

# Informatiefabrieken en meta-informatie

Ir. Hein J.M. Corstens  
URBIDATA bv  
hein.corstens@urbidata.nl

## Samenvatting

Informatisering houdt in dat fysieke productieprocessen getransformeerd worden tot informatieprocessen, processen van informatieproductie en –beheer. Deze dienen doelmatig te worden opgezet, als een soort fabriek, een informatiefabriek dus. Daarin wordt informatie geproduceerd uit gegevens, zoals die voorkomen in de veelheid en diversiteit aan bestaande applicaties en zoals die uit het internetverkeer onttrokken worden. Deze gegevens worden geïntegreerd, hetzij voor opslag in datawarehouses en operational data stores, waarna ze omgezet worden in informatie voor ondersteuning van operationele, tactische of strategische beslissingen, hetzij voor direct gebruik in applicaties (waarmee Enterprise Application Integration (EAI) vorm krijgt). Voor die integratie heeft URBIDATA een hulpmiddel ontwikkeld, de Universal Data Store (UDS). De informatiefabriek dient bestuurd te worden, net zoals een gewone fabriek. Daarvoor is informatie nodig, in dit geval meta-informatie. Meta-informatie – die gemaakt wordt uit metadata – is daarmee de onmisbare basis voor informatisering. De informatisering wordt al snel gevolgd door een volgende fase, die van de meta-informatisering. Daarin dient integratie tot stand te komen van metadata uit verschillende bronnen, gericht op de besturing van de ‘meta-informatiefabriek’.

## Besturingsmodel

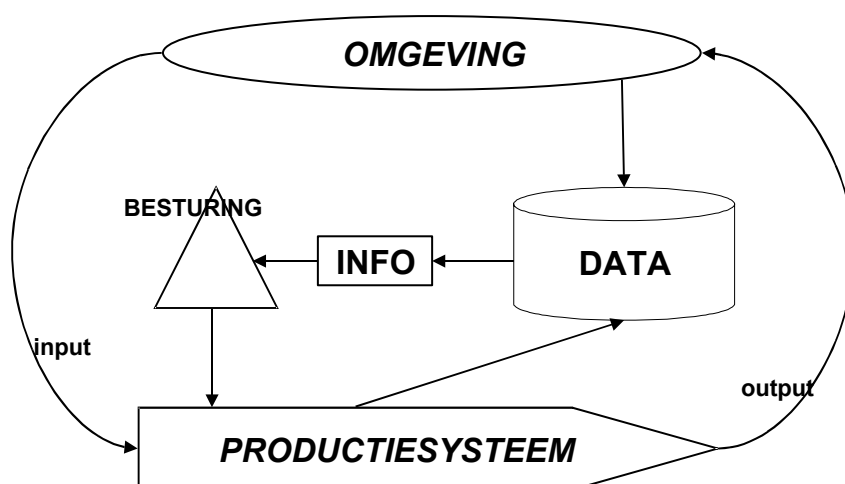
De informatie-infrastructuur wordt steeds belangrijker voor bedrijven en overheidsorganisaties. De informatie-infrastructuur is het geheel van voorzieningen voor algemeen gebruik door en toegankelijk voor alle bij het primaire proces betrokken organisatieonderdelen. De informatie-infrastructuur bestaat uit componenten voor ontwikkeling en beheer van gegevens, applicaties, configuraties, communicatie en organisatie. De informatie-infrastructuur beschouwen wij als een geheel dat BESTUURD moet worden en wel zodanig dat een optimale aanpassing tussen informatie-infrastructuur en haar omgeving gerealiseerd wordt. In dat opzicht is een informatie-infrastructuur wel te vergelijken met een ecosysteem. Net zoals ecosystemen in de natuur moet een informatie-ecosysteem aanpasbaar zijn. Op den duur veranderen zowel het evenwicht tussen verschillende componenten en hun onderlinge relaties als de omgeving. Een gezond ecosysteem ontstaat alleen als de componenten aanpasbaar, veranderbaar en in evenwicht zijn. Nu moeten die kenmerken uiteraard wel gericht zijn op een bepaald doel, bijvoorbeeld continuïteit. De onderlinge aanpassing van systeem en omgeving is daarom te zien als een besturingsproces. Daarin speelt terugkoppeling een essentiële rol. In plaats van informatiesystemen en informatie-infrastructuren te vergelijken met gebouwen en de informaticus met een architect trek ik liever een vergelijking met een landschap of desnoods een stad, welke voortdurend groeit en bloeit en waar men doelen tracht te bereiken door voortdurend kleine ingrepen te treffen, waarna men het systeem zijn gang laat gaan. Voortdurend wordt er geëvalueerd, gevolgd door bijsturing. Terzijde merk ik op dat ik in dat verband een vraagteken plaats bij de manier waarop sommigen denken de plannen van de Nederlandse overheid voor authentieke registraties te realiseren, namelijk als een geheel dat volgens een blauwdruk ‘weggezet’ moet worden.

Maar ter zake. Wat is besturing? Besturing is het voortdurend corrigeren van een systeem in de richting van een doel. Letterlijk zien we dat in de besturing van een auto. Afhankelijk van de termijn waarop het doel bereikt moet worden spreken we over strategische, tactische dan wel operationele besturing. Voorbeeld: bij het geven van een inleiding is het strategische doel de aanwezigen te informeren. Het tactische doel is de aanwezigen te vermaken. Als de inleider ziet dat er meerdere aanwezigen in slaap vallen moet er bijgestuurd worden, bijvoorbeeld door het luid vertellen van een leuke anekdote. Als nu de aandacht teveel naar bijzaken gaat moet er weer bijgestuurd worden richting onderwerp. De operationele besturing betreft onder meer de bediening van de laptop en de projector.

De toepassing van genoemde driedeling op een bedrijf levert de volgende indeling in informatiesystemen op:

- business intelligence systemen voor de strategische besturing (strategische planning; research, analyse en voorspelling);
- business management systemen voor de tactische besturing;
- business operations systemen voor de operationele besturing (dit zijn veelal legacy systemen).

Het besturingproces wordt gerepresenteerd door onderstaand plaatje.



*Besturingsmodel.*

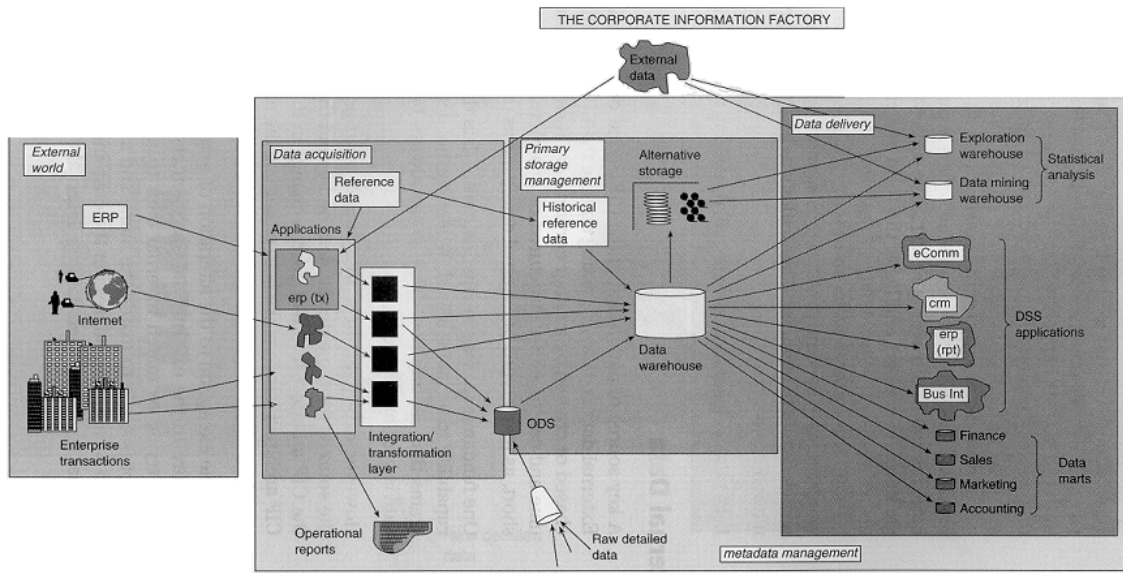
De besturing heeft betrekking op een productiesysteem, dat via input en output in relatie staat met de omgeving. De besturing – symbolisch als een piramide weergegeven met de strategische besturing aan de top en de operationele besturing aan de basis – vindt plaats op basis van informatie, die gemaakt wordt uit data, die betrokken worden uit het productiesysteem zelf en de omgeving.

## **Informatisering en informatiefabrieken**

Informatisering is een ontwikkeling, waarbij fysieke productieprocessen langzaam maar zeker getransformeerd worden tot informatieprocessen. Daarin bestaat de dagelijkse activiteit uit verwerking van gegevens over de productie zelf én over de omgeving waarin dit plaatsvindt tot informatie, waarmee het productieproces wordt bestuurd. Het productieproces zelf is geautomatiseerd. Een fase verder is de besturing van de informatieprocessen. Daarbij worden gegevens

over het informatieproces verwerkt tot informatie op een hoger niveau. Het informatieproces zelf wordt daarbij geautomatiseerd. Uiteindelijk ontstaat een ‘informatiefabriek’, een geheel van processen waarin de grondstof gegevens verwerkt wordt tot informatie.

Door Bill Inmon, één van de grondleggers van Data Warehousing, is het concept Corporate Information Factory (CIF) uitgewerkt [Inmon e.a.]. In onderstaande figuur wordt de CIF weergegeven.



Corporate Information Factory (bron: [Inmon e.a.]).

Een CIF lijkt veel op een gewone fabriek \*. Grondstoffen en onderdelen gaan een fabriek in en worden opgeslagen. De ruwe goederen worden in assembleerlijnen omgevormd tot een diversiteit van halffabrikaten en eindproducten. De samenstellende delen van de CIF zijn:

- Buitenwereld: dit is het geheel van processen, waarvoor de informatie uiteindelijk bedoeld is en van waaruit de gegevens het systeem binnenkomen.
- Applicaties deze zorgen voor verzameling van gedetailleerde transactiegegevens, directe interactie met de gebruiker, controle en aanpassing van gegevens en redactie van data. Veelal vormen de applicaties een onsamenhangend geheel met data die niet geïntegreerd zijn, het is ‘legacy’. De nadruk ligt op snelle response. In toenemende mate is het internet de bron van de data die tot grondstof van de informatiefabriek dienen.
- Integratie- en transformatielaag: deze bestaat uit programma’s die de niet geïntegreerde data uit de applicaties combineren en transformeren tot corporate data.
- Operational Data Store (ODS): dit is een hybride structuur, die enerzijds op operationele processen, anderzijds op besluitvormingsondersteuning is gericht. Deze structuur maakt de ODS de meest complexe component van de CIF. De data in de ODS zijn geïntegreerd, vluchtig (ze worden geactualiseerd als onderdeel van operationele processen), actueel en gedetailleerd.
- Data Warehouse, de basis voor alle strategische beslissingsondersteunende processen. De data in het data warehouse zijn: ‘subject oriented’ (gericht op onderwerpen zoals Klanten,

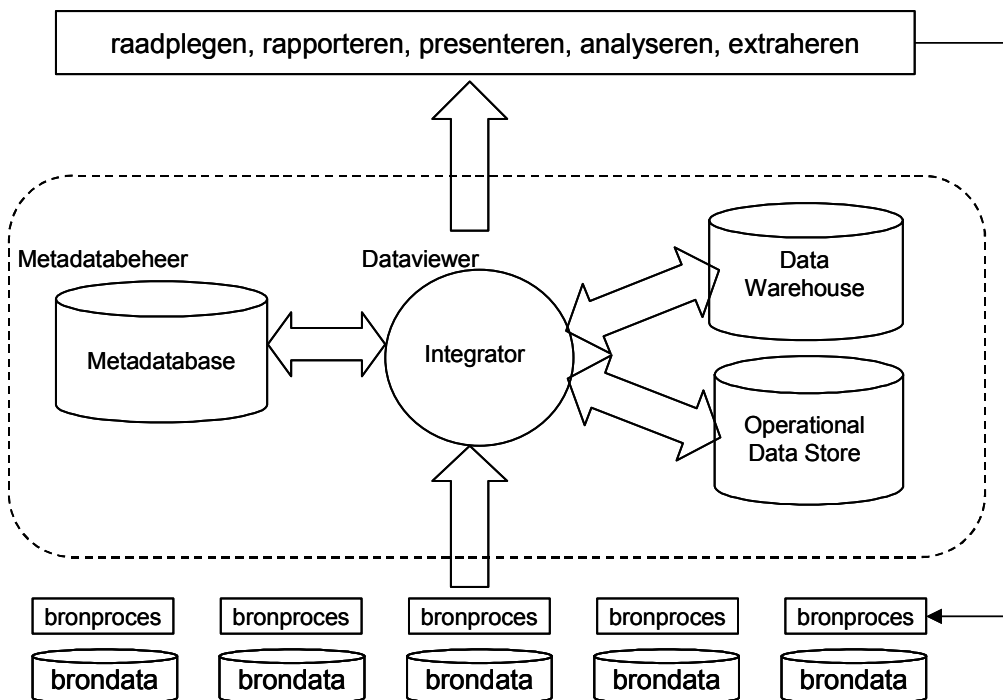
\* Inmiddels hebben Inmon en de zijnen een variant van de CIF voor de overheid bedacht: de GIF, wat staat voor ‘Governmental Information Factory’

Producten, Verkopers, Transacties, Orders, Policies, Accounts, Verplaatsingen), dus NIET functie- of applicatiegericht, geïntegreerd en tijdsvariant (ieder record heeft een geldigheidsperiode of -moment). Ze worden eenmalig opgeslagen en blijven onveranderd in het Data Warehouse staan. Ze zijn niet-vluchtig (er wordt gewerkt met snapshots) en zowel samenvattend als gedetailleerd. De datastructuur is in principe genormaliseerd en wel om redundantie te elimineren.

- Data Marts: een mart is een collectie data, toegesneden op de besluitvormingsondersteuning van een bepaalde afdeling. Een data mart is meestal gedegenormaliseerd. Data Marts worden in het leven geroepen ten behoeve van beheersbaarheid, kosten en aanpasbaarheid aan wensen van de gebruiker. Data Marts worden vanuit het Data Warehouse gevuld.
- Exploration and Data Mining Warehouses: deze componenten worden gecreëerd in verband met de zeer grote queries die alle capaciteit verbruiken. Het Exploration Warehouse is een fysiek afgescheiden structuur die exclusief gericht is op onderzoeksprocessen. Het is breed opgezet en bevat veel soorten data en relaties tussen de data; doel is hier het onderkennen van patronen en het formuleren van hypothesen. Daarna ontstaat er behoefte aan een Data Mining Warehouse. Dat bevat veel feitelijke gegevens met weinig variatie in datatype; de data miner is meer gericht op het testen en valideren van hypothesen.
- Alternative Storage Component voor de tijdelijke of permanente archivering van de data.
- Internet/Intranet Components. In toenemende mate vinden interactie en communicatie van de CIF met de buitenwereld plaats via Internet. Intern geschiedt dit via Intranet.
- Metadata; hierop komen we later terug.

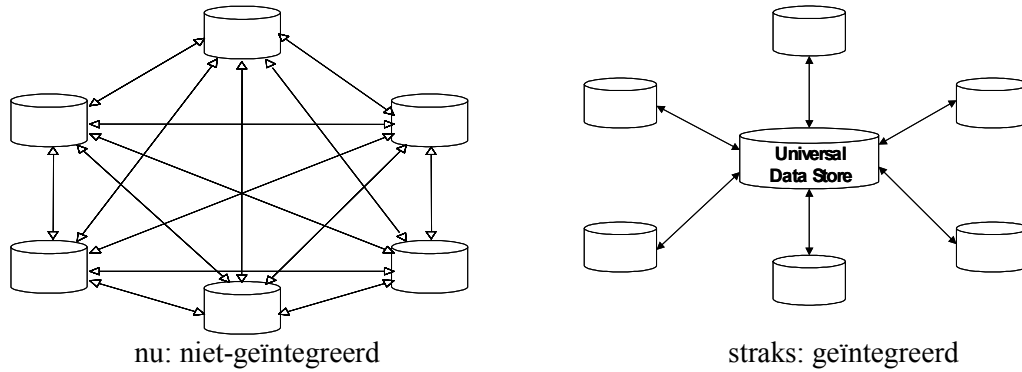
## Universal Data Store

URBIDATA heeft als basis voor informatiefabrieken integratiesoftware ontwikkeld, UDS – Universal Data Store – genaamd.



*Universal Data Store.*

De UDS is specifiek gericht op zowel RUIMTELIJKE als ADMINISTRATIEVE DATA en op TERUGKOPPELING naar de bronsystemen. Daardoor kan een vereenvoudiging in de informatie-infrastructuur bewerkstelligd worden, zoals de volgende figuur illustreert.

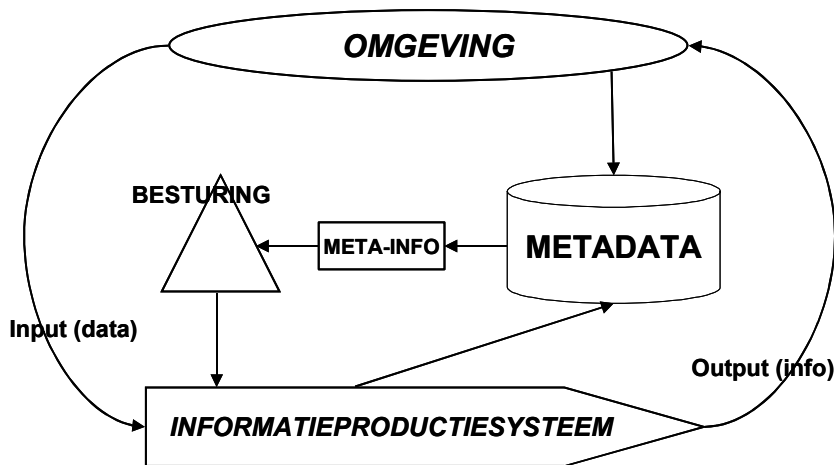


*Integratie met behulp van de UDS.*

Terzijde: de niet-geïntegreerde situatie wordt wel aangeduid als 'spaghetti'. De geïntegreerde oplossing is 'ravioli', een gestructureerd geheel van objecten, waarin alle gegevens en methoden betreffende één objecttype zijn samengebald. In de wereld van de geografische informatiesystemen wordt wel een tussenoplossing gebruikt, waarbij door de koppeling van kaartlagen op grond van ligging administratieve gegevens onderling gerelateerd worden. Deze handige 20/80 oplossing noemen we 'lasagne'.

### Meta-informatie

Ook de informatiefabriek moet bestuurd worden. Dit wordt in onderstaand plaatje weergegeven.



*Metabesturingsmodel.*

We zien dat er eigenlijk geen verschil bestaat met het besturingsmodel voor een gewone fabriek. We moeten alleen de fysieke objecten vervangen door informatieobjecten.

Meta-informatie wordt gevormd uit metadata, data over data. Deze worden in de UDS opgeslagen in een metadatabase en beheerd in de metadatamanager. In de metadata worden beschreven:

- ontwerpgegevens: datamodellen, functionele ontwerpen, technische ontwerpen, kwaliteitseisen, enz.;
- productiegegevens: fysieke modellen, testen, planning en voortgang;
- gegevens over levering en gebruik: kostenverrekening, productencatalogus, planning en logistiek.

In concreto wordt de kern van de metadata gevormd door:

- gedetailleerde beschrijving van de data (datasets, objecten, attributen, relaties, constraints), zowel qua syntax (vorm) als semantiek (inhoud);
- specificatie van de applicaties waarmee de data beheerd en benaderd worden, alsmede de bijbehorende platforms;
- relatie met de organisatie en de taken die vervuld worden;
- transformatieregels en processen;
- verversingsplanning;
- gebruik;
- verwijzingen naar documentatie (literatuur, websites);
- specifieke kenmerken; bijvoorbeeld specifieke geodetische eigenschappen.

Ook metadata worden in toenemende mate via het internet betrokken.

Metadata vormen de lijm die de componenten van de CIF bij elkaar houdt.

Het beheer van metadata heeft zeer veel aandacht, net zo als dat voor andere data geldt. In de UDS is de MetaDataManager dan ook het hart van het systeem. Belangrijk is de afbakening tussen centraal en decentraal beheer van metadata, tussen de behoefte aan autonomie en de behoefte aan samenwerken en elkaar informeren. Er dient daarin een evenwicht bereikt te worden. In iedere component van de CIF dient er een scheiding aangebracht te worden tussen gemeenschappelijke en autonome metadata. Per gemeenschappelijk metadata-element dient vastgesteld te worden welke component dit mag muteren. Het is voor alle andere componenten zichtbaar.

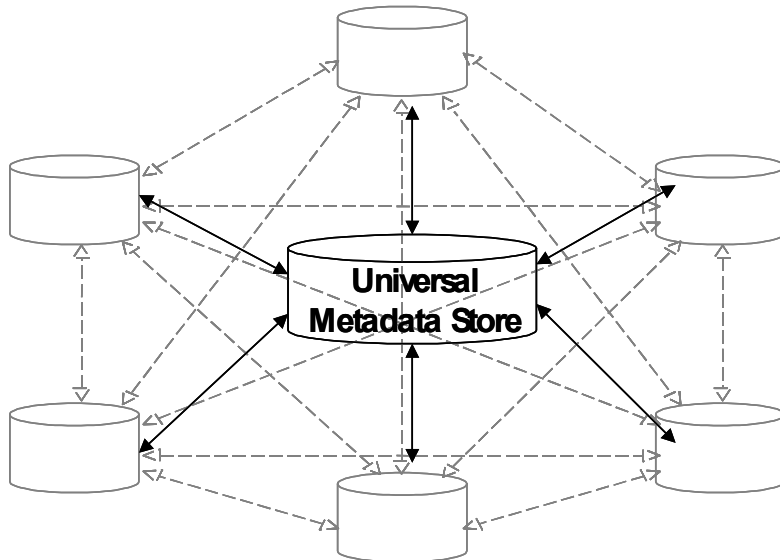
Het verkrijgen van metadata is in het algemeen een groot probleem. Aanbevolen wordt bij de ontwikkeling van de CIF te werken met tools die automatisch metadata produceren.

Overigens wordt de UDS ook ingezet om metadata te integreren. De diverse bronmetadata worden door de UDS geïntegreerd en volgens bepaalde regels aan verschillende gebruikersgroepen ter beschikking gesteld. De UDS werkt in dat verband dus als 'UmetaDS' of 'UmDS'. Dit brengt ons tenslotte op Enterprise Application Integration.

## **Enterprise Application Integration (EAI)**

De UDS kan een belangrijke rol spelen in 'Enterprise Application Integration' (EAI). Dit is het integreren van de verschillende systemen tot virtueel één systeem. Voor een gebruiker maakt het daarbij niet uit of er één groot systeem is, dan wel een verzameling van met elkaar verbonden kleine systemen. Er kan zelfs per gebruiker – door de creatie van 'portals' – een eigen view op het geheel van informatiesystemen gemaakt worden. Met zo'n portal is er integratie op user interface niveau ('sceenscraping'). EAI vergt echter integratie op dataniveau. Daarbij lezen applicaties elkaars gegevens en schrijven er ook naar weg. Dit gebeurt door 'message brokers', die gegevens routeren en transformeren (aangepast aan het datamodel van de doelapplicatie). Door gebruik van de UDS wordt consistentie gegarandeerd door toepassing van de in de metadatabase opgeslagen transformatieregels.

In feite is een centrale data store dan niet altijd meer nodig: in plaats daarvan kan men een centrale metadata store inrichten, waarin alle metadata, transformatieregels en de planning en scheduling daarvan zijn opgenomen. Met behulp daarvan kan op dataniveau onderlinge uitwisseling gestuurd worden.



*Structurering door centralisatie van metadata.*

Het zal duidelijk zijn dat zo iets ook voor de metadata zelf gerealiseerd zou kunnen worden. Uiteindelijk blijft er dan alleen nog centraal beheer op metaniveau over.

## **Conclusie**

Informatisering, de transformatie van fysieke productieprocessen tot informatieprocessen, krijgt een gestructureerde vorm in informatiefabrieken, waarin informatie wordt geproduceerd uit bestaande gegevens. De informatiefabriek dient bestuurd te worden op basis van meta-informatie. De informatisering wordt al snel gevolgd door een volgende fase, die van de meta-informatisering. Daarin dient integratie tot stand te komen van metadata uit verschillende bronnen, gericht op de besturing van de 'meta-informatiefabriek'.

## **Literatuur**

[Inmon e.a.]

W.H. Inmon, C. Imhoff, R. Sousa, 'Corporate Information Factory', 2001 ISBN: 0-471-39962-2