

Gegevensmanagement - een inleiding

Jan Berends MSc, RI
mei 2009

Vooraf

Deze tekst is geschreven als opmaat voor de BISL thema middag Gegevensmanagement (14 mei 2009). Ze bevat een aantal stellingen en begripsomschrijvingen, hier en daar kort-door-de-bocht, om onze gesprekken en discussies op te starten.

In de periode 2002-2007 gaf ik voor de landelijke Nederlandse politie (het CIP) leiding aan de afdeling gegevensmanagement, momenteel werk ik als gegevensmanager bij de vtsPN.

Uiteraard houd ik me, ter plekke en achteraf, graag aanbevolen voor aanvullingen en verbeteringen.

I. Introductie

Gegevensmanagement is een taak van groeiend belang.

Zowel in individuele organisaties als binnen hun samenwerkingsverbanden wordt de noodzaak om deskundigheid en vaardigheden op dit terrein te ontwikkelen steeds groter.

Schijnbaar bestaat het vakgebied in organisaties al tientallen jaren.

Inderdaad: de term “gegevens” bestaat ongeveer even lang als er computerapplicaties zijn, en een zekere mate van *gegevensbeheer* heeft daarbij altijd wel een rol gespeeld.

Maar de feitelijkheid van 2009 is, dat gegevensmanagement in de meeste organisatie nog zeer in de kinderschoenen staat. En daar doen de activiteiten onder de noemer van BI of van Bisl nauwelijks iets aan af.

Gegevensmanagement als vakgebied is *in ontwikkeling*, en nog lang niet klaar.¹

Dit korte, inleidende, artikel is bedoeld om een aantal begrippen te poneren en aan te scherpen.

We stippen diverse termen en knelpunten aan.

‘Gegevensmanagement’ is daarin de dominante *manier van kijken*, dat wil zeggen: een *organiseerprincipe*.

Met behulp van deze optiek kan men exacter omgaan met vraagstukken op het terrein van automatisering, informatiearchitectuur en informatiemanagement.

Naast andere nuttige ordeningen, zoals Bisl.

Door gegevensmanagement als vakgebied te ontwikkelen en bewust in te richten, door krachtig voor de invalshoek van gegevensmanagement te kiezen, voegen we een flinke kostenbespaarder en productiemotor aan ons IM / ICT arsenaal toe. Een motor die in een tijd van toenemende keten- en netwerkvervlochten niet gemist kan worden.

II. Begripsbepaling

Na deze ietwat polemische introductie, nu enkele begrippen die in de kern van ons model moeten vormen. Binnen de landelijke Nederlandse politie heb ik vanaf 2002 een reeks termen geformuleerd, om een organisatorisch begin te kunnen maken met het vakgebied, en om duidelijk te maken waar het team aan werkte.²

We beperken ons hier tot enkele begrippen, die we hier slechts beperkt uitwerken, en na de korte definitie vooral áfbakenen van de gangbare misverstanden.

Om het begrip gegevensmanagement GM de juiste kracht te geven het volgende voorstel voor een definitie:

<p>Gegevensmanagement is het geheel van activiteiten die ertoe leiden dat een organisatie [of keten] op het juiste moment over de benodigde gegevens beschikt.</p>

Eenmaal geformuleerd, lijkt het haast vanzelfsprekend...

Maar als we er bij stil staan, betekent het een grote reeks van mogelijke activiteiten.

Meteen is duidelijk, dat het om veel méér gaat dan om ‘gegevensbeheer’.

Gegevensbeheer is een systeem van activiteiten die gericht zijn op de geldigheid en de kwaliteit van gegevens. Heel belangrijk; maar [slechts] één van de activiteitenclusters binnen GM!

Andere activiteiten zijn bijvoorbeeld:

- Kiezen welke gegevens de organisatie inkoop dan wel zelf produceert.

¹ Bijvoorbeeld: waar kan men een opleiding tot gegevenmanager volgen? Hoe is de certificatie geregeld?

² Voor een vollediger begrippen overzicht zie “Wat is Gegevensmanagement”.

- Zorgen voor een 'gecontroleerd' intern begrippenbeheer, zodat parallelle activiteiten vergelijkbare gegevens opleveren.
- Organiseren dat gegevensarchitectuur en gegevensmodellen volgend zijn aan de informatie-behoefte (en flexibel ...).
- Bepalen wat de verwervings- en beheer-kosten zijn van de huidige gegevens; en welke normen daarin moeten gelden.
- Bepalen wat de baten van de huidige/toekomstige gegevens zijn, en daarop sturen.

Gegevensmanagement is een *integrale* en een *integrerende* activiteit binnen [verbanden van] organisaties. De uitdaging is: te *managen* dat de organisatie³ te allen tijde, voorspelbaar, zal beschikken over alle gegevens die zij nodig heeft, van de juiste kwaliteit, tegen de verantwoorde kosten, en binnen wet en regel.

Als samenvatting van de term **Gegeven** het volgende voorstel:

Een gegeven is de registratie van een feit.

&

Een feit is: een verschijnsel in de werkelijkheid.

De essentie is, dat een gegeven een *verwijzing* is naar de werkelijkheid. Direct of indirect. Deze verwijzing *dient* plaats te vinden op de manier waarop de organisatie deze werkelijkheid interpreteert en relevant vindt. Een gegeven heeft meer *inhoudelijke kwaliteit* naarmate die verwijzing plaatsvindt op deze beoogde manier.

Deze opvatting van het begrip gegeven is *primair* gebaseerd op een functionele optiek. Je moet er wat aan hebben. Gegevens zijn er niet voor zichzelf, ze dienen generieke en specifieke doelen.

Natuurlijk is er de afgelopen 30 jaren op het gebied van gegevensmodellering en gegevenstechniek veel analytische kennis ontwikkeld.

Van daaruit zijn belangrijke inzichten en termen beschikbaar m.b.t.⁴:

- gestructureerde en ongestructureerde gegevens;
- ER- en OO-modellering;
- elementaire en samengestelde gegevens;
- verbale en non-verbale gegevens;
- metagegevens;
- standaardgegevens;
- gegevensstandaarden;
- referentiegegevens;
- analoge en/of digitale gegevens.

De wijze waarop de organisatie gebruik maakt van deze ICT-kennis, heeft vaak een enorme invloed op het speelveld van gegevensmanagement. Zo leren alle applicatiegebruikers om te "denken" in de termen die de applicatie hen aanbiedt en vraagt; uiteenlopend van toetscombinaties tot gegevenstypen. Een ander voorbeeld is de afstemmingsproblematiek tussen applicaties met ongelijke gegevensmodellen. Of de last die je kunt hebben van niet meer actuele geïmplementeerde gegevensarchitectuur.

De keuze voor een functionele optiek betekent, dat de gebruiker / werkprocessen leidend moeten zijn bij het vaststellen (en beoordelen) van de kwaliteit en noodzakelijkheid (bijv. de kostprijs) van de gegevens.

Dit lijkt logisch. Maar in de werkelijkheid van organisaties komt dit (na de initiële implementatie van een applicatie) haast *nooit* voor.

Het begrip **Informatie** moet onderscheiden worden van gegevens en van 'Informatieproducten'.

Het voorstel is, om Informatie te analyseren als:

Informatie is het antwoord op een vraag

De vraag is in het algemeen handelingsgericht.

Het antwoord dient als trigger voor een handelingskeuze

In het algemeen bevindt de vraag (en ook het antwoord) zich in het denksysteem van een beslisser.

³ en ook [nog maar eens een keer gezegd]: de ketens en netwerken waarin zij acteert

⁴ Een beschrijving van deze begrippen is op veel plekken elders beschikbaar.

Hier stippen we deze termen slechts even aan.

Dat is natuurlijk meestal het hoofd van een mens; maar het kan ook een beslissend computersysteem zijn, of en medisch meetinstrument. De vraag kan 'latent' zijn. Het antwoord functioneert dan als een impuls. Informatie bestaat dus alleen maar gedurende het proces van vraagbeantwoording, en is vaak kwalitatief en onzichtbaar. Gegevens daarentegen bevinden zich op gegevensdragers (papier, computerbestanden, schermen) en zijn er vaak ook nog als de vraag al is afgehandeld.

Een **Informatieproduct** is de praktische brug tussen beide.

Een informatieproduct is:
een presentatiewijze van gegevens,
gericht op overdracht van informatie.

Het bekendste informatieproduct is het Formulier. Informatieproducten zijn 'voorgekookte' beantwoordingen. Hun bestaan berust op het feit dat veel vragen in een organisatie *herhalend* zijn.

Bekend probleem is, dat formulieren: niet meer geldig zijn, verkeerd begrepen worden, of tot overbodig invullen / teruglezen leiden. Met formulieren en hun misfits gaat veel geld verloren.

De definitie van gegevensmanagement (boven) betekent, dat ook het optimaliseren (managen) van de set van informatieproducten een gegevensmanagement- taak is: nl. het ter beschikking stellen van gegevens.

Ketens

Het uitwisselen van informatie ... eigenlijk: het uitwisselen van gegevens, met het doel de wederzijdse ontvangers aan informatie te helpen ... gebeurt altijd in een keten.

Er is een productieproces gaande, binnen een organisatie danwel organisatieoverstijgend. Elk productieproces vraagt zowel qua 'onderstroom' (input) als qua 'bovenstroom' (managerial, stuurinformatie) om gegevens.

Deze keteninteracties vragen om *standaardisatie*. Geen van de afzonderlijke partners is eigenstandig bevoegd of in staat om zijn communicatiegewoonten [of gegevens vastleggewoonten] aan de andere partners op te leggen, al zal dit vaak wel geprobeerd worden...

Zo goed als er binnen een organisatie 'verdiend' kan worden, door de afzonderlijke gegevensactiviteiten van medewerkers, afdelingen en systemen in te bedden in een gegevensmanagement raamwerk, zo goed kan er ook in samenwerkingsverbanden nog veel verdiend worden door gegevens effectief te delen/verwerven/beheren/uit te nutten etc.

Een voorbeeld van een maatschappelijke aanzet daartoe zijn de landelijke Basisregisters; het EPD; de Europese Geo-standaarden.

Applicaties

Gegevensmanagement is niet denkbaar zonder een verhouding met applicaties en architecturen.

Klassieke bedrijfsapplicaties (bedrijfsinformatiesystemen) doen in hoofdzaak twee zaken: ze ondersteunen een procesgang en ze behandelen gegevens. Dat laatste resulteert dan vaak in informatieproducten (output), fysiek of virtueel. Historisch waren (zijn) applicaties vaak afdelings-gebonden, en waren (zijn) gegevens vaak applicatiegebonden.

Gegevensmanagement is noodzakelijk om deze twee hobbels te laten verdwijnen.

Er is een driehoeksverhouding tussen applicaties, bedrijfsprocessen en gegevens.

Daarin zijn bedrijfsprocessen in principe leidend.

(1)

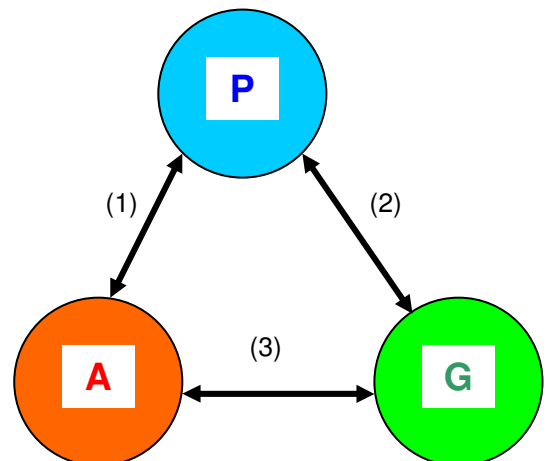
In principe bepaalt het proces, welke functionaliteiten de applicatie moet bieden. Veranderingen in de werkwijze moeten leiden tot aanpassingen in de applicatie.

In de praktijk is deze relatie niet zo duidelijk. Eerder is er sprake van een wisselwerking: de applicatie heeft grote invloed op de manier waarop het proces wordt uitgevoerd

(2)

In principe bepaalt het proces, welke informatie er nodig is, en hoe deze informatie in de vorm van gegevens wordt vastgelegd en teruggehaald. Vooral de keuze van het gegevensmodel (de gegevenstypen + gegevensstructuren) en de standaardgegevens zijn voor het bedrijfsproces belangrijk.

In de praktijk wordt sturing op het *gegevensmodel* alleen enigszins bereikt in de fase van systeembouw.



(3)

Er is een wisselwerking tussen applicatie en gegevens. Het gegevensmodel van de applicatie kent inhoudelijke en procesmatige gegevens. Dit model bepaalt welke gegevens kunnen worden vastgelegd. Omgekeerd: de behoefte aan nieuwe gegevens heeft een sturende invloed; de applicatie wordt (in principe) aangepast aan de nieuwe eisen.

In de praktijk is het aanpassen van een operationele applicatie aan gegevens-wensen om drie redenen moeizaam. * Het huidige gegevensmodel is slecht gedocumenteerd [:niet aankomen svp]; * Het systeem is niet gebouwd op aanpasbaarheid; * aanpassing van het datamodel leidt tot down-time (kosten) en tot dataconversie (risico's).

Al met al spreekt het niet vanzelf, dat applicaties “compliant” zijn ten opzichte van gegevensmanagement-beleid!

Kwaliteit

Het begrip Kwaliteit relateren we aan de diverse *functies* (nuttige eigenschappen) van een *object*.

Een object kan alles zijn wat de organisatie nodig heeft; bijv. een auto, een applicatie of een gegeven.

Het object, bijvoorbeeld het gegeven, heeft voor de organisatie een doel.

Een functie kan elke voor de organisatie wenselijke eigenschap zijn; bijv. de snelheid, de grootte.

De functie is nodig, omdat het object op die manier iets voor de organisatie kan betekenen.

We definiëren dan:

De Kwaliteit (van een functie van een object) is de mate waarin het object geschikt is om die functie werkelijk - als positieve bijdrage aan het beoogde werk - toe te passen.

Een applicatie bijv. kan allerlei functionaliteiten hebben, en van elke functionaliteit kan de kwaliteit gemeten worden [t.o.v. de ideale situatie].

Functionaliteit benoemt de voor het werk relevante ‘dimensies’. Kwaliteit geeft de ‘scores’ op die dimensies. Hoewel kwaliteit niet altijd hard gemeten kan worden, kan men er meestal wel een kwantitatieve maat aan geven, bijvoorbeeld een getal tussen de 0% en 100%. De waarde “100%” wil dan zeggen, dat het object in dit opzicht volledig functioneert.

Het vaststellen van de gewenste gegevenskwaliteit en het inrichten van beheer is deel van het gegevensmanagement van een organisatie. Dan onderscheiden we tussen kwaliteitsaspect, -norm en -afpraak. Een kwaliteitsaspect is een dimensie van kwaliteit. Er worden ongeveer 30 dimensies onderscheiden. Bijv. de *juistheid* van een gegeven. Vb.: het woonadres in GBA is echt het fysieke woonadres van betrokkene. Een kwaliteitsnorm is een referentiewaarde voor een kwaliteitsaspect. Deze wordt gebruikt in de dialoog tussen zender en ontvanger (producent en klant), om te bepalen of de levering voldoet. Een kwaliteitsnorm is bijvoorbeeld, dat de gemeente zorgt dat bij minimaal 95% van de situaties waar administratief adres verschilt van een fysiek adres, vermeld staat dat dit adres “in onderzoek” is. Een kwaliteitsafpraak is een afspraak tussen partijen met betrekking tot een kwaliteitsaspect en normen daarvoor, inclusief maatregelen. Een afspraak kan zijn, dat een gemeente die deze norm niet haalt binnen twee maanden alsnog de norm zal halen, of dat anders hun GBA wordt afgekoppeld van het gemeente-net.

III. Vormgeven aan Gegevensmanagementbeleid

In het voorgaande zijn enkele voorbeelden uit het domein van gegevensmanagement aangestipt.

Als laatste onderdeel van deze introductie besteden we kort aandacht aan enkele operationele discussies.

De bottom-line is, dat de gebruiksorganisatie middelen in handen krijgt om haar werk beter / goedkoper te doen. Gegevens zijn belangrijke middelen. Gegevens zijn namelijk de meest directe koppeling die de organisatie via systemen met de haar omringende wereld kan maken. Gegevens liggen daarom:

1) aan de basis van veel sturing en strategie; en 2) aan de basis van veel reëel operationeel handelen.⁵

Het activiteitenpalet van de gegevensmanagement optiek past bij huidige trends en ontwikkelingen, zoals:

1. interne bedrijfsintegratie.

Applicaties raken [in het verlengde van bedrijfsprocessen] toenemend verstrengeld.

Om “spaghetti” [messaging tussen applicaties] te voorkomen of tegen te gaan moeten gegevens gedeeld worden en bedrijfsbreed ontwikkeld & onderhouden.

In veel organisaties is nog de nodige rommel op te ruimen, bijv. applicaties die onvoldoende geïntegreerd werken.

⁵ Natuurlijk is er daarnaast ook wel direct contact met de omgeving (zonder tussenkomst van systemen). Bijvoorbeeld de politie agent die ingrijpt bij een verkeersongeluk; of de hoofdcommissaris die de burgemeester langs krijgt.

2. externe bedrijfsintegratie.
Integratie van productieketens leidt tot ebXML, EPD, organisatie overstijgende standaarden, e.d. Wil de organisatie daarin effectief kunnen acteren, dan moet men a) zijn eigen huiswerk doen (bijv. zijn eigen gegevens kennen) en b) zich in zijn ketens actief opstellen als gesprekspartner (bv. weten welke gegevens men nodig heeft of wil leveren).
3. internetprocessing; externe data bronnen
Beschikbaarheid van externe bronnen en verwerking stelt de organisatie voor de vraag: welke data kopen we, welke willen we zelf nog verwerken?
Alleen als men een ontwikkelde gegevensmanagement functie heeft, kan men sturing houden op dit proces, en daarin ook evt. stappen terug doen.
4. opgelegde standaarden
compliant zijn met bijv. basisregistraties of wetgeving betekent dat men grip moet hebben op zijn eigen procesgang, en niet overgeleverd aan applicaties waarvan de functionaliteit louter bepaald wordt door decentrale voorkeuren.

Gegevensmanagement is géén onderdeel van BISL en géén onderdeel van centrale bedrijfsfunctie alleen. Wel moet er nauw worden afgestemd.

Natuurlijk is het wel dénkbaar, om "functioneel beheer" of BISL als begrip zodanig te verbreden, dat ook de strategische t/m operationele aspecten van GM er allemaal onder kunnen worden begrepen.

Maar waar ben je dan mee bezig? Je zou dan ook personeelszaken weer onderdeel kunnen maken van de salarisbetaling. Het gaat er juist om, goed na te denken over de consequenties die je applicatie / systeem keuzes hebben voor de gegevens, bedrijfsbreed. Daarvoor is het belangrijk om een specifieke, gegevensgerichte invalshoek te articuleren [verbijzonderen].

Net zo goed als het belangrijk is om applicatiefunctionaliteit te beheren.

In schema zou men kunnen zeggen:

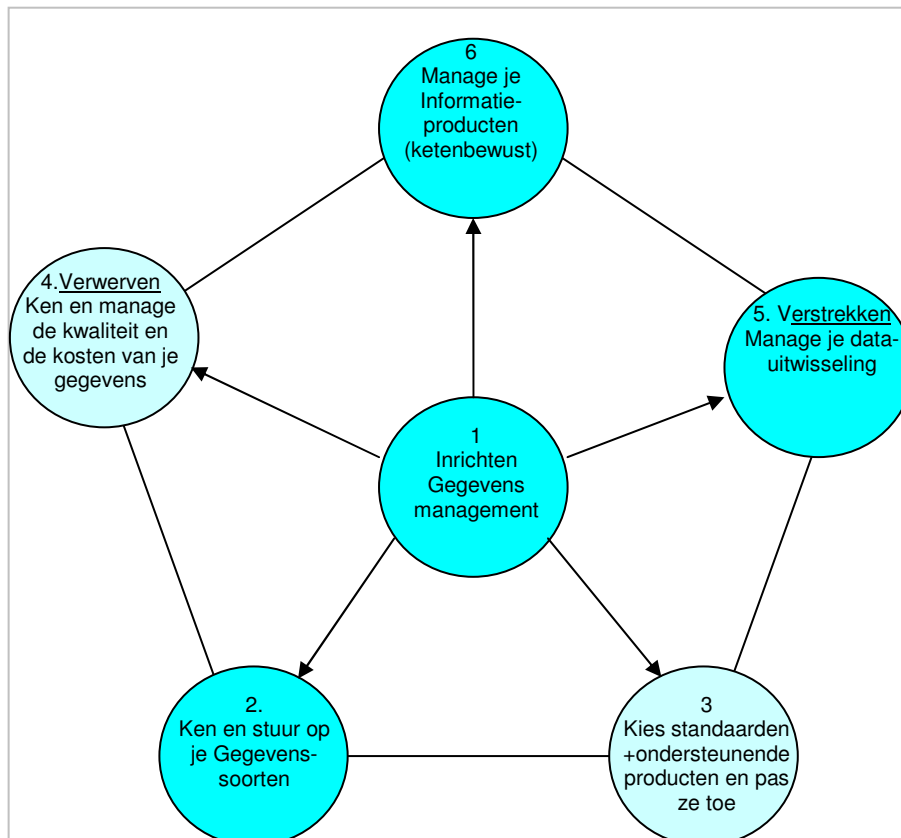
$$IM = IA + GM + FB.$$

Informatiemanagement (de gehele I-functie, incl. doelsturing van de 'net- en hardware' kant) = Informatiearchitectuur (vraagzijde) + Gegevensmanagement (antwoordzijde) + FB (procesondersteuning).

Gegevensmanagement is niet louter centraal.

Ook de decentrale gegevensbeheerder en decentrale programmeur werken aan gegevensmanagement.

Wel zou het handig zijn de visies op te schalen, en geïntegreerd te werken.



Ter afsluiting nog een overzicht van enkele belangrijke organisatorische uitdagingen. Het inrichten van gegevensmanagement (1) betekent acties in 5 taakvelden.