l","point sample","unknown",1,"computed"
"HHSK","EO 17",113757,444941,"P_inorganic","2005-01-21","13:30:00","2005-01-21","13:30:00",0.16,"mg/l","point sample","unknown",1,"computed"
"HHSK","EO 17",113757,444941,"P_organic","2005-01-21","13:30:00","2005-01-21","13:30:00",0.1,"mg/l","point sample","unknown",1,"computed"
"HHSK","EO 17",113757,444941,"N_inorganic","2005-02-17","09:40:00","2005-02-17","09:40:00",1.7,"mg/l","point sample","unknown",1,"computed"
"HHSK","EO 17",113757,444941,"N_organic","2005-02-17","09:40:00","2005-02-17","09:40:00",2.2,"mg/l","point sample","unknown",1,"computed"
"HHSK","EO 17",113757,444941,"N_total","2005-02-17","09:40:00","2005-02-17","09:40:00",3.9,"mg/l","point sample","unknown",1,"computed"
"HHSK","EO 17",113757,444941,"P_inorganic","2005-02-17","09:40:00","2005-02-17","09:40:00",0.005,"mg/l","point sample","unknown",1,"computed"
"HHSK","EO 17",113757,444941,"P_organic","2005-02-17","09:40:00","2005-02-17","09:40:00",0.205,"mg/l","point sample","unknown",1,"computed"
"HHSK","EO 17",113757,444941,"N_inorganic","2005-03-18","10:20:00","2005-03-18","10:20:00",0.61,"mg/l","point sample","unknown",1,"computed"
"HHSK","EO 17",113757,444941,"N_organic","2005-03-18","10:20:00","2005-03-18","10:20:00",3.2,"mg/l","point sample","unknown",1,"computed"
"HHSK","EO 17",113757,444941,"N_total","2005-03-18","10:20:00","2005-03-18","10:20:00",3.81,"mg/l","point sample","unknown",1,"computed"
"HHSK","EO 17",113757,444941,"P_inorganic","2005-03-18","10:20:00","2005-03-18","10:20:00",0.005,"mg/l","point sample","unknown",1,"computed"
"HHSK","EO 17",113757,444941,"P_organic","2005-03-18","10:20:00","2005-03-18","10:20:00",0.255,"mg/l","point sample","unknown",1,"computed"
"HHSK","EO 17",113757,444941,"N_inorganic","2005-04-27","09:40:00","2005-04-27","09:40:00",0.125,"mg/l","point sample","unknown",1,"computed"
"HHSK","EO 17",113757,444941,"N_organic","2005-04-27","09:40:00","2005-04-27","09:40:00",2.3,"mg/l","point sample","unknown",1,"computed"
"HHSK","EO 17",113757,444941,"N_total","2005-04-27","09:40:00","2005-04-27","09:40:00",2.425,"mg/l","point sample","unknown",1,"computed"
"HHSK","EO 17",113757,444941,"P_inorganic","2005-04-27","09:40:00","2005-04-27","09:40:00",0.23,"mg/l","point sample","unknown",1,"computed"
```

Figuur E.3: Voorbeeld van een (eel van een) csv-bestand met uitvoergegevens.

F Versiebeheer

Dit document is gebaseerd op database revisie 3 en `msdb`-package revisie 40 (versie 0.6-4).

G Dankbetuiging

De volgende personen worden hartelijk bedankt voor het kritisch doorlezen en becommentariëren van een eerdere versie van dit rapport: Erwin van Boekel (Alterra), Ariënne Oortwijn (Alterra) en Christian Siderius (Alterra).