

URBIDATA

ir H.J.M. Corstens,
dr ir R.V. Schuwer:

GEGEVENS MANAGEMENT

Overdruk van een artikel in het tijdschrift
NGT Geodesia (1993 nr 5)

Informatica-adviesbureau URBIDATA
Hasselaarstraat 16 5616 SW Eindhoven
Postbus 8595 5605 KN Eindhoven
Telefoon 040-573130 Fax 040-573935

Om een beter beeld van het management van gegevens te krijgen, zal in het volgende onderdeel dit proces nader worden geanalyseerd.

Gegevensmanagement als een cyclisch proces

Management is gericht op het permanent minimaliseren van het verschil tussen gewenste en werkelijke situaties. Het betreft concreet het permanent minimaliseren van het verschil tussen gewenste en werkelijke toestanden van de gegevenshuishouding in een organisatie (of een multidisciplinair netwerk).

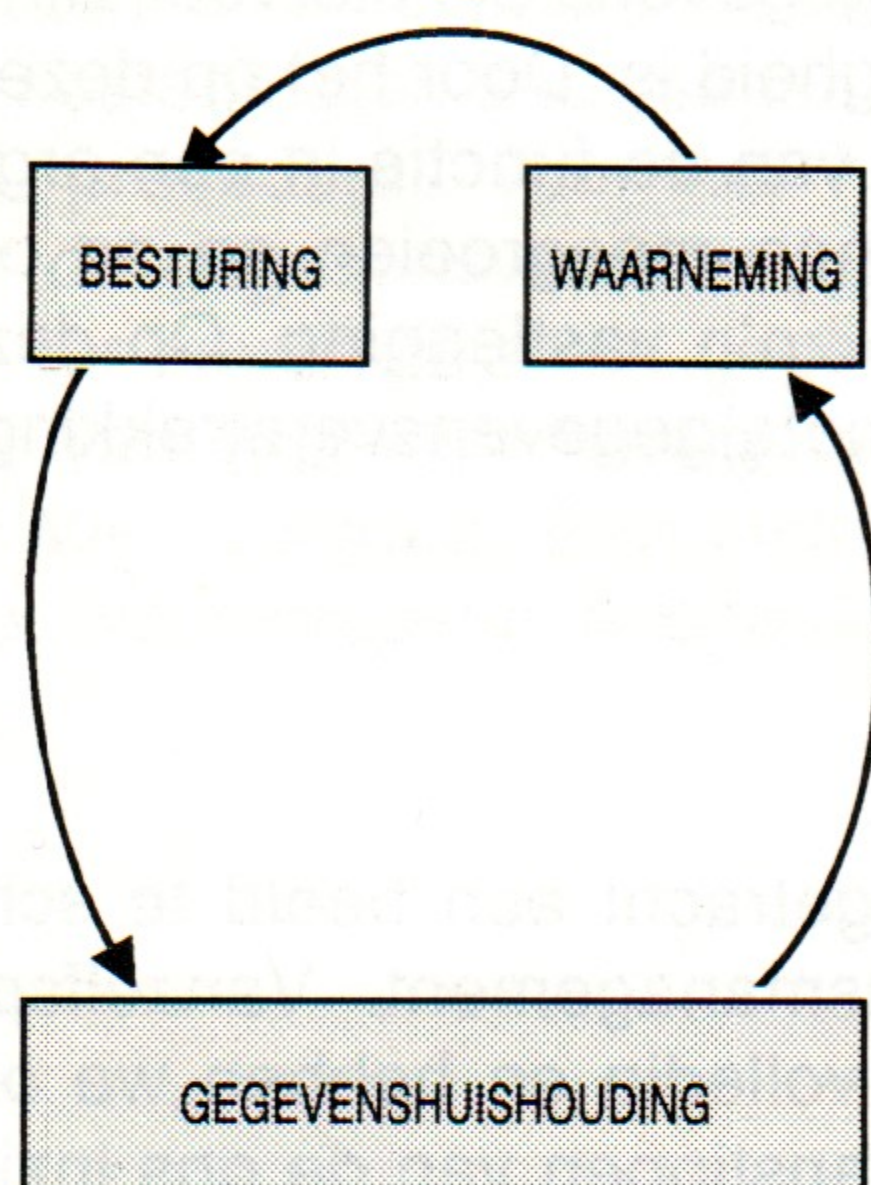


Fig. 1.
Algemeen overzicht van gegevensbeheer.

Gegevensmanagement kan dan ook worden gedefinieerd als het vaststellen van de gewenste gegevens(verzamelingen) in een organisatie, het verkrijgen van een beeld van de werkelijke gegevens(verzamelingen) en het bijsturen op basis van de geconstateerde verschillen tussen gewenste en werkelijke situatie.

Gegevensmanagement is te beschouwen als een besturingsproces, een cyclisch proces, gebaseerd op het permanent afwisselen van beleidsbepalen-waarnemen-bijsturen (fig. 1).

In fig. 2 wordt het cyclische proces verder uitgewerkt. De volgende onderdelen zijn te onderscheiden:

- *probleemsignaleren*: inventarisatie en selectie van de problemen in de gegevenshuishouding, zodat de aan te pakken problematiek wordt ingekaderd;
- *doelbepaling*: vaststellen van wat men wil bereiken, gerelateerd aan de doelstellingen van de organisatie;
- *normeren*: preciseren van de doelstellingen in termen van normen en kwaliteitseisen;
- *waarnemen*: inventarisatie en inspectie van aanwezige gegevensverzamelingen en de opslag daarvan in een (meta)gegevensbank. Resultaat is een beeld van de realiteit, in termen van de aanwezige gegevenshuishouding. In feite is waarnemen in het verband van gegevensmanagement een permanent „monitoring” proces;
- *evaluatie*: vergelijken van normen en/of eisen met de realiteit. Resultaat is een getoetst beeld van de aanwezige gegevenshuishouding, bijvoorbeeld in termen van knelpunten of afwijkingen;
- *planvorming*: ontwikkelen van oplossingen voor de gevonden problemen of wegen om de gestelde doelen te bereiken, de keuze van één oplossing of weg en een daarop gebaseerde uitwerking van een plan voor de gewenste gegevenshuishouding;

- *actie*: invoeren en realisatie van het plan. Hiertoe behoren:
 - bewaken van de performance van systemen;
 - zorgdragen voor adequate herstelprocedures en het coördineren van herstelwerkzaamheden bij het verloren gaan van gegevens;
 - bewaken van toegekende autorisaties en privacy;
 - beheren van het gegevensbankbeheersysteem (DBMS);
 - ondersteunen van het fysiek gegevensbankontwerp;
 - ontwikkelen en beheren van gegevensmodellen en het verder uitwerken en beheren van metagegevens;
 - verlenen van de benodigde autorisaties voor het gebruik van de beheerde gegevens;
 - toezien op de kwaliteit van gegevens(verzamelingen);
 - toezien op eenduidige definiëring van begrippen en relaties;
 - beschikbaar stellen van kopieën van bestanden („copy management”);
 - verstrekken van richtlijnen en instructies voor het gebruik van gegevensverzamelingen;
 - geven van voorlichting over de interpretatie van gegevens;
- *control*: permanent meten en bijsturen op basis van knelpunten of afwijkingen. Terugkoppeling kan leiden tot bijstellen van doelen, plannen en/of normen.

Implementatie van gegevensmanagement betekent onder meer het definiëren van een aantal organisatorische functies. Welke functies dit zijn en wat globaal de inhoud ervan is, wordt hierna uitgewerkt.

Activiteiten

Gegevensmanagement omvat een uitgebreid scala aan activiteiten en werkzaamheden. In navolging van [4] en [1] maken we onderscheid tussen:

- gegevensbankbeheer (Data Base Administration, DBA);
- gegevensbeheer (Data Administration, DA);

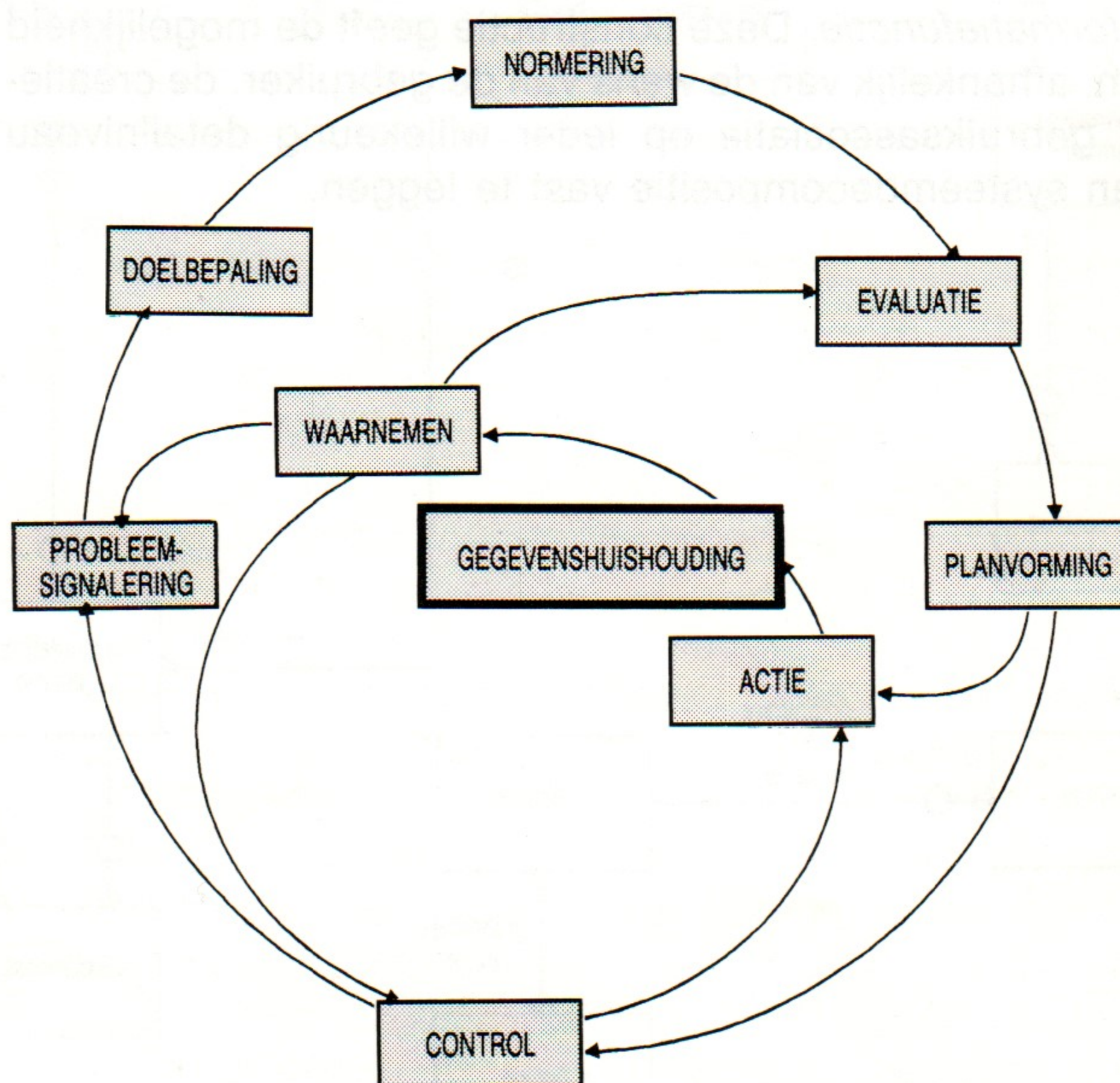


Fig. 2. Nadere uitwerking van gegevensbeheer.

- informatiebeheer (Information Resource Management, IRM).

Hierbij kan DBA als een aspect van DA, en DA als een aspect van IRM worden gezien. DBA, DA en IRM kunnen als opeenvolgende mijlpalen worden gezien op de weg naar compleet gegevensmanagement.

Hierna worden de genoemde activiteiten beschreven, die van belang zijn voor gegevensmanagement (ook wel aangeduid als „gegevensbeheer” [5]). Vooraf een opmerking over de organisatorische inrichting van gegevensmanagement. Deze dient aan te sluiten bij de algemene organisatorische inrichting in het bedrijf of de overheidsorganisatie, waar gegevensmanagement wordt ingevoerd. Een absolute ideale organisatievorm bestaat niet.

Afhankelijk van diverse factoren kan worden gekozen voor een verschillende omvang en gedifferentieerdheid van de organisatie, mate van centralisatie en mate van concentratie van activiteiten. Factoren zijn: informatiseringsgraad, aard van het bedrijf, bedrijfscultuur en omvang van het bedrijf.

Gegevensbankbeheer (Data Base Administration, DBA)

De DBA-taak is technisch georiënteerd. Wanneer de geautomatiseerde informatiesystemen gebruik gaan maken van speciale gegevensbankmanagementsystemen, zal de logische indeling van de gegevens moeten worden „vertaald” naar technische opslagstructuren. De gegevens van het informatiesysteem moeten zodanig op de opslagmedia worden vastgelegd, dat in een operationele situatie kan worden voldaan aan eisen zoals (zoek)snelheid, integriteit en flexibiliteit. De DBA-taak dient te waarborgen dat elke binnen de organisatie te bouwen gegevensbank (-beheersysteem) aan deze eisen voldoet, door middel van ontwerp, toetsing van ontwerpen of ondersteuning van systeemontwerpers. Samengevat omvat DBA de volgende deeltaken:

- bewaken van de performance;
- zorgen dat aan de voorschriften voor beveiliging en verlies wordt voldaan;
- bewaken van privacy en gebruik van de toegekende autorisatie;
- zorg voor programmatuur (DBMS) en contacten met de leveranciers;
- ondersteunen van fysiek gegevensbankontwerp.

Gegevensbeheer (Data Administration, DA)

De DA-taak ofwel gegevensadministratie is gericht op het effectief en efficiënt beheren van de gebruiksaspecten van gegevens. Dit gebeurt door het opbouwen en beheren van een metagegevensverzameling. Tot DA behoren de volgende deeltaken:

- ontwikkelen en beheren van gegevensmodellen;
- fungeren als intermediair tussen managers, gebruikers en automatiseerders;
- bevorderen van de beschikbaarheid van gegevens, bijvoorbeeld door het vervullen van de rol van informatiecentrum;
- zorg voor richtlijnen, standaards en procedures om de privacy te waarborgen, om voorzieningen te treffen in geval van beschadiging en verlies van gegevens, om de kwaliteit van de gegevens te garanderen en om de efficiency van verzamelen, invoeren, verwerken en uitvoeren van gegevens te bevorderen;

- beheer van het gegevenswoordenboek (data dictionary). De vastlegging is niet meer uitsluitend gericht op het DBMS en het technisch ontwerpproces, doch ook op systeemontwikkeling en gebruik. Vastgelegd worden zinvolle definities en entiteitstypen, attributen, relatietypen, toegestane domeinwaarden en overige beperkingsregels (constraints).

Informatiebeheer (Information Resource Management, IRM)

IRM betreft het opstellen en bewaken van een gegevensbeleid en een gegevensplan, en het zorg dragen voor de gewenste kwaliteit van gegevens. In het kader van IRM worden gegevens vastgelegd over inhoud, kwaliteit en gebruik van gegevens. Samengevat vallen de volgende deeltaken onder IRM:

- ontwikkelen van gegevensbeleid;
- maken van gegevensplannen als operationalisering van het beleid, en het (doen) uitvoeren van deze plannen;
- leiden en begeleiden van gegevensbeheer en gegevensadministratie.

Implementatie van de functies voor gegevensmanagement moet als uiteindelijk resultaat een kwalitatief hoogstaande gegevenshuishouding opleveren.

Gegevensmanagement en kwaliteit

Er is een sterke overeenkomst tussen kwaliteits- en gegevensmanagement. Kwaliteitsmanagement kan worden gedefinieerd als: *het vaststellen van de gewenste kwaliteit in een organisatie, het meten van de werkelijke kwaliteit en het bijsturen op basis van geconstateerde verschillen tussen gewenste en werkelijke kwaliteit.*

Centraal in deze definitie staat het begrip kwaliteit. Volgens het Nederlandse Normalisatie Instituut [2] moet hieronder worden verstaan: *de mate waarin een produkt voldoet aan de in specificaties vastgelegde eisen en wensen die voortvloeien uit het gebruiksdoel.* Toegepast op gegevensmanagement gaat het om kwaliteitseisen aan de gegevens. We onderscheiden de volgende kwaliteitseisen (deels volgens [3]):

- *integriteit*: de mate van overeenstemming met de werkelijkheid, en valt daarbij uiteen in de sub-eisen:
 - semantische correctheid: overeenstemming van de betekenis van een term in een gegevensbank met de voorstelling die een gebruiker van de term heeft;
 - syntactische correctheid: juistheid van de vorm van een term;
 - consistentie: afwezigheid van tegenstrijdige beweringen, afleidbaar uit een gegevensverzameling;
 - geldigheid: juistheid van een term voor een aan gegeven begrenzing;
 - nauwkeurigheid: mate van detail waarmee een term wordt uitgedrukt;
 - actualiteit: de tijd, verstreken sinds de laatste mutatie van een gegeven, in relatie tot daartoe opgestelde normen. Actualiteit wordt beschouwd als een aspect van geldigheid;
 - volledigheid: mate waarin een gegevensbank juist is gevuld met termen in relatie tot het in werkelijkheid voorkomende aantal;
- *bruikbaarheid/nut*: behoefte aan een gegeven of gegevensgroep;

- *beschikbaarheid*: gemak waarmee men over een gegeven kan beschikken;
- *exclusiviteit*: adequaatheid van afscherming van gegevens.

De eisen kunnen worden uitgewerkt tot indicatoren. Een voorbeeld: indicaties voor semantische correctheid kunnen worden geleverd door de aanwezigheid van homoniemen en synoniemen. Een homoniem is een term met verschillende betekenissen. Een synoniem is een term, die dezelfde betekenis heeft als een andere term.

Om de kwaliteit van de gegevens zo hoog mogelijk te laten zijn, staan diverse instrumenten ter beschikking. We noemen kosten/baten-analyse ter ondersteuning van planvorming, procedures voor het handhaven van de kwaliteit van de gegevensverzamelingen en technische maatregelen ter beveiliging van gegevens.

Aan de basis van gegevensmanagement dient een adequate registratie te liggen. Een instrument om deze te realiseren, is een metagegevensbank.

Metagegevensbank (metadatabase)

Een metagegevensbank bevat gegevens over gegevens. Dit type gegevens wordt ook wel metagegevens genoemd. Met behulp van de metagegevens is de gegevensmanager in staat antwoorden te geven op vragen als:

- welk bestand bevat welke gegevens;
- welk informatiesysteem maakt gebruik van welke bestanden;
- welk informatiesysteem gebruikt welk gegeven;
- welke organisatie-eenheid is verantwoordelijk voor welke gegevens;
- bij welke organisatie-eenheid ontstaan gegevens en waar gaan gegevens naartoe;
- welk bedrijfsproces gebruikt welke systemen,

en overzichten te produceren zoals:

- logisch gegevensmodel: de onderscheiden entiteitstypen en hun onderlinge associaties;
- overzicht van homoniemen en synoniemen bij entiteitstypen, attributen en gegevens.

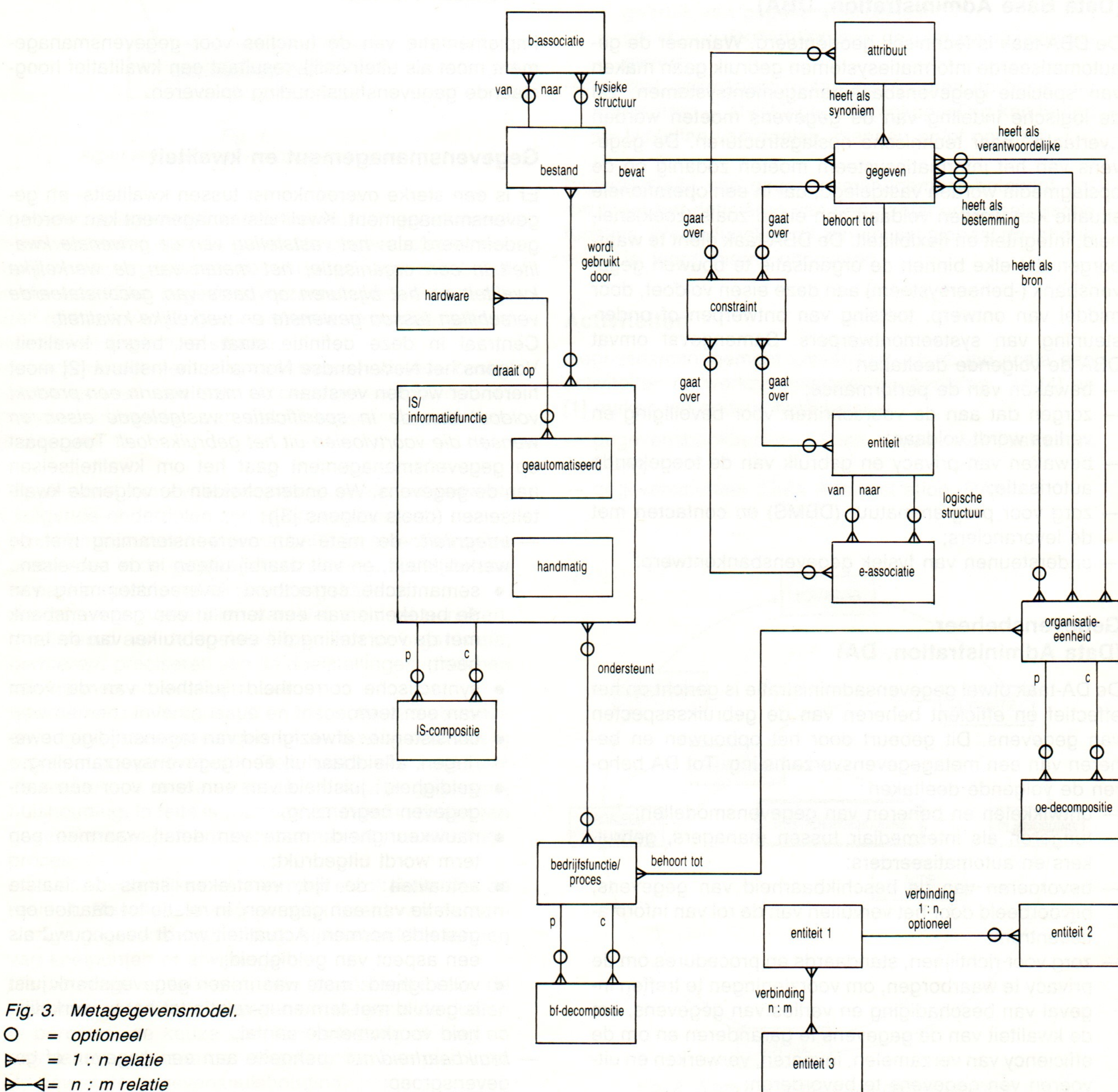


Fig. 3. Metagegevensmodel.

Om de inhoud van een metagegevensbank te verduidelijken, geven we een toelichting op een door ons ontwikkeld en toegepast logisch gegevensmodel van een metagegevensbank ofwel metagegevensmodel (fig. 3).

Centraal staat de meta-entiteit *gegeven*. Dit is een voorkomen van de meta-entiteit *attribuut* in de meta-entiteit *entiteit*. Dit voorkomen van attribuut kan in de benaming een synoniem opleveren, hetgeen de betekenis van de associatie tussen attribuut en gegeven verklaart. Attribuut moet in feite worden opgevat als de ideaaltypische definitie van een attribuut, terwijl gegeven het attribuut in een bepaalde context weergeeft.

De voorkomens van de meta-entiteit entiteit kunnen associaties met elkaar hebben (*e-associatie*), waardoor uiteindelijk de *logische structuur* van de gebruikte gegevens ontstaat. De toegestane waarden voor attribuut, gegeven, entiteit en associatie tussen entiteiten kunnen worden beperkt door *constraints* ofwel beperkingsregels.

Ieder (voorkomen van) gegeven heeft één of meer associaties met één of meer voorkomens van de meta-entiteit *organisatie-eenheid*. Voor ieder gegeven is tenminste één organisatie-eenheid verantwoordelijk. Verder zal ieder gegeven afkomstig zijn vanuit tenminste één organisatie-eenheid (de *bron-associatie*) en zal mogelijk worden gebruikt door een organisatie-eenheid (de *bestemming-associatie*). Hierbij is de beperking gemaakt, dat een gegeven niet „via” één of meerdere organisatie-eenheden naar de uiteindelijke bestemming gaat, waarbij de „tussengeschakelde” organisatie-eenheden niets toevoegen aan het gegeven. Een organisatie-eenheid kan eventueel worden gedecomposeerd.

Ieder gegeven komt voor in precies één *bestand*. Tussen bestanden onderling kunnen verbindingen bestaan. Dit is vastgelegd in *b-associatie*. Te zamen vormt dit de *fysieke structuur* van de gegevensverzameling. Een bestand wordt gebruikt door één of meer voorkomens van een *informatiesysteem* (IS). Dit informatiesysteem kan zowel *geautomatiseerd* als *handmatig* zijn. Indien het informatiesysteem is geautomatiseerd, draait het op één of meer *apparatuurlijnen*. Ieder informatiesysteem kan worden gedecomposeerd in een hiërarchie van deelsystemen. Deze hiërarchie is vastgelegd in *IS-compositie*. Een deelsysteem op het laagste niveau van deze hiërarchie is een *informatiefunctie*. Deze constructie geeft de mogelijkheid om, afhankelijk van de wens van de gebruiker, de creatie- of gebruiksassociatie op ieder willekeurig detailniveau van systeemdecompositie vast te leggen.

Een informatiesysteem wordt gerechtvaardigd door de ondersteuning van een *bedrijfsfunctie* of *proces*. De ondersteunt-associatie is optioneel om zo eventuele, overbodige informatiesystemen te kunnen opsporen (en via de gebruikt-associatie de overbodige gegevens). Iedere bedrijfsfunctie kan worden gedecomposeerd in een aantal deelfuncties.

Een metagegevensbank met al deze gegevens is een onmisbaar instrument voor de uitoefening van de IRM-functie. Met de groei van de functie vanuit een DBA via een DA naar een IRM-functie kan de metagegevensbank meegroeien. Hierbij moet worden bedacht dat het vastleggen van alle gegevens die hiervoor zijn genoemd, een niet-triviale bezigheid is. Door het op deze wijze evolutionair verankeren van de functie in een organisatie zal de organisatie kunnen meegroeien en zo overtuigd raken van het nut van zo'n vastlegging. Op deze wijze zal de noodzakelijke (meta)gegevensverstrekking beter te regelen zijn.

Afsluiting

In dit artikel is getracht een beeld te schetsen van het begrip gegevensmanagement. Vanzelfsprekend is het beeld lang niet volledig en hebben we ons moeten beperken tot het aanstippen van de ons inziens belangrijkste aspecten van gegevensmanagement. In volgende artikelen zal dieper op aspecten van gegevensmanagement worden ingegaan. Ook staan ontwikkelingen op het gebied van gegevensmanagement niet stil. Eén van de meest besproken onderwerpen is daarbij de definitie van de zogenaamde *repository* binnen het AD/CYCLE-raamwerk van IBM. De repository vervult als metagegevensbank-definitie binnen dit raamwerk van standaards een centrale rol.

Literatuur

1. Bots, J. M., E. van Heck, V. van Swede, *Bestuurlijke informatie-voorzorging, een handboek voor de mondige gebruiker*. Cap Gemini Publishing, 1990.
2. Gielen, L. J. M. W., G. J. P. Swinkels, *Kwaliteitsbeheersing bij systeemontwikkeling*. Compact, zomer 1990, p. 3 - 12.
3. Mollema, K. IJ., *Informatiekwaliteit*. Handboek Informatica, mei 1991, p. 1 - 17.
4. Nederlands Genootschap voor Informatica NGI, *Functies in de informatica*. Amsterdam 1986.
5. Vereniging van Nederlandse Gemeenten VNG, *De organisatie de baas*. Groene Reeks 113, VNG Uitgeverij, 's-Gravenhage 1991.