

# INFORMATIEMODEL EN DATADICTIONARY RUNDVEEHOUDERIJ

J. Doeksen

*In de afgelopen jaren zijn er informatiemodellen ontwikkeld voor alle takken van de agrarische sector. Van de meeste modellen zijn ondertussen (werk)versies beschikbaar. Toch zijn, na ruim 3 jaar van ontwikkeling, de resultaten anders dan men zich aanvankelijk had voorgesteld. In de loop van de tijd hebben zich onder invloed van praktisch gebruik van de modellen accentverschuivingen voorgedaan. Binnen het informatiemodel voor de rundveehouderij is de nadruk meer en meer op het datamodel komen te liggen. Dit heeft geresulteerd in een (concept) datadictionary voor de rundveehouderij. Daarmee is een produkt ontstaan dat het oorspronkelijke idee van de informatiemodellen wat naar de achtergrond heeft verdronken. In dit artikel zal worden ingegaan op de datadictionary die Taurus ontwikkelt voor de rundveehouderij.*

## Informatiemodel

Bij het opzetten van de informatiemodellen is gebruik gemaakt van Information Engineering (IE), de methode die door James Martin is ontwikkeld. Bij deze methode wordt de ontwikkeling van informatiesystemen opgesplitst in fasen, waarbij de mate van detaillering toeneemt.

Bij IE worden de volgende vijf fasen onderscheiden:

- 1. Strategieplanning;
- 2. Informatie-analyse;
- 3. Systeemontwerp;
- 4. Systeembouw;
- 5. Introductie.

De scheiding tussen de fasen is moeilijk exact aan te geven, fase 1, 2 en 3 lopen meestal geleidelijk in elkaar over.

## Detailstappen

Van fase 1 naar fase 5 neemt de mate van detaillering toe. Dit proces gaat in de regel stapsgewijs. Nadat over de hele breedte van het bedrijfsmodel een detailstap is toegevoegd wordt dit proces herhaald op een vervolgnivo.

In enkele vervolgprojecten werkt Taurus momenteel opnieuw aan een verdere detaillering van het informatiemodel rundveehouderij. In deze fase wordt iedere functie verder gedetailleerd en vindt er vooral een afstemming plaats met informatiesystemen buiten het primaire bedrijf (externe organisaties). Deze detaillering speelt zich vooral af in de overgang van fase 2 naar 3. Het gaat dan vooral om afstemming van de datadefinities in de betreffende systemen. Hierbij wordt informatie van de melkveehouder, al dan niet elektronisch, ontvangen, verwerkt en teruggestuurd. Een belangrijk deel van de data in die systemen is vaak al in het informatiemodel gedefinieerd, een ander deel is bedrijfseigen voor de externe organisatie.

## Oorspronkelijke idee

De bedoeling was dat deze informatiemodellen het uitgangspunt zouden gaan vormen voor het ontwikkelen van informatiesystemen met name voor de primaire bedrijven. Door gebruik te maken van de informatiemodellen zou de informatie gestandaardiseerd worden vastgelegd en verwerkt. Hierdoor zou er op een snelle manier software beschikbaar komen voor primaire bedrijven. Het gebruik van informatietechnologie (IT) op de primaire bedrijven zou er door moeten worden gestimuleerd. Ook moesten witte vlekken in de kennis worden opgespoord die de ontwikkeling van software op het primaire bedrijf belemmerden. Het landbouwkundig onderzoek zou deze hiaten in de kennis met doelgericht onderzoek op moeten vullen.

## Resultaat

Het uiteindelijke resultaat van het project informatiemodellen heeft een iets ander karakter gekregen dan oorspronkelijk was bedoeld. In het globale informatiemodel voor de melkveehouderij werden ongeveer 90 suggesties gedaan voor het opzetten van onderzoek. Door budgetaire problemen heeft het landbouwkundig onderzoek daarvan uiteindelijk slechts enkele overgenomen. Om die reden zijn deze voorstellen in het gedetailleerde modellen voor de rundveehouderij niet meer genoemd.

Ook de sterke stimulerende werking die er van informatiemodellen uit zou gaan op het beschikbaar komen van software voor het primaire bedrijf is (nog) niet gekomen.

Toch heeft het informatiemodel de afgelopen jaren binnen de rundveehouderij een sterk standaardiserende werking gehad. Het procesmodel heeft duidelijk de rol van referentiemodel gekregen. De meeste ontwikkelaars van boerderijsoftware hebben de structuur van het model overgenomen. Bovendien wordt het procesmodel veel gebruikt binnen onderwijs en voorlichting. Verschillende elementen van het datamodel zijn in betrekkelijk korte tijd al in verschillende communicatiebestanden opgenomen.

## Rapportage

Tijdens het detailleren van het globale model zijn er regelmatig werkversies van het model beschikbaar gekomen. Nieuwere versies zijn daarbij steeds met voorgaande functies/rapporten geïntegreerd. De beheersbaarheid van het procesmodel is voor een belangrijk deel opgelost met het gebruik van de Information Engineering Workbench (IEW). Voor de rapportage van het procesmodel bestaan binnen IEW goede mogelijkheden, voor het datamodel zijn die minder goed. De toegankelijkheid van de rapporten nam evenwel af naarmate het project dichter bij de afronding kwam. De nul-versie van het rapport beslaat ongeveer 1500 pagina's.

Vooraf voor de toegankelijkheid en het gebruik van datamodel en rekenregels was dit een handicap. De aan-

vankelijk geplande rapportage waarin het hele datamodel als tekst was weergegeven bleek voor de gebruiker ongechikt te zijn. Het gebruik van het datamodel door een groot aantal bedrijven in de sector was nieuw. De bereidheid om bij communicatiebestanden gebruik te gaan maken van het datamodel was zeker aanwezig. Maar om de doelstelling te bereiken dat er in ruime mate gebruik van het datamodel zou worden gemaakt moest iets anders worden ontwikkeld. Het zwaartepunt van het project Informatiemodel Rundveehouderij is daardoor meer en meer verschoven en is sinds 2 jaar heel sterk op het datamodel komen te liggen.

### Alfnumerieke benaming

Bij publicatie van de eerste rapporten van het datamodel hadden alle attributen een naam die uit alfanumerieke tekens bestond (woorden). Gebruik van deze methode leverde evenwel nogal wat problemen op bij de afstemming met externe organisaties. In een dergelijke notatie kan het vetpercentage van een melkgift bijvoorbeeld op verschillende manieren worden geschreven:

- percentage vet
- vetpercentage
- vetgehalte
- gehalte vet
- vet%
- vet %
- %vet
- % vet

Vooraf bij elektronische communicatie stuitte het gebruik van een alfanumerieke notatie op grote problemen. De onderlinge verschillen in benamingen die er bestaan bij informatieverwerkende organisaties leiden vaak en gemakkelijk tot misverstanden. Bij communicatie tussen twee of meer partijen zou steeds opnieuw moeten worden nagegaan of beiden inderdaad dezelfde definities gebruiken. Wanneer de definitie niet exact gelijk is is de kans op fouten en problemen erg groot. Tijdens de ontwikkeling van het informatiemodel bleek dat de beschikbaarheid van een datamodel voor de hele sector met alfanumerieke namen weinig standaardisatie zou kunnen bewerkstelligen.

### Nummering

Dit probleem was aanleiding om een nieuwe methode te ontwikkelen om attributen en rekenregels van elkaar te onderscheiden. Daartoe zijn alle in het model genoemde elementen uniek genummerd met een getal dat uit zes posities bestaat. Het hierboven genoemde probleem zal dan niet meer voorkomen omdat hiervoor het unieke nummer 200347 (voorbeeld) wordt gebruikt. De namen waarmee dit element binnen verschillende systemen wordt gebruikt is daardoor van ondergeschikt belang geworden. Wanneer twee gebruikers over hetzelfde nummer spreken dan kan over de definitie geen misverstand meer bestaan. Ook rekenregels worden op deze manier genummerd. Door deze nummering ontstaat een bijzonder aantrekkelijke structuur om te komen tot uniformiteit van begrippen, kengetallen en rekenregels. Door het samenvoegen van alle genummerde elementen ontstaat een zgn. datadictionary.

### Doel(groep) datadictionary

Het doel van een datadictionary is het op snelle en eenvoudige wijze beschikbaar stellen van de beschrijvingen en definities van data-elementen. Het kan daarbij gaan om gegevens die al in het informatiemodel zijn opgenomen, maar ook om gegevens die daarin niet voorkomen. Dit is met name van belang voor degenen die informatie verwerken (informatie-analisten, systeemontwerpers, etc). Uiteraard kunnen ook andere doelgroepen gebruik maken van de datadictionary.

Bovendien is de definitiestudie die vooraf gaat aan de ontwikkeling van een nieuw systeem uiterst kostbaar (en voor sommigen vaak ook vervelend). De beschikbaarheid van een conceptueel datamodel met DD-nummers zal binnen de sector kostenverlagend werken.

### Inhoud datadictionary

In de datadictionary voor de rundveehouderij (VeeDATA) zitten in totaal ongeveer 1800 attributen en rekenregels die willekeurig zijn genummerd. In de door Taurus gebruikte DD zitten voorlopig alleen attributen en rekenregels. In een later stadium kunnen relaties, entiteittypen, subject areas, etc. worden toegevoegd.

De DD werd aanvankelijk gevuld met de data-elementen die in het informatiemodel waren gedefinieerd. Er zijn evenwel ook elementen die niet thuishoren in de informatie-analyse (-modellen). Het betreft hierbij bijvoorbeeld specifieke elementen (gegevens) die worden gebruikt bij de communicatie. Daardoor bevat de DD meer elementen en is de inhoud van de DD groter geworden dan het datamodel van het informatiemodel.

### DBMS

Tijdens de ontwikkeling van de DD bleek dat IEW onvoldoende mogelijkheden bood om de informatie onder te brengen. Ieder DD-element heeft een aantal standaard eigenschappen: codering, de plaats waar het element wordt gebruikt, de datum waarop het is gecreëerd, de datum van update, etc. Binnen IEW zou dit moeten worden ondergebracht in een algemeen tekstveld. Een dergelijk veld biedt onvoldoende mogelijkheden voor selectie. Ook voor het vastleggen van rekenregels zijn de tekstverwerkingsfaciliteiten binnen IEW gebrekkig. Toen steeds meer elementen aan de DD moesten worden toegevoegd die niet in het informatiemodel voorkwamen is besloten tot gebruik van een DBMS, los van het informatiemodel. Bij het gebruik van twee bestanden moet er een hiërarchie worden aangebracht. De relatie informatiemodel-DD wordt daarmee fundamenteel anders dan tot op dat moment werd gehanteerd. Het informatiemodel heeft er voornamelijk voor gezorgd dat de DD werd gevuld. Nadat de DD was gecreëerd heeft Taurus besloten dat het DD bestand leidend is. Zodra een nieuw element in de DD wordt opgenomen dan worden alle relevante velden gevuld, daarna is het element beschikbaar voor gebruik elders (o.a. in IEW).

### Externe informatie

Naarmate de datacommunicatie toeneemt blijkt steeds meer dat de individuele organisaties een onderdeel zijn van een veel groter netwerk. Zij ontvangen informatie vanuit

allerlei verschillende bronnen maar verzenden ook weer informatie. Eén organisatie kan het zich niet meer permitteren om een definitie van een attribuut op een hem passend tijdstip te wijzigen. Omdat het zelfde attribuut ook bij andere organisaties in gebruik kan zijn zullen er strakke afspraken moeten worden gemaakt over een eventuele wijziging.

Het is heel goed denkbaar dat de DD in de toekomst ook de informatie gaat bevatten van externe organisaties/bedrijven. Grote voordeel daarvan is dat de betreffende organisaties zich geen zorgen meer hoeven te maken over het onderhoud van de definities. Wijzigingen worden centraal gemeld en gepubliceerd. Naarmate er meer organisaties zijn die gebruik gaan maken van de mogelijkheden die de DD biedt, wordt het draagvlak groter. Iedereen kan een abonnement nemen op de DD en is daarmee verzekerd van gebruik van gestandaardiseerde data.

Momenteel vinden de eerste verkennende gesprekken plaats om ook datadefinities van externe (boeren)organisaties van een nummer te voorzien en in de DD op te nemen. Het gaat dan om datadefinities die niet in het informatiemodel voor de rundveehouderij voorkomen omdat ze uitsluitend worden gebruikt binnen de externe-/dienstverlenende organisaties. Maar omdat deze elementen mogelijk wel worden uitgewisseld met andere externe organisaties is het aantrekkelijk om ook die gegevens van een uniek nummer te voorzien. Door verwijzing naar de DD kunnen elektronische bestanden met DD-nummers snel en efficiënt worden samengesteld.

### **DD voor land-en tuinbouw**

De problematiek rond het datamodel is natuurlijk niet uniek voor het informatiemodel rundveehouderij. Bij andere takorganisaties hebben zich soortgelijke problemen voorgedaan bij gebruik van de elementen uit de datamodellen. Enkele takorganisaties zijn er ondertussen ook toe overgegaan om hun attributen van een nummer te voorzien. Omdat beheer en onderhoud van een DD niet takgebonden is ontstaat dan meteen de vraag of de tak-DD's niet moeten worden geïntegreerd tot één DD voor land- en tuinbouw. Het opzetten van een centrale DD voor land- en tuinbouw is evenwel een complex probleem, zowel inhoudelijk als organisatorisch. Omdat deze stap te groot is om in één keer te overbruggen hebben de takorganisaties besloten om de DD eerst decentraal op te starten met slechts attributen en rekenregels. Gedurende deze fase worden gemeenschappelijke zaken al vast op elkaar afgestemd. Tegelijkertijd wordt bestudeerd hoe de centrale DD er in de toekomst uit moet gaan zien. Hierbij spelen zaken als inhoud, financiering, beheer, toegankelijkheid, etc. □

---

*Ir. J. Doeksen is werkzaam bij de Takorganisatie Automatisering Uniformering Rundveehouderij Sector (TAU-RUS), postbus 1076, 8200 BB Lelystad.*