



Een systeem is te vervangen maar de gegevens blijven

# Beter sturen door een gedegen dataconversie

Ortwin Verreck, Kees van den Berg en Arjen de Graaf

**Om technische redenen kan een overgang naar een nieuw systeem wenselijk zijn. Als men daarbij gegevenskwaliteit onderschat en de conversie niet goed uitvoert dan kan dat grote invloed hebben op de beschikbare functionaliteit en de acceptatie van het nieuwe systeem. Dit artikel beantwoordt een aantal kernvragen rond dataconversie en presenteert een aanpak waarmee planmatig gegevensconversie kan worden aangepakt.**

Bedrijfsmatige redenen voeren vandaag de dag steeds vaker de boventoon. Denk daarbij aan verdergaande internationalisatie, standaardisatie – naar oplossingen van bijvoorbeeld PeopleSoft of SAP – of schaalvergroting, door fusies en nieuwe samenwerkingsverbanden. In alle gevallen is er behoefte aan conversie van de gegevens, waarbij wel de 'winkel open moet blijven'. Er is dan vaak veel tijd besteed aan de selectie van het nieuwe systeem op basis van de nieuwe bedrijfsprocessen en daarmee samenhangende gewenste functionaliteit.

Een gevaar schuilt in onvoldoende aandacht voor de kwaliteit van de gegevens en de conversieuitdagingen die dit met zich meebrengt. De kwaliteit van de gegevens en de conversie is niet alleen sterk van invloed op de beschikbare functionaliteit in het nieuwe systeem, maar ook het gebruik en de acceptatie van het nieuwe systeem wordt sterk beïnvloed door de kwaliteit van de conversie. Het is dan ook aan te raden om een aantal kernvragen te beantwoorden alvorens de conversieaanpak te bepalen.

## Kernvraagstukken bij gegevensconversie

Op basis van ervaring in diverse conversietrajecten bij grote gegevensintensieve bedrijven is een vijftal kernvragen gedefinieerd waarop antwoord gevonden dient te worden als men gedegen conversie wil uitvoeren: wat is de kwaliteit van de huidige gegevens; hoe kan de kwaliteit worden verbeterd; hoe kom je tot een goede transformatie; welke eisen stel je aan het conversieproces; hoe gaat de organisatie migreren?

### 1. Wat is de kwaliteit van de huidige gegevens.

In de praktijk blijkt vaak dat er geen (gedegen) gegevenskwaliteitsmeting plaats heeft gevonden ter voorbereiding op de gegevensconversie. Terwijl daarmee de grootste voorspellende waarde wat betreft uitdagingen, planning, etcetera gedaan kan worden.

Zo kan het gebeuren dat men in een traject aan de slag gaat met het converteren van interne telefoonnummers terwijl deze slechts bij een enkele medewerker ingevuld zijn. De vraag kan gesteld worden wat we met gegevenskwaliteit bedoelen. Er zijn vele definities van het begrip 'kwaliteit' in omloop. We houden het graag simpel en hanteerbaar, voor ons betekent kwaliteit vooral 'duurzaam bruikbaar'. Kwaliteit is daarmee nauw verbonden met het beoogde gebruik en dus ook het gebruik na conversie. Een uitstekende nijptang is vaak een belabberde hamer. Hetzelfde geldt voor de kwaliteit van gegevensverzamelingen. Of een verzameling van voldoende kwaliteit is hangt af van wat we ermee willen gaan doen in het nieuwe systeem. Duurzame bruikbaarheid vertalen we in een tweetal kwaliteitsdimensies die aan het beoogde gebruik van de gegevens zijn gerelateerd. Die dimensies zijn betrouwbaarheid en relevantie. Als de kwaliteit van de gegevens niet voldoet aan de eisen van betrouwbaarheid en relevantie van het nieuwe systeem, dan kan het gebeuren dat die gegevens niet kunnen worden overgenomen in het nieuwe systeem.

### 2. Hoe kan de kwaliteit worden verbeterd.

Als de kwaliteit van de gegevens niet voldoet aan de nieuwe betrouwbaarheids- en relevantie-eisen van het nieuwe systeem dan is het opschonen en verrijken van de gegevens feitelijk de enige juiste weg die kan worden bewandeld. Het schonen bij de bron verdient daarbij de voorkeur, zodat vanuit de juiste context schoningsmaatregelen genomen kunnen worden en er geen onnodige uitval ontstaat tijdens de conversie. Er zijn drie vormen van vervuiling te onderkennen.

*Technische vervuiling* is het gemakkelijkste te detecteren. Het gaat daarbij altijd om vervuiling in de administratie waarvan eenvoudig is vast te stellen dat dit niet voor 'kan' komen, bijvoorbeeld een telefoonnummer met letters.

*Functionele vervuiling* heeft altijd betrekking op registraties die

kunnen voorkomen, maar niet 'mogen' voorkomen.

Praktijkvoorbeelden: medewerkers die een full-time jaarsalaris hebben onder het wettelijk minimumloon.

*Inhoudelijke vervuiling* is veruit het lastigst vast te stellen. Het gaat hier dan ook om registraties die kunnen voorkomen en ook mogen voorkomen, maar die feitelijk 'onjuist' zijn.

Uiteraard zijn alle drie vormen van vervuiling relevant. De eerste twee vormen dienen om conversieredenen te worden opgelost. De derde variant is vanuit bedrijfsoptiek belangrijk en kan het best in het oorspronkelijke systeem – de juiste context – worden opgelost. Met behulp van (analyse)tools kunnen verslagen worden gemaakt waarmee zowel handmatige als geautomatiseerde schoningsacties kunnen worden uitgevoerd. Dit hoeft echter niet in alle gevallen voldoende te zijn. Het kan nodig zijn om ook andere gegevensbronnen te benaderen om de gegevens verder te verrijken. De uiteindelijke verrijking vindt plaats tijdens het conversieproces.

### 3. Hoe kom je tot een goede transformatie.

Bij dit vraagstuk is het de opgave om een mappingstructuur te vinden die de gegevens, opgeslagen volgens de logische en technische structuur van het bronsysteem, vertaalt naar de logische en technische structuur van het doelsysteem. Bij het maken van een goede mapping tussen twee systemen is de betekenis in het bronsysteem en het gebruik van de gegevens in het doelsysteem doorslaggevend. Dit is in de regel een zodanig complexe uitdaging dat het bij voorkeur op een iteratieve manier opgelost dient te worden. Er kunnen veel bedrijfsregels zijn die ingaan op – soms erg veel – specifieke situaties waarbij afwijkend moet worden geconverteerd. Er kunnen daarbij niet alleen veel afwijkende situaties zijn maar vaak zijn deze situaties ook niet gedocumenteerd. Een iteratieve aanpak is dan aan te bevelen waarbij begonnen wordt met een zeer strak filter, zodat alleen de gedefinieerde goedsituaties worden goedgekeurd tijdens conversie en de rest zal worden gekenmerkt als conversie-uitval. Door de in elke stap opduikende probleemgevallen zorgvuldig te registreren en te definiëren als additionele bedrijfsregels, kan de uitval worden gereduceerd en tevens wordt een waardevolle testset opgebouwd waarmee de te ontwikkelen conversieprogrammatuur kan worden getest.

### 4. Welke eisen stel je aan het conversieproces.

In principe onderkent ieder conversieprogramma drie fasen, te weten: extraheren uit het bronsysteem, transformeren volgens de conversie- en bedrijfsregels en laden via de applicatielaag van het doelsysteem. In de applicatielaag van het doelsysteem worden de gegevens uiteindelijk definitief gevalideerd. Het is de bedoeling dat op die plek feitelijk geen uitval meer plaats vindt. Er is een aantal eisen te stellen aan de conversieprogrammatuur:

1. Tijdens de extractie uit het bronsysteem stellen wij als eis dat er geen bewerking van gegevens plaatsvindt. Ook blijft het bronsysteem zelf bij een extractie ongewijzigd;
2. Voor het laden van de gegevens in het doelsysteem geldt hetzelfde;

3. Alleen in de transformatiefase worden de gegevens bewerkt en omgezet naar de nieuwe situatie.

Voor alle stappen in het proces dienen controle-totaal en controle-getallen gedefinieerd te worden, zodat gegarandeerd wordt dat er een volledige conversie heeft plaatsgevonden en er als gevolg van de conversie geen financiële verschillen zijn ontstaan.

## Er zijn vele definities van het begrip 'kwaliteit' in omloop

Een belangrijke keuze bij het ontwikkelen van conversieprogrammatuur is: zelf ontwikkelen of kopen? De keuze voor het zelf ontwikkelen van conversieprogrammatuur wordt meestal ingegeven door de veronderstelde (lage) complexiteit. Bij de op de markt verkrijgbare tools is onderscheid te maken tussen de zogenaamde codegeneratietools en de ETL-tooling. Zeker door de kracht van de huidige tools ontstaat een betrouwbare en snelle oplossing.

### 5. Hoe gaat de organisatie migreren.

Met het uitvoeren van de daadwerkelijke conversie en het overzetten van de geconverteerde gegevens naar productie, stapt de organisatie meestal ook over op een nieuwe werkwijze. Dit blijkt in de praktijk een onderschatte operatie te zijn. Enkele aandachtspunten hierbij:

- Zorg voor een duidelijke knip tussen de oude en de nieuwe situatie. Er mogen geen gegevens meer in de pijplijn zitten;
- Bekijk wat de consequenties zijn voor partners die voor of achter op de eigen keten aansluiten. Ook zij gebruiken de geconverteerde gegevens;
- Werk, bij een gefaseerde invoering, met duidelijk afgebakende releases;
- Spaar de geconstateerde fouten waar mogelijk op, om ze bij de volgende release te herstellen;
- Een stap binnen het conversiedraaiboek mag pas worden uitgevoerd na controle en accordering van de voorgaande stap;
- Indien er gedurende de conversie gewoon wordt doorgewerkt dienen de gegevens die dan ontstaan te worden opgespaard, om later in het doelsysteem te kunnen worden geplaatst.

## Conversieaanpak

Om een gedegen gegevensconversie te kunnen uitvoeren is het van groot belang zorgvuldig te werk te gaan, niet alleen tijdens de daadwerkelijke conversie zelf maar ook tijdens de voorbereiding en nazorg. Hoewel in theorie een gegevensconversie natuurlijk herhaalbaar is, gaat men er in de praktijk meestal vanuit dat het een éénmalige actie betreft. Wat overigens niet betekent dat een gegevensconversie altijd daadwerkelijk op één moment plaatsvindt.

De werkwijze voor het uitvoeren van gegevensconversies kan wor-

den onderverdeeld in een drietal stappen; voorbereiding, uitvoering en nazorg.

## Stap 1: Voorbereiding.

Bij de voorbereiding hoort een gegevenskwaliteitsmeting. Op basis van deze meting kunnen op voorhand potentiële conversieuitdagingen worden geïdentificeerd. De conversieuitdagingen behoeven niet altijd per definitie tijdens het conversietraject opgelost te worden. Er kan ook voor gekozen worden dat deze door de lijnorganisatie worden opgelost en dat vervolgens een periodieke meting plaatsvindt op de nog vigerende uitdagingen.

Tijdens de gegevenskwaliteitsmeting wordt van de te converteren gegevens de veronderstelde kwaliteit in kaart gebracht. Dit kan door diverse soorten controles, waarvan de belangrijkste zijn:

- Domeincontroles: hierbij wordt gekeken of de waarden die in een rubriek voorkomen geldige waarden zijn;
- Integriteitscontroles: hierbij wordt getoetst of aan allerlei validiteitsregels is voldaan. Zo weten we dat het gewicht van een auto niet elke waarde kan aannemen, dat een order altijd bij een klant hoort, een trouwdatum ruim na de geboortedatum valt enzovoort. In deze controle wordt ook vaak de technische samenhang tussen een aantal velden bekeken. Zijn de sleutelwaarden uniek? Bestaat de klant die in de order wordt genoemd? Heeft elke levering geleid tot een factuur?;
- Bedrijfsregelcontroles: hierbij wordt gecontroleerd op overtredingen van bedrijfsregels. Voorbeelden zijn: niet toegestane combinaties van arbeidsvoorwaarden, niet gefiatteerde leveringen van boven de € 1000,-.

## Zeker in het begin moeten gebruikers wennen aan de nieuwe situatie

Na de kwaliteitsmeting zal, ter voorbereiding op de daadwerkelijke gegevensconversie (Uitvoering), een aantal andere onderdelen in kaart worden gebracht. De nadruk ligt hierbij niet alleen in het technisch omzetten, maar meer in het definiëren van de conversieregels en de functionele uitdagingen die daarbij boven tafel komen.

De volgende activiteiten worden na de gegevenskwaliteitsmeting uitgevoerd:

- Analyse en vastlegging van de structuur van de bronbestanden en de doelbestanden;
- Opstellen van de conversieregels ten aanzien van selectie, mapping en transformatie;
- Definiëren van de controle-eisen. Hoe wordt de volledigheid en de juistheid van de conversie vastgesteld. Wat zijn bijvoorbeeld de eisen van accountants, toezichhouders en actuarissen.

Controletellingen, aangevuld met steekproeven, zijn hierbij het meest gangbaar;

- Bepalen van de schonings- en verrijkingsmaatregelen.

## Stap 2: Uitvoering.

De uitvoering is als stap onder te verdelen in een aantal afzonderlijke activiteiten, te beginnen met de realisatie van alle voor de conversie benodigde componenten. Daarna komt de uitvoering van de daadwerkelijke gegevensconversie, gevolgd door een controle van de resultaten.

De te realiseren componenten bestaan uit maatwerkconversie-programmatuur en/of uit het inrichten van conversietools voor een meer geautomatiseerde gegevensconversie (Op basis van de in de tool vastgelegde mappingregels en afspraken met betrekking tot de conversie worden dan de conversieprogramma's gegenereerd).

De volgende activiteiten worden uitgevoerd:

- Er worden procedures ontwikkeld voor het uitvoeren van de gegevensconversie en de daaraan gerelateerde activiteiten als ontladen, selecteren en extraheren van gegevens;
- Er zal programmatuur worden ontwikkeld voor het uitvoeren van controles die voor, tijdens en na de gegevensconversie plaatsvinden. Hierdoor wordt duidelijk welke controletotalen en controle-aantallen gebruikt worden. Zo moet bijvoorbeeld in beide systemen direct na de conversie hetzelfde totaalbedrag aan openstaande facturen vermeld staan. Ook kan het zijn dat er procedures en instructies gemaakt moeten worden om (een deel van) de conversie handmatig uit te voeren;
- Om alle stappen van de conversie gestructureerd, herhaalbaar en controleerbaar uit te kunnen voeren wordt een conversie-draaiboek opgesteld. Daarbij wordt ook rekening gehouden met fallback-scenario's.

Nadat alle benodigde componenten akkoord zijn bevonden kan de daadwerkelijke conversie volgens het opgestelde draaiboek worden uitgevoerd. Daarbij is het ook zaak dat organisatorische maatregelen zijn genomen waarbij zoveel mogelijk de 'pijplijn is leeggemanaged', zodat er zo min mogelijk onderhanden werk is.

## Stap 3: Nazorg.

Na uitvoeringsfase dient een aantal afrondende activiteiten te worden uitgevoerd, door middel van een kwaliteitsmeting op het doelsysteem en een afsluitende controle op de uitgevoerde conversie. Daarbij wordt een aansluitcontrole gedaan tussen de controletotalen en controle-aantallen. Na deze controle kan het nieuwe systeem worden vrijgegeven.

Zeker in het begin moeten gebruikers wennen aan de nieuwe situatie en kunnen ze gemakkelijk invoerfouten maken, en dus nieuwe vervuiling creëren. Het plannen van een aantal gegevenskwaliteitsmetingen kan tijdig zicht geven op dit probleem. Tenslotte worden de maatregelen bepaald waarmee vervuiling in de toekomst voorkomen kan worden. Ook hier spelen de verkregen bedrijfsregels weer een belangrijke rol. Met behulp van deze regels kunnen de invoerprocessen zo ingericht worden dat er

zoveel mogelijk schoning plaatsvindt bij de bron zelf, maar nu wel in het nieuwe systeem. De kwaliteitsmeting is gebaseerd op de criteria die ook gebruikt zijn bij de initiële meting tijdens de voorbereiding. Het is verstandig periodiek kwaliteitsmetingen te blijven doen om zo vroegtijdig mogelijk onjuiste interpretaties en verkeerde invoer op te sporen.

## Conclusie

Voor al de drie gepresenteerde stappen geldt dat sturing door het business management essentieel is. Per slot van rekening is het management de klant. Daarnaast leert de ervaring dat de politieke factor hierdoor niet onnodig sterk wordt. Een 'betrokken sturing' kan, als het business management weet 'wat-waarom-wanneer' dient plaats te vinden, inclusief bijbehorende output. Hiervoor dient gedurende het conversietraject tijd voor ingepland te worden.

### Ortwin Verreck, Kees van den Berg en Arjen de Graaf

Ortwin Verreck (ortwin.verreck@arvix.nl) en Arjen de Graaf (arjen.de.graaf@arvix.nl) zijn respectievelijk technisch en algemeen directeur van Arvix. Kees van den Berg is management consultant bij Arvix.

## In Memoriam Lex de Haan (1954-2006)

Op 1 februari 2006 overleed Lex de Haan. Lex vervulde een belangrijke rol binnen de Oracle-gemeenschap in Nederland en daarbuiten. Hij presenteerde op talloze conferenties en gaf les over de hele wereld. Bovendien schreef hij twee boeken over SQL. Het 'Leerboek Oracle SQL' verscheen bij Academic Service en beleefde een paar herdrukken. Vorig jaar verscheen zijn nieuwe boek 'Mastering Oracle SQL en SQL\*Plus', uitgebracht onder het 'Oaktable Press' label bij Apress.

Een goede vriend van Lex was Chris Date, met wie hij sterk van mening verschilde over het gebruik van de NULL-value in SQL. Lex was, in tegenstelling tot Chris Date, een overtuigend voorstander van het gebruik van de NULL-value. Veel van zijn lezingen op conferenties behandelden dit onderwerp, wat hem de bijnaam 'Mr. Null' opleverde.

Medio 2005 besloot Lex om een nieuw boek op stapel te zetten: 'Applied Mathematics for Database Professionals'. Dit boek heeft hij helaas niet kunnen afmaken; co-auteur Toon Koppelaars heeft inmiddels gemeld dat hij het boek zal afronden.

Lex de Haan was veelzijdig, briljant en bescheiden. Hij wordt node gemist.



## Bent u ICT-professional of ICT-Decision maker?

Computable is de meest gelezen informatiebron voor ICT-professionals en ICT-managers in Nederland. Met altijd het belangrijkste nieuws, objectieve scherpe analyses, vacatures en onafhankelijke productinformatie. Bent u als ICT'er werkzaam, dan heeft u recht op een kosteloos abonnement.

**Ga nu naar [www.abonneren.nl/computable](http://www.abonneren.nl/computable) en meld u aan!**

Computable, dé ICT-informatiebron

**Computable**