

Volledig nieuwe generatie BI-oplossingen dient zich aan

# BI 2.0: fixing the bugs?

Erik Franssen

**Het is ogenschijnlijk goed gebruik geworden om ontwikkeling in de IT (maar inmiddels ook ver daarbuiten) aan te duiden met de toevoeging 2.0, 3.0, 4.0 etcetera. De oorsprong van de versieaanduiding ligt uiteraard in de software-ontwikkeling.**

Versie 1.0 van een softwareproduct is meestal de eerste versie beschikbaar voor een groot publiek, versie 1.1 (of 1.2, 1.3 etcetera) geeft aan dat er beperkte functionaliteit is toegevoegd of fikse fouten zijn aangepast. Een softwareproduct verandert naar versie 2.x (of 3.x, 4.x etcetera) indien significante functionaliteit is toegevoegd. Zo bezien impliceert de toevoeging 2.0 een significante verbetering in functionaliteit, dus dat belooft veel goeds voor de eindgebruikers! Maar is dat wel zo? In dit artikel wil ik daarom de nieuwe 'versie' van Business Intelligence eens onder de loep nemen: BI 2.0 en deze plaatsen in relatie tot ontwikkelingen als Web 2.0, Web 3.0 en DW2.0.

### Wat is BI 2.0?

Een snelle internetzoektocht op de termen 'BI 2.0' en 'Business Intelligence 2.0' levert zo'n zeventigduizend treffers op, 'DW2.0' ruim elfduizend en 'Web 2.0' 308 miljoen. De term BI 2.0 is geïntroduceerd in 2006 door Charles Nicholls, schrijver van het lezenswaardige boekje 'BI 2.0: In Search of Insight'<sup>1</sup>. In navolging hiervan hebben velen zich gestort op BI 2.0: Charles Nicholls<sup>2</sup> heeft het daarbij bovenal over Operational BI (event-driven, real-time); Cubegeek heeft het over AJAX en licentiemodellen; James Taylor heeft het onder andere over "intelligent processes"; SharedInsight organiseert een congres over BI 2.0 met thema's als: enterprise search, SaaS, open source, Web 2.0, communities en SOA; ITToolbox heeft het over "process oriented" BI. Barry Devlin ziet met name ontwikkeling op SOA-gebied en geavanceerde user-interfaces; Neil Raden heeft het over "proactive, real-time, operational, integrated with business processes, and extending beyond the boundaries of the organization". Diaz Nesamoney, CEO van Celequest, schrijft: "BI 2.0 is the next generation of BI powered by Web 2.0 technologies".

Als je deze beschrijvingen en definities doorleest dan bekruipt je het gevoel dat BI 2.0 een soort vergaarbak is waarin alles kan

worden gegooid dat mist aan BI in zijn huidige vorm. Laten we deze huidige vorm BI 1.0 noemen. Nu is er over BI 1.0 nauwelijks iets te vinden op internet, de term bestaat dus eigenlijk niet.

Als ik dan zelf een gooi mag doen naar kenmerken van BI 1.0, dan ziet dat lijstje er als volgt uit:

- datagedreven rapportages en, in beperkte mate, analyse met OLAP, visualisatie, datamining en statistiek;
- op basis van het datawarehouse en/of dimensioneel gemodelleerde datamarts;
- vullen van het datawarehouse met data-integratie-tools;
- data komen met (veel) vertraging in het datawarehouse;
- beperkte interactiviteit met en tussen eindgebruikers.

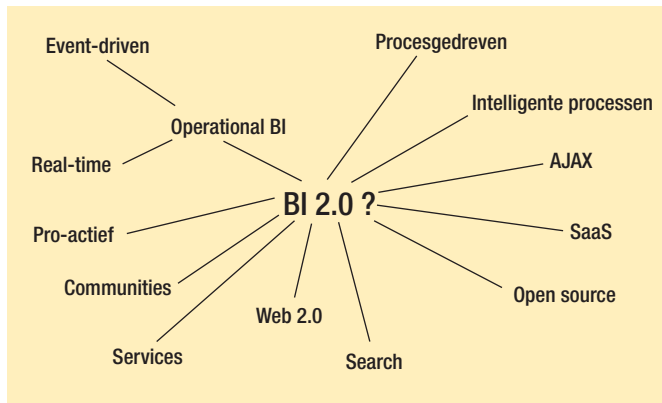
BI 1.0 kent zeker zijn toegevoegde waarde voor de business, maar het blijft beperkt in proces en techniek: data samenvoegen vanuit verschillende bronnen in een datawarehouse met daarop rapportages, applicaties zijn nauwelijks interactief, er is een sterke afhankelijkheid van IT bij het inladen en integreren van nieuwe data, toekennen van eenduidige metadata is erg lastig. BI 1.0 is daarbij sterk gebaseerd op de ideeën uit de jaren negentig van Ralph Kimball (dimensioneel modelleren/bus-architectuur) en Bill Inmon (Corporate Information Factory CIF). Er lijkt echter zoveel meer mogelijk (en noodzakelijk) op het vlak van datawarehousing en BI en dat is waarop BI 2.0 inspringt.

### BI 2.0 en Web 2.0, Web 3.0

Web 2.0 wordt gezien als een volledig platform voor interactieve webapplicaties voor eindgebruikers op het World Wide Web. Afbeelding 2 geeft een mindmap<sup>3</sup> weer van de kenmerken van Web 2.0.

Kenmerken van Web 2.0 zijn:

- "Network as platform", delivering (and allowing users to use) applications entirely through a browser;



Afbeelding 1: Een vergaarbak?

- Users owning the data on a site and exercising control over that data;
- An architecture of participation that encourages users to add value to the application as they use it. This is in sharp contrast to hierarchical access-control in applications, in which systems categorize users into roles with varying degrees of functionality;
- A rich, interactive, user-friendly interface based on Ajax or similar frameworks;
- Some social-networking aspects.

Bijzondere aandacht vragen de zogenaamde mashups<sup>4</sup>: websites en/of applicaties waarin content uit verschillende bronnen wordt samengevoegd, waardoor de websites/applicaties een nieuwe functionaliteit krijgen. Voorbeelden zijn de API's van Google Maps, Teglo, Yahoo Maps en Yahoo Pipes waarmee mashups ontwikkeld worden. Zie [www.programmableweb.com/popular](http://www.programmableweb.com/popular) voor een actueel overzicht.

Web 3.0<sup>5</sup> gaat nog een stap verder: het web wordt een database, kunstmatige intelligentie wordt een standaardapplicatie en Web 3.0 evolueert in het Semantische Web<sup>6</sup>, waarin data gevonden, geïntegreerd en geïnterpreteerd kunnen worden door software agents, omdat de semantiek van en relatie tussen de gegevens bekend zijn. Op <http://knowledge-hunter.blogspot.com/> zijn actuele ontwikkelingen over de verschillen en overeenkomsten tussen Web 2.0 en 3.0 te vinden.

De relevantie van Web 2.0 en 3.0 voor BI 2.0 is:

- Interactieve webapplicaties op basis van AJAX en aanverwante frameworks;
- De gebruiker kan data zelfstandig manipuleren;
- De gebruiker kan functionaliteit en data toevoegen (user generated content);
- De gebruiker bevindt zich binnen een netwerk van andere gebruikers;
- Folksonomy: de gebruikers kunnen data, rapportages en analyses 'taggen' met eigen termen/metadata;
- De gebruiker kan data, rapporten en analyses eenvoudig en snel terugvinden en delen met anderen.
- De betekenis van data en de tussenliggende relaties is bekend;
- BI Mashups: BI-content (gestructureerde data) wordt gecombi-

neerd met andere content (geografisch data, ongestructureerde data, audio, video etcetera).

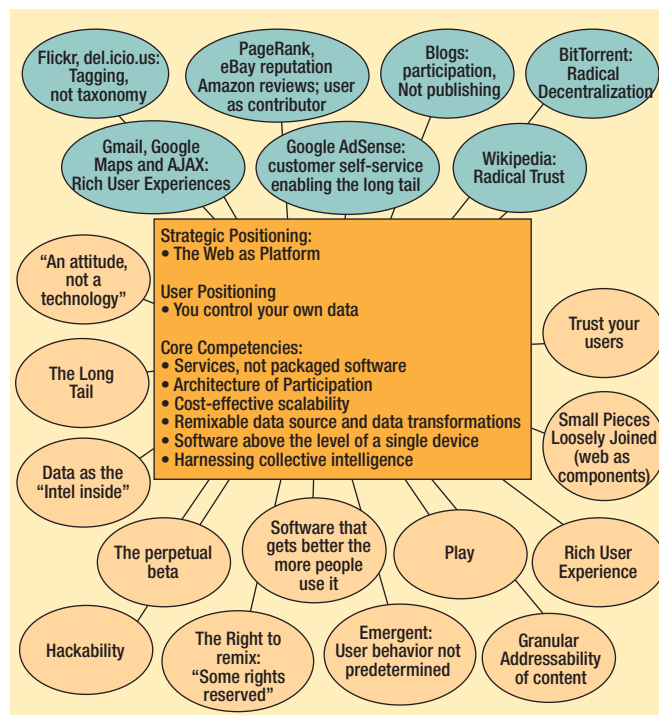
## BI 2.0 en DW2.0

Een discussie over BI 2.0 kan niet zonder een introductie van DW2.0, een term bedacht en geregistreerd door Bill Inmon, de bekende datawarehouse goeroe. DW2.0 is door Inmon geregistreerd als handelsmerk. Het is een datawarehouse-architectuur en de opvolger van de bekende Corporate Information Factory.

De kenmerken van DW2.0 volgens Inmon:

- Hold data at the lowest detail;
- Hold data to infinity (or at least to your retirement);
- Not costing huge amounts of money;
- Have integrity of data and have online high-performance transaction processing;
- Link structured data and unstructured data;
- Tightly couple metadata to the data warehouse environment;
- Support different kinds of processing without sacrificing response time;
- Support changes of data over time.

DW2.0 is een de facto datawarehouse-architectuur en heeft als zodanig niet veel met BI te maken: de architectuur zorgt voor het tijdig en volledig beschikbaar maken van relevante data. Het zegt weinig over wat er vervolgens mee gedaan wordt. DW2.0 heeft enige overeenkomsten met de definitie van BI 2.0 van Charles Nicholls: de data worden in real-time beschikbaar gesteld voor analyse, met een zo hoog mogelijk detailniveau. Daar houden de overeenkomsten ook op. Sterker nog: Charles Nicholls ziet het datawarehouse het liefst helemaal verdwijnen! BI vindt volgens



Afbeelding 2: Mindmap.

hem plaats op de events die plaatsvinden op de service bus. Deze events worden in real-time geanalyseerd in het interne geheugen, waardoor een datawarehouse overbodig wordt. De BI 2.0 definitie van Charles Nicholls en DW2.0 van Bill Inmon zijn op dit vlak dus tegenstrijdig.

De relevantie van DW2.0 voor BI 2.0 is:

- Data komen in real-time beschikbaar voor analyse;
- Data worden met een hoog detailniveau beschikbaar gesteld;
- Data worden geïntegreerd uit de bronsystemen;
- Data worden 'voor eeuwig' vastgelegd in het datawarehouse;
- Ongestructureerde data (tekstdocumenten) worden ook vastgelegd in het datawarehouse.

Zo gezien is DW2.0 een datawarehouse-architectuur die door veel organisaties omarmd zal worden, omdat het simpelweg noodzakelijk is voor een gezonde bedrijfsvoering: actuele en historische geïntegreerde gegevens vormen de basis voor Business Intelligence. Ten aanzien van de ongestuctureerde data kan nog het volgende opgemerkt worden: Inmon heeft in DW2.0 deze component formeel geïntroduceerd. De achtergrond daarvan is dat in het kader van risico management, compliance, wet- en regelgeving etcetera, inzichten verkregen moeten worden in gegevens die niet altijd in een gestructureerde database zijn opgeslagen. Denk hierbij aan e-mail, documenten, spreadsheets waarin relevante, relatief ongestructureerde informatie is opgeslagen over bijvoorbeeld consolidatie van financiële gegevens. In het kader van nieuwe wet- en regelgeving en de noodzaak voor audit trails kan het noodzakelijk zijn dat deze informatie samen met de financiële data in de database ontsloten moet worden. Inmon stelt voor om tekstanalyse toe te passen, waardoor structuren uit e-mail, documenten, spreadsheets gegenereerd worden. Deze structuren, in de vorm van kenmerkende trefwoorden, worden opgeslagen in een tabel en zijn daardoor met een querytaal als SQL te benaderen. De vraag is echter of dit type gegevens wel in het datawarehouse thuishoort: dit type informatie past toch veel beter in daartoe speciaal ontwikkelde content management systemen in combinatie met zoekmachines en tekstanalyse-software? Het vakgebied rondom zoeken en content management

stelt daarnaast heel andere eisen aan IT en de eindgebruikers. Ik krijg sterk de indruk dat Inmon ook alle relevante tekstuele gegevens in het datawarehouse wil opslaan, hoewel de huidige beschrijving van DW2.0 daar niet expliciet in is. Dit zal de complexiteit van DW2.0 alleen maar doen toenemen. De vraag is eigenlijk: moeten we dit soort gegevens in het datawarehouse opslaan? Of kiest men ervoor deze gegevens in het content management systeem op te slaan?

In de tweede variant worden met een portaal en een zoekmachine de gegevens uit het datawarehouse en content management systeem integraal ontsloten. IBM heeft overigens met UIMA<sup>7</sup> een framework ontwikkeld waarmee ongestructureerde gegevens worden gestructureerd en deze vervolgens beschikbaar gesteld in een database, waardoor deze van oorsprong ongestructureerde gegevens gecombineerd worden met de gestructureerde gegevens. Als we de voorgaande analyse van BI 2.0 en de invloeden van Web 2.0, 3.0 en DW2.0 op een rij zetten, dan komen we tot de kenmerken van BI 2.0 zoals te zien in afbeelding 3.

## BI 2.0 Tools

Na deze verhandeling over BI 2.0 is het natuurlijk zaak om eens te kijken of er al leveranciers zijn die één of meerdere van de kenmerken van BI 2.0 al hebben omgezet naar concrete producten.

*Swivel.*

Eén van de vooruitstrevendste leveranciers van BI 2.0-oplossingen is Swivel. Het heeft heel wat kenmerken van BI 2.0 geïntegreerd op de website [www.swivel.com](http://www.swivel.com). Gebruikers kunnen zelf data toevoegen en daarmee direct delen met anderen, een grote hoeveelheid grafieken maken, toelichting geven, tags toevoegen, data sets en grafieken selecteren op basis van populariteit, commentaar geven op data sets en grafieken en Swivel integreren met andere applicaties via een API. Tekstbestanden kunnen zonder problemen ingeladen worden: Swivel gebruikt tekstanalyse om de structuur aan te brengen. Wat ontbreekt bij Swivel is omgaan met operationele, real-time gegevens, dat is niet mogelijk omdat gebruikers handmatig bestanden moeten laden. Echter, met de API is automatisch laden van nieuwe gegevens weer wel mogelijk.

Kenmerk	Oorsprong	Toelichting
Interactieve webapplicaties	Web 2.0	Rich Internet Applications op basis van Ajax, Flash, Apollo, Silverlight
Datamanipulatie door gebruiker	Web 2.0	De gebruiker draagt actief bij tijdens het opbouwen en verfijnen van de data
Functionaliteit en data toevoegen	Web 2.0	De gebruiker kan, zonder tussenkomst van IT, data toevoegen en nieuwe functionaliteiten (OLAP, mining, visualisatie) zonder veel omwegen en in samenhang gebruiken
Taggen	Web 2.0	Trefwoorden/metadata toekennen aan data, rapporten, analyses, kubussen etc.
Zoeken, vinden en delen van data	Web 2.0	Zoekmachines, navigatiestructuren
Mashups	Web 2.0	Content combineren via API's
Netwerken van en interactie tussen gebruikers	Web 2.0	communities, wiki's, weblogs
Ongestructureerde data	DW 2.0	Integreren van documenten en zoekmachines via UIMA, Open Search Initiative
Data in real-time beschikbaar en procesgericht	DW 2.0, BI 2.0 (CN)	Operational BI (BI 2.0 volgens Charles Nicholls)
Historie van data	DW 2.0	Datawarehouse voor opslag van historische data
Betekenis van data en relaties (semantiek)	Web 3.0	Semantisch Web

**Afbeelding 3:** Tabel kenmerken van BI 2.0.

---

### *Many Eyes.*

Een soortgelijke oplossing als Swivel is Many Eyes, een product van IBM. Many Eyes heeft echter zeer uitgebreide visualisatiemogelijkheden, dit in tegenstelling tot Swivel. Verder kent Many Eyes veel functionele overeenkomsten met Swivel (zie <http://services.alphaworks.ibm.com/manyeyes/browse/data>).

### *BI-specialisten.*

Een hele reeks Business Intelligence leveranciers heeft zich de afgelopen jaren gespecialiseerd en zich daarmee afgezet tegen leiders als Business Objects, Cognos, SAS, Hyperion/Oracle, Information Builders en Microsoft. Bekende voorbeelden zijn BI-leveranciers als Advizor Solutions, TIBCO Spotfire, QlikTech, Visual Sciences, Anago, Panorama enzovoort. Waar de leiders zich steeds meer profileren als leveranciers die het gehele scala van tools voor BI (voornamelijk door middel van doorlopende overnames) afdekken, specialiseren de anderen zich vaak met een focus op een beperkt aantal gebieden. Deze specialisatie uit zich in het algemeen door rijkere functionaliteit en meer gebruiksgemak. Wat daarbij veelal kenmerkend is, is de hoge mate van interactiviteit en gemak waarmee rapporten, analyses en visualisaties worden uitgevoerd. Een prachtig voorbeeld hiervan is overigens Gapminder, een tool waarmee vrij beschikbare data op een eenvoudige, maar zeer krachtige wijze beschikbaar worden gesteld en via visualisatie geanalyseerd kunnen worden. Inmiddels heeft Google Gapminder via <http://tools.google.com/gapminder/> beschikbaar gesteld.

### *Business Objects Masher.*

Er zijn inmiddels diverse BI-tools waarmee mashups gemaakt kunnen worden: BI-informatie, bijvoorbeeld bestaande uit omzetcijfers of demografische gegevens, wordt geïntegreerd met een applicatie als Google Maps, waardoor een gebruiker op een geheel nieuwe en interactieve wijze informatie analyseert. Business Objects is bijvoorbeeld medio 2007 bezig met een tool genaamd Business Objects Masher, waarmee informatie in rapporten gekoppeld kan worden met geografische applicaties en waarin een zeer uitgebreide grafische engine beschikbaar is gesteld. Meer informatie is in het Lab van Business Objects (via <http://labs.businessobjects.com/mashup/>) te vinden.

### *Open Source BI.*

Een groep van BI-leveranciers waarvan bij uitstek verwacht mag worden dat hun producten menig kenmerk van BI 2.0 zal herbergen betreft de open source BI-leveranciers. Pentaho, JasperSoft, Ingres zijn voorbeelden van open source BI-leveranciers waar zich een grote groep ontwikkelaars omheen is gaan scharen, waardoor in een relatief hoog tempo producten worden doorontwikkeld. Hierdoor worden technologieën, vaak uit de Web 2.0 wereld, zoals AJAX en mashups, snel opgenomen in nieuwe releases. Een voorbeeld van een mashup is Pentaho BI Maps (zie [www.pentaho.com/products/demos/](http://www.pentaho.com/products/demos/)). Overigens is ook UIMA van IBM in een open source versie beschikbaar

(via <http://uima-framework.sourceforge.net/>). Het is een mooi voorbeeld van open source software waarmee gestructureerde en ongestructureerde data geïntegreerd aangeboden kunnen worden.

### *Operational BI.*

Operational BI, de invulling van BI 2.0 door Charles Nicholls, is inmiddels wijdverbreid geïntegreerd in BI-tools. Met name de data integratie-, middleware- en service bus-leveranciers zijn actief op dit gebied en werken nauw samen met BI-leveranciers. Zij zijn bij uitstek in staat om gegevens in (near) real-time uit bronsystemen te extraheren en beschikbaar te stellen aan de eindgebruiker, in de vorm van rapporten, analyses, operational data stores of real-time datawarehouses en datamarts.

### *Gestructureerde en ongestructureerde data.*

Een boeiende ontwikkeling van de laatste tijd is de steeds verdergaande integratie van gestructureerde en ongestructureerde data. UIMA is al genoemd. Leveranciers als Business Objects, Cognos, SAS, Informatica, IBM hebben zoekmachines, veelal die van Google, gekoppeld aan hun producten. Hierdoor wordt het mogelijk om via een zoekmachine rapporten terug te vinden, voorzien van *tags*. Interessanter wordt het als databases geïndexeerd worden en vervolgens met een zoekmachine bevraagd kunnen worden, dus zonder SQL. De zoekresultaten worden vervolgens gegroepeerd in clusters of op basis van tags gesorteerd en gerelateerd. Een stap verder nog is het afvuren van zoekvragen op de database en op een documentenverzameling, bijvoorbeeld in een content management systeem. Hierdoor worden bijvoorbeeld alle BI-rapporten van klant X getoond en worden eveneens de contracten behorende bij deze klant zichtbaar. Business Objects richt zich bijvoorbeeld momenteel sterk op dit gebied en noemt dat Enterprise Information Management.

## **BI 2.0: fixing the bugs?**

BI 2.0 openbaart zich via een nieuwe generatie BI-oplossingen, samen te stellen uit meerdere heterogene, maar samenwerkende tools. De steeds groter worden groep BI-eindgebruikers krijgt veel meer mogelijkheden om (grote hoeveelheden) pluriforme data, op een interactieve en toegankelijke wijze om te zetten naar bruikbare informatie. Met uiteindelijk als doel effectiever en efficiënter ondersteunen van beslisprocessen. Zo bezien is BI 2.0 geen bugfix, maar een volledig nieuwe release! En een nieuwe release rechtvaardigt een nieuwe naam. Ik nodig de lezers hierbij uit om op [www.dbm.nl](http://www.dbm.nl) hun suggesties te plaatsen.

### **Noten**

1. [www.seewhy.com/component?option=com\\_docman/task/doc\\_details/gid,18/Itemid,118/](http://www.seewhy.com/component?option=com_docman/task/doc_details/gid,18/Itemid,118/)
2. [www.dmreview.com/article\\_sub.cfm?articleId=1066763](http://www.dmreview.com/article_sub.cfm?articleId=1066763)
3. [www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html](http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html)
4. <http://en.wikipedia.org/wiki/Mashup>
5. [http://en.wikipedia.org/wiki/Web\\_3.0](http://en.wikipedia.org/wiki/Web_3.0)
6. [http://en.wikipedia.org/wiki/Semantic\\_web](http://en.wikipedia.org/wiki/Semantic_web)
7. [www.research.ibm.com/UIMA/](http://www.research.ibm.com/UIMA/)

**Erik Fransen** is BI consultant.