

Afstudeerscriptie over toepassing van MDM bij PepsiCo



PEPSICO

Student: Henk van Beek
ID 1578767

Student begeleider: Jos van Dongen

Hogeschool Utrecht, richting BI

Eerste examiner: Jos van Dongen
28 mei 2012

Voorwoord

Ter afsluiting van mijn studie Bedrijfskundige Informatica aan de Hogeschool Utrecht is dit rapport geschreven. Hiermee heb ik dieper in de systemen van Pepsico kunnen kijken en kennis gemaakt met projectmatig werken. Daarnaast heb ik de samenwerking met de collega's als erg aangenaam ervaren. Ik ben dan ook erg dankbaar dat Pepsico mij de gelegenheid heeft geboden om aan dit onderwerp te mogen deelnemen. Dit is zeker ook een mooie opstap naar het komende SAP systeem integratie project, waar veel MDM aspecten aan bod zullen komen.

Graag wil ik de volgende personen bedanken voor hun bijdrage aan dit rapport.

Allereerst Piet Mooij, mijn leidinggevende, opdrachtgever en bedrijfsbegeleider voor de adviezen en ondersteuning tijdens de gehele afstudeeropdracht.

Daarnaast Tjalling Dijkstra voor het toezeggen en steunen van de praktijkopdracht en Irene uit het Broek voor de deskundige begeleiding ervan.

Verder Carla Liekens, Dirk Goffin en Raj Sall, Karine Verstappen en Yvonne Postma voor hun expertise en ondersteuning bij het project.

Tenslotte Jos van Dongen, docent begeleider van de HU, voor de vlotte en deskundige begeleiding tijdens deze afstudeeropdracht.

Managementsamenvatting

Deze afstudeeropdracht is gemaakt in opdracht van de Hogeschool Utrecht (HU) ter afsluiting van de studie Bedrijfskundige Informatica (BI) en uitgevoerd bij Pepsico Nederland.

Voor Pepsico Nederland bestaat de opdracht uit het in kaart brengen van Master Data Management (MDM) voor de afdeling Data Management Organization (DMO) en het uitwerken van een praktische toepassing voor de afdeling IT.

De doelstellingen van de opdracht zijn:

- De student heeft zowel voor Pepsico als voor de Hogeschool Utrecht de afstudeeropdracht met goed resultaat afgerond.
- De afdeling DataManagement beschikt over een beschrijving van de huidige stand van MDM en wat dit kan betekenen voor de afdeling/organisatie.
- De afdeling IT beschikt over een beschrijving van de te bouwen interface, welke artikelinformatie van Oracle beverages naar PDM stuurt.

Hiervoor is het rapport verdeeld in drie secties, namelijk:

- Een literatuurstudie om te onderzoeken wat het begrip MDM inhoudt
- Welke methoden belang kunnen zijn voor Pepsico
- Een praktijkvoorbeeld van een proces van data overdracht: het ontwerpen van een interface

Een MDM programma is bedoeld ter ondersteuning van bedrijfsbehoeften door het leveren van toegang tot consistente overzichten van uniek te identificeren master data entiteiten door de gehele operationele applicatie infrastructuur. Het zorgt voor “één versie van de waarheid” met een hoge data kwaliteit.

De hoofdvraag voor deze opdracht luidt:

Wat zijn goede methoden om correcte data aan de klant aan te bieden?

Deze hoofdvraag is aan de hand van een vijftal deelvragen beantwoord, maar het komt er op neer dat de volgende methoden van data verwerking en controle van belang zijn:

- 1 Data consolidatie en integratie, het verzamelen en toegankelijk maken van data
- 2 Data synchronisatie, afstemming tussen applicaties en de master data set
- 3 Data governance
 - a Beheer van data standaarden, gestandaardiseerde definities
 - b Beheer van data kwaliteit (data integratie, matching, analyse, controle)
 - c Beheer van metadata (identificeren, kwalificeren)
- 4 Services, automatisering van voorgaande processen

Daarnaast is gekeken naar de methoden die Pepsico nu gebruikt, welke eisen gesteld worden aan data governance en welke methoden zijn het meest geschikt voor Pepsico.

In principe zijn alle methoden essentieel voor MDM binnen Pepsico.

MDM is, zeker voor zo'n grote internationale organisatie met honderden bedrijven en systemen, een belangrijk onderdeel om goed te kunnen functioneren. De twee belangrijkste zijn:

- Data governance – met daarin het beheer van data kwaliteit als speerpunt
- Omdat controles handmatig worden uitgevoerd, is het een logische vervolgstap om meer van services gebruik te maken.

Inhoudsopgave

1 Inleiding	6
2 Probleemstelling	7
3 Doelstelling	8
3.1 Opdrachtschrijving en deelopdrachten	8
4 Resultaten	11
5 Master Data Management	12
5.1 Inleiding	12
5.2 Coördinatie: belanghebbenden, eisen en planning	15
5.3 MDM componenten en het volwassenheidsmodel	15
5.4 Data modellering	17
5.5 Architectuur	17
5.6 Welke methoden van dataverwerking en controle zijn er?	18
5.6.1 Consolidatie en integratie	18
5.6.2 Synchronisatie	19
5.6.3 Data governance	20
5.6.4 Data standaarden	21
5.6.5 Data kwaliteit	21
5.6.6 Metadata management	22
5.6.7 Identificeren van masterdata en metadata	23
5.7 Functionele servicelaag	24
5.8 Management richtlijnen	25
5.9 Deelconclusie	26
6 Welke delen van MDM zijn voor Pepsico interessant?	27
6.1 Inleiding	27
6.2 Welke methoden gebruikt Pepsico nu?	27
6.3 Mate van volwassenheid	28
6.4 Data governance	29
6.5 Data kwaliteit	30
6.6 Business Process Management	32
6.7 Deelconclusie	32
7 Praktijkproject MDM, het ontwerpen van een data interface	34
7.1 Inleiding	34
7.2 Projectafbakening	35
7.3 Het ontwerp	36
7.4 Bouwen en testen	37
7.5 Implementatie en nazorg	37
7.6 Support model	38
7.7 MDM aspecten	38
7.8 Deelconclusie	39
8 Conclusies en aanbevelingen	40
9 Evaluatie van het proces	42

Bronvermeldingen	43
Bijlagen:	44
1 Plan van Aanpak	44
2 Evaluatie van het eigen functioneren	59
3 MDM Theorie	60
4 MDM Maturity Matrix	94
5 Lijst met afkortingen en begrippen	95
6 Informatie theorie	96
7 Mapping Oracle to PDM	97
8 Package type and material code	101
9 Bepaling PDM veldwaarden voor “Brand, Subbrand, Functional name, Product group and GPC code”	102
10 Specificaties en eisen van de BODS interface	103
11 Functioneel ontwerp van de BODS interface	104
12 Incident and change management	106

1 Inleiding

Deze afstudeeropdracht is gemaakt in opdracht van de Hogeschool Utrecht (HU) ter afsluiting van de studie Bedrijfskundige Informatica (BI) en uitgevoerd bij Pepsico Nederland.

Voor Pepsico Nederland bestaat de opdracht uit het in kaart brengen van Master Data Management (MDM) voor de afdeling Data Management Organization en het uitwerken van een praktische toepassing voor de afdeling IT.

Het doel is, om het fenomeen MDM te onderzoeken en mogelijk bruikbare methoden te destilleren voor intern gebruik.

Deze scriptie is als volgt opgebouwd:

Eerst wordt de probleemstelling geformuleerd en vervolgens de doelstellingen met een hoofdvraag en deelvragen. Daarna worden de resultaten beschreven, opgesplitst in drie secties.

De theorie wordt beschreven, vervolgens wordt de theorie vertaald naar de situatie bij Pepsico en tenslotte in praktijk gebracht door het ontwerpen van een data interface.

Het rapport wordt afgesloten met de conclusies en aanbevelingen.

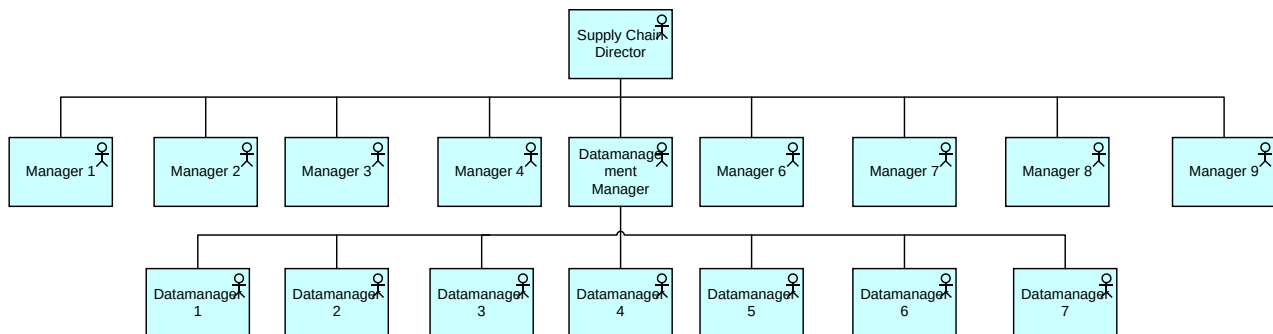
In dit rapport worden veel specifieke termen en afkortingen gebruikt. In bijlage 5 is een overzicht daarvan opgenomen.

Beknorte context waarbinnen het afstudeerproject wordt uitgevoerd

PepsiCo is één van de grootste voedingsmiddelenbedrijven ter wereld met merken in hartige tussendoortjes, ontbijtgranen en dranken en met 285.000 medewerkers met activiteiten in meer dan 200 landen. PepsiCo Nederland maakt onderdeel uit van PepsiCo Europa.

De afdeling DataManagement Organization (DMO) is onderdeel van de supply chain afdeling, welke de gegevens van alle productie eenheden in Nederland en België verzamelt en beheert.

Deze gegevens betreffen informatie over leveranciers, materialen, productie, klanten en prijzen en worden in verschillende systemen opgeslagen en bijgehouden.



Figuur 1 Organogram supply chain afdeling

Positie student binnen de organisatie

Ik werk als één van de zeven Datamanagers binnen de afdeling DMO.

Deze functie omvat het verwerken en beheren van alle supply chain data in de verschillende systemen, zoals SAP en Oracle. Mijn bedrijfsbegeleider Piet Mooij is Data Management manager.

2 Probleemstelling

Vanwege de recente nadruk op naleving van de regelgeving, maar ook door fusies en overnames, is het voor Pepsico een noodzaak om accurate en complete Master Data te creëren en te onderhouden.

Omdat er nog diverse legacy systemen op diverse locaties actief zijn, met vaak dezelfde gegevens, is het risico aanwezig, dat die gegevens niet consistent zijn. Er moeten hiervoor overkoepelende data-onderhoud en governance processen en procedures ontwikkeld worden om “wildgroei” te voorkomen.

3 Doelstelling(en) van de opdracht

Na afloop van het afstudeerproject moet het volgende zijn bereikt:

- De student heeft zowel voor Pepsico als voor de Hogeschool Utrecht de afstudeeropdracht met goed resultaat afgerond.
- De afdeling DataManagement beschikt over een beschrijving van de huidige stand van MDM en wat dit kan betekenen voor de afdeling/organisatie. Ook welke onderdelen kunnen worden gebruikt en hoe die kunnen worden toegepast.
- De afdeling IT beschikt over een beschrijving van de te bouwen interface, welke artikelinformatie van Oracle beverages naar PDM stuurt.

3.1 Opdrachtomschrijving met deelopdrachten

Een middel om goede controle over Master Data te hebben is Master Data Management (MDM). Een definitie van Master Data Management (MDM) is: de benodigde technologie, tools en processen om consistente en accurate master data te creëren en te onderhouden.

MDM is niet alleen een technologisch probleem. In veel gevallen zijn fundamentele veranderingen van bedrijfsprocessen nodig om zuivere master data te behouden en vaak zijn de lastigste MDM problemen meer van politieke aard dan technisch.

Met deze afstudeeropdracht ga ik onderzoeken, op welke manier MDM van toepassing kan zijn voor twee afdelingen binnen Pepsico: DataManagement Organization (DMO) en IT.

DMO

Voor DMO gaat het om het binnenhalen van kennis van MDM om te inventariseren welke toepassingen er in de toekomst van belang kunnen zijn.

Hiervoor moet eerst kennis en inzicht worden verkregen van MDM: wat is het, wat houdt het in, wat zijn de huidige best practices? Dit is deelopdracht 1 (hoofdstuk 5).

Vervolgens wordt het huidige volwassenheidsniveau van Pepsico Nederland bepaald op het gebied van Data Governance. Waar staan we nu op het gebied van Data standaarden, Metadata management, Data kwaliteit en Data validatie, de vier deelgebieden van Data Governance.

Tenslotte wordt gekeken naar welk volwassenheidsniveau Pepsico wil groeien en op basis daarvan bekeken wat er moet gebeuren om dat niveau te behalen. Dit is deelopdracht 2 (hoofdstuk 6).

IT

Ten behoeve van de informatievoorziening naar hun klanten, wil Pepsico productgegevens van de Business Unit “Dranken” aan hen verstrekken.

Deze productinformatie staat in het systeem Oracle, maar is nog niet beschikbaar voor klanten.

De bedoeling is, om die informatie automatisch door te sturen naar een zogenaamde datapool, waarin deze artikelgegevens centraal beschikbaar zijn en beheerd worden.

Het systeem dat Pepsico daarvoor gebruikt heet PDM, Product Data Management, zie bijlage 1 van bijlage 1, het Plan van aanpak.

De gegevens in PDM worden binnen Pepsico beheerd en worden elektronisch naar de externe datapool (1Sync) gestuurd.

Voor deze afstudeeropdracht gaat het om het beschrijven van een ontwerp om MDM te gebruiken bij het interfaceren van data vanuit Oracle naar PDM met behulp van BO Data Services, zie bijlage 2 van bijlage 1, het Plan van aanpak. Dit is deelopdracht 3 (hoofdstuk 7).

Van het IT project ben ik zelf de projectleider en heb ik het project gestuurd aan de hand van PLM, Project Lifecycle Management, zie bijlage 3 van bijlage 1, het Plan van aanpak.

In de IT wereld zijn er vele Project Management methoden bekend, waarbij Prince2 en PLM de bekendste zijn. Prince2 is van Engelse origine en wordt veel in West-Europa toegepast. PLM is de Amerikaanse variant en wordt veel in Noord-Amerika gebruikt. Pepsico hanteert als Amerikaans bedrijf binnen de organisatie de PLM methode en daarom zal ik deze ook toepassen op dit project.

Hoofdvraag

De hoofdvraag luidt als volgt:

Wat zijn goede methoden om correcte data aan de klant aan te bieden?

MDM is een veelomvattend begrip, maar welke MDM methoden dragen bij aan juiste gegevens?

Afgeleide vragen hierbij zijn:

1. Welke methoden van databewerking en controle bestaan er?
2. Welke methoden gebruikt Pepsico nu?
3. Welke eisen worden gesteld aan Data Governance?
4. Welke methoden zijn het meest geschikt voor Pepsico en waarom?
5. Hoe weet ik dat de aangeboden data juist zijn?

Activiteiten

Om deze vragen te beantwoorden, zijn de volgende acties nodig:

1. Onderzoeken wat MDM inhoudt en welke onderliggende methoden gebruikt worden.
2. Onderzoeken welke MDM methoden momenteel binnen de afdeling DMO gebruikt worden.
3. Onderzoeken welke eisen worden gesteld aan Data Governance.
4. Een voorstel maken van welke methoden bij Pepsico toe te passen zijn en hoe.
5. Onderzoeken, aan de hand van dit project, hoe MDM methoden tot de juiste data kunnen leiden.

Eindproduct en deelproducten

Het eindproduct is een rapport dat bestaat uit twee delen:

DMO

Een beschrijving van de huidige stand van MDM en welke onderliggende methoden bruikbaar zijn voor Pepsico en in het bijzonder de afdeling DMO. Ook wordt ingegaan op hoe die toegepast kunnen worden.

Dit omvat een:

1. Inventarisatie van MDM
2. Inventarisatie van momenteel gebruikte methoden bij de afdeling DMO
3. Eisenpakket van Pepsico
4. Voorstel van mogelijk toe te passen methoden

IT

Een beschrijving van de te bouwen interface, welke artikelinformatie van Oracle beverages naar PDM stuurt. Daarbij wordt onderzocht, welke MDM toepassingen kunnen bijdragen aan juiste data (deelproduct 5).

Methoden en technieken

Voor het DMO gedeelte wordt een literatuurstudie uitgevoerd en van internetsites actuele informatie verzameld, zoals white papers en artikelen, zie “bronvermeldingen”.

Voor het IT gedeelte wordt Business Objects Data Services gebruikt van SAP, zie bijlage 2 van bijlage 1, het Plan van aanpak. Dit is een conversie / interface tool, welke door Pepsico gebruikt wordt.

Voor project management wordt PLM gebruikt, zie bijlage 3 van bijlage 1, het Plan van aanpak.

4 Resultaten

De opdracht bestaat uit drie onderdelen (zie § 3.1), te weten:

- A. Beschrijf wat MDM inhoudt
- B. Welke onderdelen hieruit kunnen van belang zijn voor Pepsico
- C. Ontwerp een interface voor data overdracht

Elk onderdeel bevat een of meerdere deelvragen, welke in de volgende hoofdstukken aan bod komen:

- A Beschrijf wat MDM inhoudt, hoofdstuk 5
 - 1 Welke methoden van databewerking en controle bestaan er?
- B Welke onderdelen hieruit kunnen van belang zijn voor Pepsico?, hoofdstuk 6
 - 2 Welk methoden gebruikt Pepsico nu?
 - 3 Welke eisen worden gesteld aan Data Governance?
 - 4 Welke methoden zijn het meest geschikt voor Pepsico en waarom?
- C Ontwerp een interface voor data overdracht, hoofdstuk 7
 - 5 Hoe weet ik dat de aangeboden data juist zijn?

5 Master Data Management, MDM

In dit hoofdstuk wordt de eerste deelvraag behandeld vanuit de theoretische achtergrond van Master Data Management, kortweg MDM.

Die eerste deelvraag is: **Welke methoden van dataverwerking en controle zijn er?**

5.1 Inleiding

In dit hoofdstuk wordt er eerst ingegaan op het begrip Master Data en Master Data Management. Omdat MDM niet alleen een bedrijfsbreed programma is, maar ook veel omvattend wat betreft reikwijdte in de organisatie, betrokken personen, processen, applicaties en gegevens(bewerking), wordt er ook toelichting gegeven op:

- wie de betrokkenen zijn (§ 5.2)
- uit welke componenten een MDM programma bestaat om zo een goed overzicht te krijgen (§ 5.3)
- een volwassenheidsmodel, om te bepalen hoe volwassen of hoe ver een organisatie is op het gebied van MDM (§ 5.3)
- data modellen en architectuur (§ 5.4 en 5.5)
- methoden van dataverwerking en controle, als antwoord op de deelvraag (§ 5.6)
- service en een management richtlijn bij het ontwikkelen van een MDM programma (§ 5.7 en 5.8)

In bijlage 3 staat een uitgebreide versie van de beschrijving van MDM. Indien nadere informatie gewenst is, kan deze geraadpleegd worden.

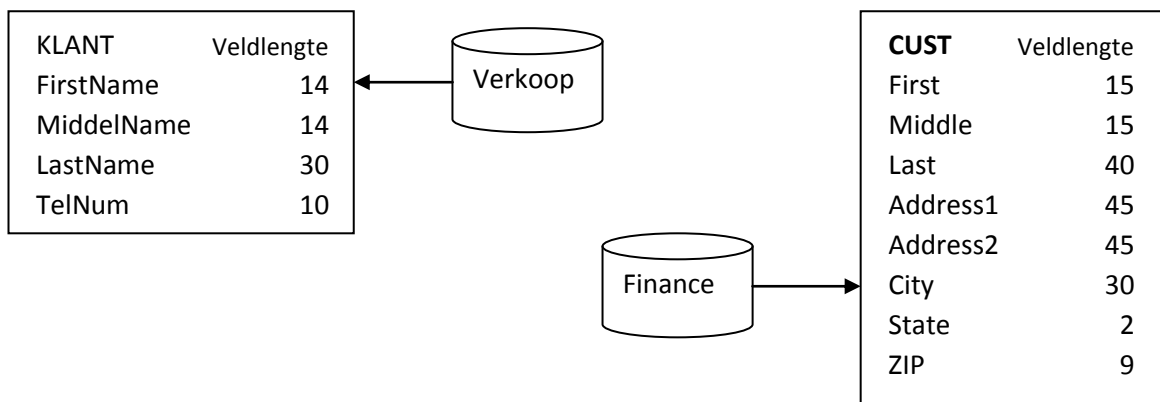
In dit hoofdstuk wordt er voor het begrip “gegevens” steeds de term “data” gehanteerd, aangezien (master) data management een internationaal gebruikte term is en voor vele bedrijven als een algemeen bekende term wordt gehanteerd.

Alvorens ingegaan wordt op voorgaande punten, wordt in deze paragraaf eerst uitgelegd wat master data en master data management is.

Flexibele organisaties die zich kunnen aanpassen aan een vloed van redundante en soms conflicterende gegevens, merken dat de combinatie van informatie uitwisseling en operationele samenwerking een onderscheidende factor is voor organisatorisch succes.

Aangezien partnerbedrijven meer informatie delen en uitwisselen, neemt verbondenheid ook toe. Het zijn niet alleen systemen die beter samenwerken, maar ook mensen die de systemen beheren vormen betere werkrelaties, hetgeen leidt tot effectievere bedrijfsleiding en uiteindelijk tot concurrentievoordeel.

In het verleden werden applicaties ontwikkeld om in operationele behoeften te voorzien op een specifiek gebied. De applicatie architectuur bood ondersteuning in die deelgebieden. Dat gold ook voor de afgeleide IT ondersteuning met data definities, tabellen, structuren, functionaliteiten etc.. Verschillende applicaties kunnen echter refereren naar meerdere datasets met hetzelfde bedrijfsconcept, maar dezelfde namen kunnen ook refereren naar verschillende concepten, zie figuur 1. Om tot een organisatiebrede gegevensuitwisseling te komen, zal er eerst gewerkt moeten worden aan een consistent beeld van alle data uit de verschillende databronnen.



Figuur 1 Verschillende applicaties kunnen refereren naar dezelfde bedrijfsconcepten

Om dit te realiseren moet een organisatie in staat zijn om de bedrijfsconcepten helder te definiëren, de verschillende manieren te identificeren waarop data sets een bedrijfsconcept vertegenwoordigen, de data integreren in een consistent beeld en vervolgens het consistente beeld beschikbaar maken binnen de gehele organisatie.

Dit is een enorme uitdaging voor integratie van bedrijfsinformatie, management en uitwisseling om een echt geïntegreerde organisatie te vormen.

De essentie van de MDM uitdaging is om een bedrijfsbreed overzicht te organiseren van de belangrijkste informatieobjecten en om de kwaliteit, het gebruik en synchronisatie te waarborgen, daarmee het gebruik van informatie te optimaliseren en zo de operationele en strategische doelstellingen van de organisatie te bereiken.

Wat wordt verstaan onder Master data?

Het verkrijgen van “één versie van de waarheid” en het beheren van de kwaliteit van gegevens, is een probleem voor bijna elke organisatie.

Master data zijn die entiteiten, relaties en attributen, welke kritisch zijn voor een onderneming en fundamenteel voor de belangrijkste bedrijfsprocessen en applicaties.

Master data wordt gedefinieerd door master data objecten.

Master data objecten zijn de kern bedrijfsobjecten, welke in verschillende applicaties in de organisatie gebruikt worden met de bijbehorende metadata, attributen, definities, rollen, verbindingen en taxonomieën (manier waarop bedrijfsconcepten worden beschreven, gekarakteriseerd en geclassificeerd, zie ook bijlage 5). In figuur 2 staat een voorbeeld van master data objecten.

Master data objecten zijn de sleutel objecten die er het meest toe doen, de zaken die worden vastgelegd in transactionele systemen, gemeten en gerapporteerd in rapportagesystemen en geanalyseerd in analytische systemen. Het gaat dus om klanten, medewerkers, leveranciers, onderdelen, producten, locaties, contracten, etc.

Master Data Object	Waarde
Klant	Jumbo
Product	Cruesli 12x500g
Locatie	Utrecht
Hoogte doos	200 mm

Figuur 2 Master Data Elementen

Een master data systeem, dat bestaat uit een master data set, is een (potentieel virtuele) opslag of index van uniek geïdentificeerde entiteiten met hun kritische data attributen, gesynchroniseerd vanuit de originele databronnen en beschikbaar gemaakt voor bedrijfsbreed gebruik.

Met geschikte controle en overzicht, kunnen de data in het master data systeem (of opslag) gekwalificeerd worden als een eendrachtig en coherente databron, waar alle applicaties op kunnen vertrouwen voor consistente informatie van hoge kwaliteit.

Wat is MDM?

MDM is een verzameling van methoden, welke belanghebbenden, deelnemers en interne klanten leiden in het opnemen van bedrijfsapplicaties, informatie management methoden en Data Management tools om het beleid, procedures, service en infrastructuur uit te voeren ter ondersteuning van het laden, integreren en gedeeld gebruik van accurate, tijdige, consistente en complete master data.

Met andere woorden, een MDM programma is bedoeld ter ondersteuning van bedrijfsbehoeften door het leveren van toegang tot consistente overzichten van uniek te identificeren master data entiteiten door de gehele operationele applicatie infrastructuur.
MDM moet beschouwd worden als een “programma”, niet als een project of een applicatie.

Data kwaliteit

Het succes van MDM hangt af van de hoge kwaliteit van data. Data kwaliteit kan verbeterd worden door MDM.

Data kwaliteit is kritisch voor het succes van MDM en het is van invloed op bijna alle aspecten van het MDM programma.

Technologie

MDM is geen technologisch project, maar het ontwikkelen van een MDM programma is onmogelijk zonder tools en technologie. Het gaat in deze bijlage dus niet om de tools, maar meer hoe de tools worden gebruikt om een geconsolideerd overzicht van uniek te identificeren master dat objecten te krijgen.

5.2 Coördinatie: belanghebbenden, eisen en planning

Het beheren van de master data is een kritische component van de bedrijfsinformatie management strategie. Dit vergt veel energie met betrekking tot coördinatie: het identificeren van belanghebbenden, hun steun krijgen, onderlinge samenwerking bewerkstelligen, eisen opstellen en rollen en verantwoordelijkheden aan de juiste mensen toekennen.

Alvorens een MDM wordt gebouwd en geconfigureerd, moeten er door de MDM teamleden een aantal taken uitgevoerd worden:

- Identificeer de personen in de organisatie die profijt hebben van MDM
- Omschrijf de toegevoegde waarde van MDM en ontwikkel de rechtvaardiging ervan
- Verzamel de data eisen en ken daaraan prioriteiten toe, welke het ontwerp en de ontwikkeling van de onderliggende infrastructuur bepalen
- Ontwerp een plan voor bedrijfsbrede data integratie
- Ontwerp een migratieplan voor de betreffende applicaties

Wie zijn de belanghebbenden?

De belanghebbenden zijn o.a. het hoger management, interne gebruikers, eigenaars van applicaties, informatie architecten en systeemontwikkelaars. Wat hun rol en bijdrage is aan de ontwikkeling van het MDM programma, staat beschreven in bijlage 3.

Ontwikkel een project handvest

Om een duidelijk beeld te schetsen en voor acceptatie van het MDM programma door het hoger management, moet het MDM team een handvest opstellen.

In een project handvest staat, wat er bereikt gaat worden met het MDM programma en hoe de diverse mijlpalen bereikt worden.

Het project handvest beschrijft de business case, het doel en de waarde van MDM en de belangrijkste personen in het project. Verder staan er de doelstellingen in, de scope, de acceptatie- en succescriteria en potentiële risico's.

5.3 MDM componenten en het volwassenheidsmodel

Voordat aan een MDM programma begonnen wordt, is het zaak om eerst de essentiële bekwaamheden te begrijpen om MDM succesvol uit te rollen en de volwassenheid van die bekwaamheden te bepalen om MDM uitvoerbaar te maken.

Er is geen enkele lijst die alle benodigde services bevat, welke een bedrijf nodig heeft voor zijn master data set, maar het is waardevol om high level de MDM bekwaamheden te inventariseren.

In deze paragraaf worden een aantal technische MDM componenten besproken en wordt het volwassenheidsniveau van elk van die componenten bekeken. Aan de hand van de huidige situatie en vervolgens de gewenste situatie van een onderneming, kan een MDM programma manager een implementatie roadmap ontwerpen met de te nemen stappen in het MDM programma.

Om een MDM oplossing in te zetten, moet de onderneming voorbereid zijn op de technische, operationele en management uitdagingen, waarbij de ontwikkeling in meerdere iteraties zal verlopen. Een volwassen MDM oplossing zal de bekwaamheden en services, benoemd in figuur 3, bevatten.

Architectuur	Service laag (controle)			
	Masterdata model (structuur)		Systeemarchitectuur (kracht)	
Governance				
	Data standaarden	Metadata management	Data kwaliteit	Databeheer
Management	Migratieplan			
	Hierarchie		Identiteit	
	Administratie en configuratie			
Identificatie	Zoek en bepaal unieke omschrijving			
	Koppeling records		Samenvoegen en versterken/completeren	
Integratie	Applicatie integratie en synchronisatie laag			
	MDM component service laag			
Business Process Management	Business Process Integratie			
	Business rules			
	MDM business component laag			

Figuur 3 MDM componenten en service model (service lagen)

In bijlage 3 wordt nader ingegaan op deze componenten.

Het model

Het volwassenheidsniveau beschrijft hoe reeds ingezette componenten en services kunnen worden benut ten behoeve van een MDM programma en geeft aan welke ontbrekende bekwaamheden moet worden verkregen, zodat een applicatie op een betrouwbare manier gebruik kan maken van master data. Er zijn een vijftal volwassenheidsniveaus te onderscheiden:

- 1 Initieel
- 2 Reactief
- 3 Beheerst
- 4 Proactief
- 5 Strategisch

In bijlage 4 staat een voorbeeld van een Maturity Matrix, waarin een onderneming het huidige en het gewenste volwassenheidsniveau kan aangeven.

Ontwikkel en implementeer een roadmap

Hoewel elk bedrijf uiteindelijk een strategische volwassenheid wil bereiken, is het raadzaam om met behulp van een roadmap door de verschillende volwassenheidsniveaus te lopen, om vervolgens de verwachte waarde voor het bedrijf bij elk bereikt niveau te beschrijven.

Het doel van het toepassen van een roadmap is om de bedrijfsdoelen helder te krijgen, weten welke bekwaamheden er nodig zijn om die doelen te bereiken, bepalen welke bekwaamheden al in de organisatie aanwezig zijn, welke kloven zijn er, bepaal waar de organisatie moet zijn en beslis hoe daar te komen.

Het volwassenheidsmodel wordt gebruikt als de maatstaf waarmee de organisatie zijn MDM bekwaamheid en gereedheid worden gemeten. Het wijst een eindstadium aan, welke de bedrijfsanalisten helpt de uitvoering van het plan vorm te geven naar de huidige applicatie architectuur en bekijk de beste manier om naar het eindstadium te komen.

5.4 Data modellering

Hierbij wordt gekeken naar aspecten met betrekking tot het ontwikkelen van modellen voor master data, van extractie, consolidatie, vastleggen en verdelen.

Om elke master entiteit te kunnen onderscheiden moeten hun identiteit attributen gemodelleerd worden.

Identificatie attributen zijn die data elementen met hun waarden die gebruikt worden om één entiteit voorval te onderscheiden van een andere entiteit voorval.

Dat kan bijvoorbeeld door het toevoegen van een datum (timestamp) of een ordernummer bij een overeenkomst.

Onafhankelijk van welke architectuurbenadering wordt toegepast, de naamgeving is een set data attributen voor het formuleren van een “gesleutelde” master index, welke gebruikt kan worden om unieke attributen te vormen. De master index bevat een aantal gebruikelijke data attributen. De minimale set van attributen is de set van identificerende attributen.

Informatie delen en uitwisselen

Hierbij wordt gekeken naar de conceptuele aspecten van het modelleren van extractie en uitwisselingsmodel, als ook beheersbare gecentraliseerde opslag, welke gebruikt kan worden voor consolidatie, identiteitbeheer, publicatie en toegangsbeheer.

5.5 Architectuur

In deze paragraaf wordt gekeken naar master data gebruikscenario's en architectuurvormen.

Deze beschrijven de criteria die gebruikt kunnen worden om te bepalen welk architectuur raamwerk het best bij de onderneming past.

Er zijn drie basis architecturen voor het beheren van master data.

- Register, een virtuele master data index
- Transactie Hub, een volledig geconsolideerde master data set voor gesynchroniseerde toegang
- Hybride / gecentraliseerde master, een combinatie van voorgaande benaderingen

Een gedetailleerde uitleg hiervan is te vinden in bijlage 3.

Om een goede selectie te kunnen maken, welke benadering het meest geschikt is voor een onderneming, zijn er een zestal variabelen van belang:

- Aantal attributen
- Mate van consolidatie
- Mate van synchronisatie
- Toegang
- Service complexiteit
- Prestaties

Grofweg kan gehanteerd worden dat hoe groter het aantal attributen beheerd moet worden en hoe hoger de mate van consolidatie (dus hoe strakker de koppeling), hoe meer de architectuur wijst richting de transactie hub. Het tegenovergesteld geldt voor het register, terwijl de hybride er tussenin zit.

5.6 Welke methoden van dataverwerking en controle zijn er?

In deze paragraaf wordt de eerste deelvraag behandeld.

De methoden van dataverwerking in het MDM programma zijn consolidatie, integratie en synchronisatie. De controlemethoden die daarop uitgevoerd worden, vallen onder data governance.

5.6.1 Consolidatie en integratie

Zodra het master data model helder is, moet bekeken worden hoe de gegevens van de geïdentificeerde bron naar een geconsolideerde master omgeving worden gebracht. Het gaat daarbij om de initiële migratie van data vanuit de bron en om de “lopende” integratie van data voorvallen die worden gecreëerd, veranderd of verwijderd van de master omgeving.

Data integratie omvat de processen voor het verzamelen van data uit verschillende bronnen en deze data toegankelijk maken voor verschillende applicaties.

Informatie delen

De benadering voor het delen van informatie is afhankelijk van de bedrijfseisen aan applicaties. Informatie wordt gedeeld op drie verschillende manieren met behulp van data integratie tools:

1. Extractie en consolidatie, met data integratie producten:
 - Transformatie, op basis van regels naar een juist formaat
 - Monitoring, conformiteit aan bedrijfseisen, validatie
 - Consolidatie, m.b.v. data kwaliteit technologieën
2. Data verenigen, door het gebruik van een register met een index in de master database, kunnen gegevens bij de bron blijven en op verzoek getoond worden.
3. Data verspreiden, vanuit de master terug naar de bronnen.

Informatie identificeren

Onafhankelijk van welke architectuur wordt gekozen voor het opbouwen van een MDM systeem, er moet altijd een identificatie service zijn die wordt opgevraagd, telkens als een applicatie een master record wil benaderen.

De identificatie service wordt ondersteund door het gebruik van een register van entiteiten die geïndexeerd is met behulp van de waarden in de geïdentificeerde attributen.

Consolidatie technieken

Identiteit herkenning gaat over de bepaling dat twee of meer data representaties samengebracht kunnen worden tot één representatie van een uniek object.

Het gaat daarbij vaak over namen en adressen, maar ook veel over productnamen, productcodes, objectomschrijvingen, etc.

Dit proces omvat twee fases:

- Ontdekking, deze combineert data profilering activiteiten met manuele beoordeling van data.
- Overeenkomsten bekijken en matchen voor het koppelen van records.

De volgende data kwaliteit tool functies zijn van belang voor het master data integratie proces:

1. Ontleding en standaardisatie
2. Data transformatie
3. Matchen en koppelen

Classificatie

Master data objecten kunnen vaak geclassificeerd worden in een logische hiërarchie of taxonomie. Classificatie heeft te maken met de applicatie eisen en hoe de entiteiten gebruikt worden. Het verbetert het consolidatieproces door de "zoekruimte" te reduceren bij het lokaliseren van overeenkomsten voor een specifieke zoekopdracht.

Consolidatie

Dit is het resultaat van de toegepaste taken van data integratie.

- Identifierende waarden zijn ontleed, gestandaardiseerd en genormaliseerd over alle data domeinen
- Deze zijn onderworpen aan classificatie en blokkeringschema's
- Deze zijn voorgelegd aan de unieke identificatie service voor het analyseren van duplicaten, kijken naar hiërarchische groepering, lokaliseren van een bestaand record of bepalen dat die niet bestaat.

5.6.2 Synchronisatie

De beschikbaarheid van master data hangt af van de mate van synchronisatie tussen applicaties en de master data set, evenals tussen de componenten van de master data set zelf. Hierbij zijn specifieke aspecten, de synchronisatie dimensies, van belang.

Synchronisatie dimensies:

- Tijdigheid
- Actualiteit
- Vertraging
- Consistentie (overal hetzelfde)
- Coherentie (samenhangend)
- Reproduceerbaarheid

Elk van de karakteristieken beïnvloed het onderhoud van een consistente weergave van master data door de applicatie infrastructuur en elk van de verschillende MDM architecturen zullen voldoen aan de eisen binnen die verschillende dimensies op verschillende manieren.

Conceptuele modellen voor het delen van data

In paragraaf 5.5 zijn de drie verschillende MDM architecturen, register, transactie hub en hybride, reeds besproken.

Als we nu de synchronisatie dimensies bekijken van elke architectuur, wordt zichtbaar op welke vlakken de architecturen zich van elkaar onderscheiden.

	Register	Transactie hub	Hybride
Tijdigheid	+	+	?
Actualiteit	+	+	?
Consistentie	-	+	?
Coherentie	-	+	?
Reproduceerbaarheid	-	+	?

Tabel 1 Synchronisatie karakteristieken van verschillende modellen

Bij de hybride architectuur staan alleen vraagtekens, omdat deze tussenvorm van gegevensuitwisseling meerdere kopieën bevat van unieke gegevens. Hierbij sturen de applicaties de volgorde van transacties en daarnaast doorkruisen de gegevens meerdere onderliggende database platformen , hetgeen de complexiteit van recordblokkering verhoogt. Hierdoor is het moeilijk om de niveaus van de synchronisatie dimensies te bepalen.

5.6.3 Data governance

Data governance is het beheren en bewaken van de risico’s die zich voordoen in een bedrijfsinformatie portfolio en het reduceren van impact als gevolg van het ontbreken van overzicht. Van data governance wordt verwacht dat de data voldoen aan alle bedrijfsdoelen in de context van databeheer, eigenaarschap, compliance (nakomen van het bedrijfsbeleid), privacy, veiligheid, data risico’s, gevoeligheid van data, metadata management en MDM. Elk van deze problemen moeten worden geïntegreerd met controle voor management overzicht samen met verificatie van organisatorisch toezicht en informatiebeleid. Als een bedrijf meerdere informatiesystemen heeft, is het verstandig dat de directie / het management uitspraken doet over de wijze waarop daarmee omgegaan moet worden. Dit noemen we informatiebeleid. Met andere woorden, elk aspect van data governance is gerelateerd aan de specificatie van informatiebeleid dat de bedrijfsbehoefte en verwachtingen reflecteert, samen met de processen voor het monitoren van conformiteit naar dat informatiebeleid.

Afstemming informatie doelstellingen met de bedrijfsstrategie

Elke organisatie heeft een bedrijfsstrategie dat de bedrijfsdoelstellingen reflecteert, evenals de risico management doelstellingen en de compliance management doelstellingen. Uiteindelijk hangt het succes van de onderneming af van het beheer over hoe alle processen conform de bedrijfsstrategie verlopen. De uitdaging ligt in de bekwaamheid van het management om de bedrijfsstrategie effectief te communiceren.

Data governance en data kwaliteit

Een data kwaliteit en data governance assessment verklaart hoe de informatiearchitectuur wordt gebruikt ter ondersteuning van compliance met gedefinieerde informatiebeleid. Het geeft aan, dat data kwaliteit en data standaard management een onderdeel zijn van een veel groter geheel met betrekking tot overzicht van bedrijfsinformatie.

Voor dit overzicht is nodig:

- Beoordeling van data kwaliteit dimensies (zie § 5.6.5) in relatie tot de data elementen
- Bepalen hoe kritisch deze zijn voor bedrijfsprocessen
- Gegevensregels bepalen, welke compliance beïnvloeden
- Kwantitatieve metingen definiëren voor conformiteit aan informatiebeleid
- Manieren bepalen om het voorgaande te integreren in een data governance raamwerk

Risico's

Een belangrijke peiler voor de noodzaak van data governance is risico, zowel bedrijfsrisico als compliance risico. Het is goed om te kijken naar enkele risicogebieden in relatie tot master data, welke informatie management vereisen.

Risicogebieden

Enkele risicogebieden zijn:

- Bedrijf en financieel
bijv. onvoldoende controle op financiële rapporten heeft negatief effect op het bedrijf en kan leiden tot waardevermindering, onvoldoende overzicht over betalingen kan leiden tot boekhoudkundige afwijkingen.
- Rapportage
t.b.v. voorschriften zoals SOx, onjuiste data kunnen leiden tot waardevermindering van het bedrijf, rechtszaken, etc.
- Kennis van entiteiten
Kennis van partijen waar de onderneming zaken mee doet is kritisch voor reductie van risico's (kunnen zij hun beloftes nakomen?)
- Bescherming, privacy en beveiligde informatie
- Beperking van gebruik, toegangsrechten

Data governance is onder te verdelen in de volgende gebieden:

- Beheer van data standaarden
- Beheer van data kwaliteit (data integratie, matching, analyse, controle)
- Beheer van metadata (identificeren, kwalificeren)

5.6.4 Data standaarden

Als sommige data elementen met dezelfde naam niet dezelfde betekenis hebben, of met verschillende namen welke juist wel behoren tot hetzelfde concept, dan vormt dat een gevaar voor het integratie- en consolidatieproces. Gestandaardiseerde definities moeten de naamgeving van de master data object catalogus sturen en controleren en hoe die tot uniek overzicht komen.

5.6.5 Data kwaliteit

Datakwaliteit is een belangrijk onderdeel van Data governance.

De uitdagingen voor het monitoren van en zorg dragen over data kwaliteit binnen een MDM omgeving zijn:

- het identificeren van kritische data elementen
- het bepalen van welke data elementen bijdragen aan master data
- lokaliseren en isoleren van master data objecten binnen het bedrijf
- beoordelen en wegnemen van variaties tussen de verschillende representaties

om zo tot een geconsolideerde weergave te komen.

Zelfs na een initiële integratie van data naar een master data omgeving, is er nog steeds behoefte aan data inspectie, monitoring en controles ter identificatie van potentiële gevaren voor datakwaliteit.

Dimensies

Dimensies van datakwaliteit zijn afgestemd op de te meten bedrijfsprocessen. Dimensies geassocieerd met datawaarden lenen zich goed voor systeemautomatisering, maken zich geschikt voor dataregels binnen datakwaliteit tools voor data validatie.

Dit zijn o.a. de volgende dimensies:

- **Uniciteit**
Er is geen entiteit in de MDM omgeving die meer dan één keer voorkomt en deze kan ook uniek benaderd worden (sleutel) in een data set
- **Accuraatheid**
De mate waarin een gegeven op juiste wijze een “dagelijks” object vertegenwoordigen
- **Consistentie**
Twee datawaarden uit verschillende data sets mogen niet met elkaar conflicteren
- **Compleetheid**
Specifieke data elementen mogen niet de waarde nul hebben
- **Tijdigheid**
De verwachte tijd voor benadering en beschikbaarheid van informatie
- **Actualiteit**
De mate waarin informatie actueel is

Daarnaast wordt er gekeken naar data formaten en patronen en referentiële integriteit.

Data entiteiten die als master data objecten worden beheerd krijgen een uniek label.

Er moet geverifieerd worden of elke gelabelde entiteit refereert naar een bestaande entiteit binnen de omgeving.

Data integratie tools

Data kwaliteit en integratie tools zijn geëvolueerd van eenvoudige standaardisatie en patroonvergelijking naar toolsets voor complexe automatisering van data analyse, standaardisatie, matching en aggregatie. Bijvoorbeeld data profilering heeft zich ontwikkeld tot een set van complex geautomatiseerde analysetechnieken, welke gebruikt kunnen worden voor identificeren, isoleren, monitoren, auditen en het helpen vinden van afwijkingen welke de waarde van een bedrijfsinformatie master degraderen.

Data controles

Data kwaliteitscontroles zijn bedoeld voor het beoordelen van afwijkingen in data, het vaststellen van de oorzaak en deze, indien mogelijk wegnemen.

Ondanks alle moeite voor het bereiken van data kwaliteit, zullen er altijd afwijkingen zijn die aandacht vragen en waarvoor een oplossing gezocht moet worden. Hiervoor moeten protocollen opgesteld worden om fouten zo snel mogelijk op te sporen, naar de juiste persoon te sturen en binnen een redelijk tijd op te lossen.

5.6.6 Metadata management

Het volgende onderdeel van data governance is metadata management.

Alle aspecten om de noodzaak van MDM te bepalen vereisen een helder beeld van de informatie over data, welke in de organisatie gebruikt worden, de metadata.

Vaak wordt metadata omschreven als “gegevens over gegevens”. Het is echter meer dan alleen maten en types van elke data element, bijvoorbeeld ook betekenissen en domeinen.

Metadata kunnen worden ingedeeld in zeven niveaus, welke kritisch zijn voor MDM, zie figuur 5.

Bedrijfsdefinities			
Concepten	Bedrijfstermen	Definities	Semantiek
Referentie metadata			
Conceptuele domeinen Provincie	Waarde domeinen Utrecht	Referentie tabellen Onderlinge relaties van waardedomeinen	Mappings
Data elementen			
Kritische data elementen attributen en velden	Data element definities	Data formaten	Aliassen en synoniemen
Informatie architectuur			
Entiteitsmodellen	Relationele tabellen	Master object directory tracking object types bij data migratie	
Data Governance			
Informatiegebruik	Informatiekwaliteit	Datakwaliteit SLA's	Toegangbeheer
Services			
Service directory	Service gebruikers	Interfaces	
Bedrijfsmetadata			
Bedrijfsbeleid	Informatiebeleid	Bedrijfsregels	

Figuur 5 MDM Metadata stapel

De uitdaging is om te kijken naar de wijze waarop deze niveaus op elkaar in werken als onderdeel van een algehele metadata management strategie. In bijlage 3 wordt elke laag verder uitgewerkt.

5.6.7 Identificeren van Master data en Meta data

Het identificatieproces bestaat uit twee activiteiten, een top-down proces voor beoordeling van bedrijf datamodellen en procesmodellen om te identificeren welke data objecten kritisch zijn binnen de bedrijfsapplicaties. De andere activiteit is een bottom-up proces welke de bedrijfsgegevens evalueert voor het vinden van applicaties die data structuren gebruiken ter identificatie van master data objecten en deze samenbrengt naar een master data omgeving. Naast de omschrijving van master data objecten in paragraaf 5.1, kunnen deze ook nog als volgt gekarakteriseerd worden:

- dienen als referentie in meerdere bedrijfsgebieden en processen
- dienen als referentie in zowel transactionele als analytische systeemrecords
- kunnen geclassificeerd worden binnen een hiërarchie met verschillende niveaus van classificatie, attributen en specialisatie (bijv. “partij” kan betekenen “persoon”, “organisatie”, “klant”, etc.)

Top-down benadering, bedrijfsproces modellering
Organisaties moeten de relaties onderhouden tussen bedrijfsstrategie en de componenten van de applicatiemodellen. Door het beoordelen van bedrijfsprocesmodellen en vervolgens identificeren welke data objecten kritisch zijn in de processtromen van meerdere applicaties. De objecten die gedeeld worden met diverse bedrijfsactiviteiten zullen analisten gaan beheren als master data.

Bottom-up benadering, evaluatie van gegevens

Deze benadering identificeert master data objecten die reeds gebruikt worden binnen de organisatie. Het zoekt naar bedrijfsgegevens in applicaties welke in feite kopieën zijn van master objecten en gaat deze samenbrengen in een master data omgeving.

Het gaat daarbij om de volgende drie stappen:

1 Identificeren en centraliseren van semantische metadata

Om master data te beheren, moet men eerst in staat zijn om master metadata te beheren.

Hiervoor moeten dan eerst de volgende zaken op orde zijn:

- Formaat op element niveau: de juiste syntax
- Structuur op record niveau: dezelfde informatie over het data element
- Semantiek op elk niveau: dezelfde betekenis of definitie

2 Unificeren van data object semantiek

Dit omvat het begrijpen wat de gegevens betekenen, hoe die betekenis wordt overgebracht, hoe die betekenis data sets binnen de organisatie koppelt, hoe de gegevens gebruikt worden en benaderingen voor het omschrijven van semantiek als een attribuut voor het metadata raamwerk.

3 Identificeren en kwalificeren van master data

Door het toevoegen van karakteriseringinformatie voor elke metadata profiel van een data set, wordt er meer kennis toegevoegd aan het proces voor het bepalen van bron data sets, die geschikt zijn voor toevoegen aan de master data database, hetgeen de taak van de analist verlicht.

Het gebruik van een eenvoudig classificatieschema reduceert de complexiteit van gegevensvariatie en verlicht de beperkingen bij het vergelijken van meerdere metadata records.

De technieken en tools die gebruikt kunnen worden voor het bepalen van de bronnen van master data objecten zijn in essentie dezelfde type tools die later gebruikt worden om data te consolideren naar een master database. Het gebruik van data profilering, ontleding, standaardisatie en matching kan bijdragen aan het proces van het identificeren welke data sets (tabellen, bestanden, spreadsheets, etc) refereren naar welke master data objecten.

5.7 Functionele service laag

Onder services wordt verstaan: software tools, interfaces of applicaties die bepaalde acties of bewerkingen in een systeem automatiseren. Elk proces heeft zijn eigen service en al die services komen samen in een service raamwerk.

Master data kunnen op twee manieren geïntegreerd worden in een service raamwerk:

- Tactisch, een servicelaag om de transitie van applicaties naar de master opslag te faciliteren.
- Strategisch, een hiërarchische set van informatie services ter ondersteuning van snelle en efficiënte ontwikkeling van geautomatiseerde applicaties.

Beide noodzakelijke aspecten worden opgelost door een SOA, Service Orientated Architecture.

Een belangrijk aspect van een service georiënteerde benadering is dat de eisen aan een bedrijfsapplicatie de behoefte stuurt om “mogelijkheden” te identificeren en consolideren, welke door het bedrijf heen beheerd worden, op een manier dat het gedeeld en hergebruikt kan worden binnen de grenzen van een bestuurlijke omgeving.

Een organisatie die SOA toepast, is er een die zich aanpast van levering van applicaties naar een van levering van bedrijfsoplossingen.

Identificatie van services

Het identificeren van applicatie functionaliteit en categoriseren hoe die functionaliteit te abstraheren in een service laag, toont hoe de applicaties de data objecten raken die gerelateerd zijn aan master object typen.

Dit proces wordt gebruikt om die functionaliteit te standaardiseren voor vereenvoudiging en hergebruik.

Er zijn applicaties die gebruik maken van dezelfde set van kern mogelijkheden. Het doel is om deze mogelijkheden te identificeren in relatie tot hoe ze gebruik maken van master data objecten en deze te consolideren naar een set van services.

Met een service georiënteerde benadering is er een grotere snelheid en flexibiliteit in de mogelijkheid om te reageren op veranderingen in de bedrijfsomgeving. Het integreren van nieuwe mogelijkheden in de bedrijf service laag, zorgt er voor dat alle applicaties daar profijt van hebben, hetgeen ontwikkelingskosten reduceert en dat nieuwe mogelijkheden sneller operationeel zijn.

Enkele voorbeelden van kern services zijn:

- Levenscyclus services (creëren, lezen, wijzigen, afvloeien)
- Governance (audit, veiligheid, legitimatie, toegangscontrole)
- Integratie (integratie, data kwaliteit, metadata management)
- Consolidatie (modellering, match/samenvoegen, identiteit en hiërarchie management)
- Workflow/regels (zoeken en aanpassen, regels, workflow, invoeren en synchroniseren)

5.8 Management richtlijnen

Deze paragraaf behandelt een high-level overzicht van activiteiten voor het ontwikkelen van een MDM programma. Een van de moeilijkste delen van het totale proces is om te bepalen waar te beginnen. Het is vaak effectiever om externe experts in te schakelen tijdens de initiële fasen van het project, om daarmee zaken te bespoedigen. Voor meer detailinformatie, kan bijlage 3 geraadpleegd worden.

Om een MDM programma uit te voeren, moeten de volgende stappen genomen worden:

- 1 Rechtvaardiging verkrijgen voor MDM van het hoger management
Stel een business case op voor alle betrokkenen in de organisatie.
- 2 Ontwikkel een MDM roadmap en uitrolplan
 - 2a De implementatie van een complex programma vereist strategische planning, samen met details voor de gefaseerde uitvoering. Het opstellen van een roadmap voor het ontwerp en ontwikkeling van de infrastructuur en processen ter ondersteuning van MDM, schept duidelijkheid aan het management.
De verschillende stappen zijn in paragraaf 3 al aan de orde geweest.
 - 2b Het uitrol plan wordt ontworpen om de stappen uit de roadmap uit te voeren.
- 3 Beschrijf rollen en verantwoordelijkheden
Dit is in paragraaf 5.2 beschreven.
- 4 Project planning
Deze activiteiten zijn in voorgaande paragrafen beschreven (data governance, metadata, architectuur, etc..)

5.9 Deelconclusie

Er zijn in dit hoofdstuk vele aspecten van MDM aan de orde gekomen.

Naast veel omschrijvingen van begrippen, zijn bijvoorbeeld ook integratie, management en architectuur beschreven. De deelvraag die echter beantwoord moet worden is:

Welke methoden van dataverwerking en controle zijn er?

Dat zijn de aspecten, die in paragraaf 5.6 en 5.7 zijn behandeld:

- 1 Data consolidatie en integratie, het verzamelen en toegankelijk maken van data
- 2 Data synchronisatie, afstemming tussen applicaties en de master data set
- 3 Data governance
 - a Beheer van data standaarden, gestandaardiseerde definities
 - b Beheer van data kwaliteit (data integratie, matching, analyse, controle)
 - c Beheer van metadata (identificeren, kwalificeren)
- 4 Services, automatisering van voorgaande processen

6 Welke delen van MDM zijn voor Pepsico interessant?

Hier wordt gekeken naar aspecten uit het voorgaande hoofdstuk, die van belang kunnen zijn voor, met name de Data Management Organisatie (DMO) van Pepsico Nederland.

6.1 Inleiding

In dit hoofdstuk wordt eerst gekeken naar de huidige status van MDM bij Pepsico Nederland, vervolgens naar de gewenste situatie en op basis daarvan wordt bepaald hoe die eindsituatie bereikt kan worden.

Dit is gedaan met behulp van de MDM Maturity Matrix, het volwassenheidsmodel dat in paragraaf 5.3 behandeld is en in bijlage 3 verder uitgeschreven is. Die Maturity Matrix omvat alle aspecten van MDM, maar voor dit hoofdstuk is alleen het gedeelte data governance en business rules er uitgenomen, aangezien dit aspecten zijn die door de afdeling DataManagement te beïnvloeden zijn. Daarnaast staan deze zaken hoog op de agenda voor DMO, vanwege toename van de hoeveelheid data en het verspreiden en delen daarvan.

Zaken als identificatie en integratie zijn meer IT gerelateerd, architectuur en management worden meer op een strategische niveau bepaald.

De deelvragen die hier beantwoord worden zijn:

1. Welke methoden gebruikt Pepsico nu?
2. Welke eisen worden gesteld aan Data Governance?
3. Welke methoden zijn het meest geschikt voor Pepsico en waarom?

6.2 Welke methoden gebruikt Pepsico nu?

De afdeling datamanagement

De datamanagement afdeling verwerkt de supply chain gegevens van Pepsico Benelux. Aanvraagformulieren met nieuwe gegevens van leveranciers, producten, prijzen en klanten worden in een standaard formaat aangeleverd. Bij de verwerking van die gegevens worden de volgende controles uitgevoerd:

Controle tijdens verwerking:

- Business rules (regionaal bepaald)
De velden in de diverse systemen worden handmatig ingevoerd. De business rules staan in een Excel blad en zijn veldspecifiek opgesteld. Deze worden bij elke batch invoer geraadpleegd.
- Systeemregels, metadata (internationaal bepaald, regionaal aangepast)
SAP is een van de weinige systemen binnen Pepsico die gebruikt maakt van systeemregels. Dat houdt in, dat tijdens handmatige invoer veldspecifiek wordt gekeken naar bepaalde voorwaarden.
Zo zijn voor leveranciersgegevens bijvoorbeeld de velden “naam” en “plaats” verplicht, maar wordt er niet gecontroleerd op inhoud of naamgeving.

Controle achteraf:

- Validatie op basis van business rules (inclusief naming convention).
De gegevens worden gevalideerd door periodiek een rapport te draaien van alle aangemaakte en gewijzigde data objecten. In dat overzicht wordt nogmaals gekeken of de waarden conform de business rules zijn, maar er wordt ook gecontroleerd op consistentie.

- Data kwaliteit check op basis van PIRT¹ (op Europees niveau) en naming convention. Er worden periodiek centraal in Europa data kwaliteit rapporten gedraaid en vervolgens naar alle key users gestuurd. Zij gaan hun resultaten bekijken en zo nodig bijsturen.

Momenteel wordt er bij Pepsico Nederland gewerkt aan standaarden en definities van 140 data elementen, maar dit aantal zal binnen korte tijd verdubbelen. Er zijn een beperkt aantal data kwaliteit regels om te controleren of de waarden van die data elementen juist zijn.

6.3 Mate van volwassenheid

In paragraaf 5.3 is het MDM volwassenheidsmodel besproken. In bijlage 4 is een uitgebreider model opgesteld en heeft Pepsico Data Management daarin aangegeven wat de huidige status is van MDM, evenals de gewenste situatie. Uit dat overzicht zijn twee blokken genomen om die nader te bestuderen.

Die blokken staan in tabel 2 aangegeven.

	Technische component	Beheerst	Proactief	Strategisch
Governance overzicht verkrijgen	Algemeen	Business gerelateerd databeheer	Enterprise data governance programma aanwezig	Data governance door de hele organisatie verzekert hoge kwaliteit van informatie uitwisseling
	Data standards	Processen aanwezig	Bedrijfsbreed aanwezig	
	Metadata management		Bedrijfsbreed aanwezig	
	Data kwaliteit	Beleid en procedures voor data quality management, data quality tools aanwezig, data quality issues	Proactieve monitoring voor data quality controles in het governance programma	
	Data Stewardship Operationele aspecten	Line of business niveau		
Business Process Management Hoogste abstractieniveau	Business rules Combineert de evaluatie van (gedeelde) data object waarden met de beschreven waarden in gedefinieerde controles	Integratie van business rules met masterdata operations	Business rules worden geïntegreerd in validatiesoftware en beschikbaar gemaakt op bedrijfsproces niveau	Applicaties integreren validatiesoftware op grote schaal

Tabel 2 Aandachtsgebieden van Pepsico DMO uit de MDM Maturity Matrix

De aandachtsgebieden die verdere aandacht behoeven zijn dus:

- Data governance, met de componenten: data standaarden, metadata management, data kwaliteit en data stewardship/gegevensbeheer.
- Business Process Management met de component: business rules.

De huidige status van die gebieden zijn over het algemeen beheerst. De wens van Pepsico Nederland is om die volwassenheidsniveaus naar proactief en uiteindelijk strategisch te brengen. Aangezien bovenstaande aandachtsgebieden elk op zich ook weer bestaat uit een wereld van processen, mogelijkheden en toepassingen, worden hierna slechts twee gebieden uitgewerkt, namelijk data kwaliteit en business rules. Data kwaliteit is een erg belangrijk aspect voor Pepsico DMO, omdat niet alleen interne processen hiervan afhankelijk zijn, maar bijvoorbeeld ook klanten en consumenten.

¹ PIRT = Pepsico International Reference Template

Er worden gegevens gepubliceerd naar een externe datapool waaruit klanten hun logistieke gegevens halen (maten, gewichten) en consumenten de voedingswaarden van producten en eventuele allergieën kunnen uitlezen.

Business rules dragen bij aan data kwaliteit, aangezien die aangeven hoe bepaalde gegevensvelden ingevuld moeten worden.

6.4 Data governance

Algemeen

Data governance bij Pepsico Nederland is momenteel op het beheerste niveau: business gerelateerd databeheer. De volgende stap die gemaakt moet worden is uitbreiding en afstemming met meerdere business units, eerst Europa, later wereldwijd.

Het uiteindelijke doel is om data governance door de hele organisatie te brengen om hiermee een hoge kwaliteit van informatie uitwisseling te bereiken.

Welke eisen worden door Pepsico Nederland gesteld aan data governance?

Het is lastig om een eisenpakket op te stellen voor data governance.

“Er is geen enkele afdeling die een uitgebreide opsomming kan geven van eisen aan MDM in een organisatie en sommige eisen op afdelingsniveau kunnen met elkaar conflicteren.” (1)

Het is een zaak van het hoger management om bij te dragen aan data governance en te zorgen voor een succesvolle uitvoering ervan.

Enkele eisen die naar voren gebracht worden zijn:

- Duidelijke identificatie en koppeling van data objecten
Sleutel attributen aanwezig tussen verschillende systemen, product hiërarchie en klant hiërarchie aanwezig
- Gebruikers ondersteunen het data governance programma
- Governance processen centraal in de organisatie aanwezig
- Voldoen aan de informatiebehoefte, meerdere data elementen meenemen in de standaardisatie

Welke methoden zijn het meest geschikt voor Pepsico en waarom?

De eerder vastgestelde methoden van gegevenswerking waren:

- 1 Data consolidatie en integratie, het verzamelen en toegankelijk maken van data
- 2 Data synchronisatie, afstemming tussen applicaties en de master data set
- 3 Data governance
 - a Beheer van data standaarden, gestandaardiseerde definities
 - b Beheer van data kwaliteit (data integratie, matching, analyse, controle)
 - c Beheer van metadata (identificeren, kwalificeren)
- 4 Services, automatisering van voorgaande processen

In principe zijn alle methoden essentieel voor MDM binnen Pepsico.

MDM is, zeker voor zo'n grote internationale organisatie met honderden bedrijven en systemen, een belangrijk onderdeel om goed te kunnen functioneren.

1 Berson, Dubov 2011, p. 409.

Als er gekeken wordt naar bovenstaande punten, gericht op de afdeling DMO, is data governance een speerpunt om redenen die in de inleiding (§ 6.1) genoemd zijn.

Zoals eerder vermeld, is data kwaliteit een hoofdzaak voor data governance, dus wordt hieronder een voorbeeld uitgewerkt om het volwassenheidsniveau van data kwaliteit bij Pepsico Nederland te verhogen.

Aangezien controles handmatig worden uitgevoerd, is het interessant om ook naar services te kijken.

6.5 Data kwaliteit

Het huidige volwassenheidsniveau van data kwaliteit zit nu ergens tussen beheerst en proactief in. Beleid, procedures en tools voor data kwaliteit management zijn aanwezig, evenals proactieve monitoring voor data kwaliteitscontroles in het governance programma (audits).

De wens is om uiteindelijk een strategisch niveau te bereiken. Dat kan met behulp van meer geavanceerde technieken, maar efficiënter en effectiever is het toepassen van services.

Onder services wordt verstaan: software tools, interfaces of applicaties die bepaalde acties of bewerkingen in een systeem automatiseren. Zo kan bijvoorbeeld invoeren van data, integreren en synchroniseren geautomatiseerd worden, maar ook data kwaliteit of metadata management.

Technieken

Zoals gezegd, is de effectiviteit om bepaalde (handmatige of semiautomatische) technieken toe te voegen ter verhoging van de data kwaliteit niet zo hoog als bij het toepassen van services.

In de eerste plaats omdat er al veel technieken beschikbaar zijn, ook geautomatiseerd, ten tweede, zijn veel nieuwe technieken zo complex, dat ze beter meteen geautomatiseerd (als een service) toegepast kunnen worden.

Er is een steeds grotere behoefte aan het meten van data governance c.q. data kwaliteit.

Het aantal regels die gedefinieerd en toegepast zouden moeten worden om de kwaliteit van bedrijfsgegevens te meten en te verbeteren is buitensporig. Daarom is er een dringende vraag naar generieke, wetenschappelijke en model gedreven benaderingen van data kwaliteit om tot het volwassenheidsniveau van proactief of strategisch te komen. Om toch een voorbeeld te geven van een goede techniek, wordt hieronder de informatie theorie beschreven.

Informatie theorie

Een voorbeeld van een wetenschappelijke benadering van data kwaliteit voor MDM is gebaseerd op Informatie theorie, een statistische methode om de velden in een record te vergelijken met diezelfde velden in een referentie record. In bijlage 6 wordt deze methode nader toegelicht.

Het voordeel van deze meting is dat het niet alleen kwantificeerbaar is, maar ook dat het gebruikt kan worden voor het bepalen van de data kwaliteit van delen van attributen (tokens), een functie die nieuwe inzichten verschaft in de oorzaak van de meest significante data kwaliteitsproblemen.

Records met een lage score zijn doelen voor gegevensbeheer. Op het moment dat er additionele gegevens zijn, zullen complete metingen voldoen aan de data governance eisen. Dit voorbeeld illustreert, hoe data governance kan zorgen voor continue verbeterprocessen van data kwaliteit door middel van een paar opgestelde metingen.

Organisaties kunnen het voorgaande toepassen om zo het data governance beleid uit te breiden en zorgen voor processen die de volwassenheid verhogen.

Door het consistent gebruik van deze kwantitatieve benadering door de organisatie zullen data governance organisaties waardevolle inzichten krijgen in hoe metingen toegepast worden op data karakteristieken en scenario's.

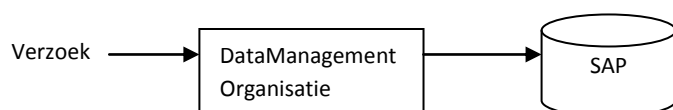
De beschreven benaderingen kunnen worden geïntegreerd in toekomstige producten, zoals dashboards voor het hoger management en data kwaliteit tools. Data governance zal in staat zijn informatiebronnen te selecteren die gemeten moeten worden, deze te kwantificeren door middel van de voorgestelde metingen, de voortgang te meten door de tijd, de bevindingen te rapporteren en acties aan te bevelen aan het hoger management.

Services

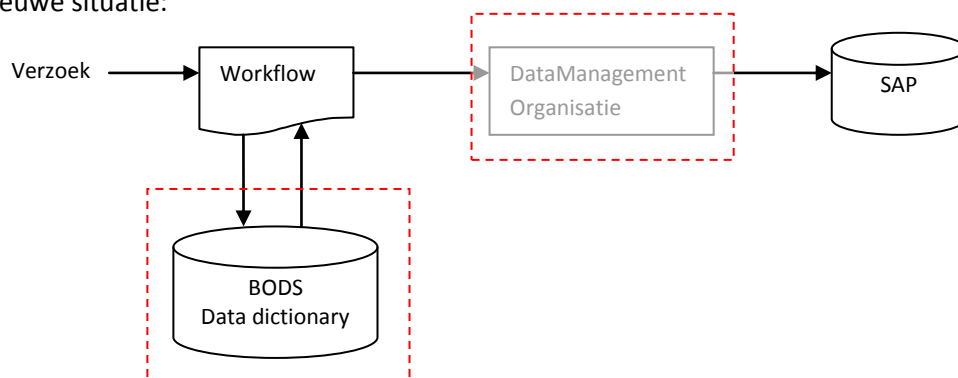
Aangezien DataManagement zijn governance processen goed op orde heeft, maar vrijwel alles nog handmatig wordt uitgevoerd, is de stap naar automatisering van die processen een logisch vervolg. Figuur 6 illustreert de huidige situatie en een toekomstige situatie.

De controle wordt zowel tijdens als na de verwerking door de datamanagement medewerker uitgevoerd,.

Huidige situatie:



Nieuwe situatie:



Figuur 6 Huidig en nieuwe situatie van gegevensverwerking, data kwaliteit controle in de workflow

Als er op een later tijdstip de legacy systemen worden geïntegreerd naar SAP, biedt dat mogelijkheden voor kwaliteitscontrole. Door gebruik te maken van de workflow functie in SAP, kunnen aanvragen automatisch verwerkt worden en is dus digitale verificatie mogelijk.

Allereerst zullen data kwaliteit regels opgesteld moeten worden en op basis daarvan controleren of de data daarmee in overeenstemming zijn.

Een data dictionary (metadata opslag) in BODS (interface, zie § 7.1) wordt gevuld met regels op het gebied van naming convention, zoals afkortingen, maar ook categorieën.

Als er een aanvraag in de workflow wordt toegevoegd, zal er een verificatie plaatsvinden met de metadata gegevens in BODS. Vervolgens wordt de aanvraag doorgestuurd naar de afdeling datamanagement voor verder verwerking. Een volgende stap naar volwassenheid zal de automatische verwerking van de aanvraag worden, geheel op de gegevens in BODS.

De rol van datamanagement zal veranderen in steeds meer een validerende functie, het bekijken van data die met behulp van de workflow automatisch in SAP zijn ingevoerd. Het gaat dan voornamelijk om vals negatieve of vals positieve gegevens te onderzoeken, data elementen die grote gelijkenis vertonen, maar niet gelijk zijn, of andersom. Datamanagement zou dan migreren naar data governance.

6.6 Business Proces Management

Business rules dragen bij aan data kwaliteit, aangezien die aangeven hoe bepaalde gegevensvelden ingevuld moeten worden. Ook dit is geschikt om als een service te implementeren.

Business rules

Om de bedrijfsgegevens te kunnen valideren, wordt gebruik gemaakt van business rules. Hierin worden per data attribuut de eisen en voorwaarden beschreven, waaraan volgens de organisatie moet worden voldaan.

Voor data synchronisatie uit verschillende systemen, kan gebruik gemaakt worden van een Business Rules Engine.

Een Business Rules Engine (BRE) is een software toepassing of een systeem dat ontworpen is om business rules te beheren op basis van een bepaalde prikkel (bijvoorbeeld verandering van een attribuutwaarde). Deze is apart ontworpen van de applicatie, waarmee het onderhoud van de applicatie minder tijd, moeite en kosten vraagt, aangezien gebruikers de business rules kunnen wijzigen zonder dat de applicatie wordt gewijzigd.

Om business process management naar een meer volwassen niveau te brengen is automatisering van het toepassen van business rules een sterke optie. Dit moet dan ook bedrijfsbreed gebeuren. Als de huidige business rules (in Excel formaat) bij datamanagement in een BRE worden gezet, kan dit ook als een workflow service(zie § 5.7) fungeren en steeds verder ontwikkeld worden.

6.7 Deelconclusie

De deelvragen die in dit hoofdstuk beantwoord worden zijn:

1. Welk methoden gebruikt Pepsico nu?

Pepsico Nederland gebruikt nog veel handmatige verwerkings- en controlemethoden. Dit zijn:

- Business rules, processen aanwezig
- Metadata, systeemregels, nog niet ten volle benut
- Data validatie (achteraf) op basis van business rules
- Data kwaliteit check, periodiek, nog in ontwikkeling

2. Welke eisen worden gesteld aan Data Governance?

Hoewel het lastig is om heldere eisen te stellen aan data governance zijn de volgende benoemd:

- Duidelijke identificatie en koppeling van data objecten
- Gebruikers ondersteunen het data governance programma
- Governance processen centraal in de organisatie aanwezig
- Voldoen aan de informatiebehoefte, meerdere data elementen meenemen in de standaardisatie

3. Welke methoden zijn het meest geschikt voor Pepsico en waarom?

In principe zijn alle methoden essentieel voor MDM binnen Pepsico.

MDM is, zeker voor zo'n grote internationale organisatie met honderden bedrijven en systemen, een belangrijk onderdeel om goed te kunnen functioneren.

- Data governance - beheer van data kwaliteit is een speerpunt
- Omdat controles handmatig worden uitgevoerd, is het een logische vervolgstap om meer van services gebruik te maken.

7 Praktijkproject MDM, het ontwerpen van een data interface

Als praktijkproject voor MDM wordt een interface ontworpen, om te kijken op welke manier MDM methoden zijn toe te passen als gegevens vanuit een bestaand ERP systeem overgezet worden naar een datapool.

7.1 Inleiding

In dit hoofdstuk wordt eerst de context beschreven en komen enkele begrippen aan bod. Na de projectafbakening volgt een procesbeschrijving van het ontwerp, wordt de bouw en test toegelicht, de implementatie en nazorg en het support model. Het support model gaat over de ondersteuning van de interface voor gebruikers op het moment dat er zich storingen of foutmeldingen voordoen.

Tenslotte worden MDM aspecten van dit project toegelicht en wordt afgesloten met een deelconclusie.

Het gaat bij het ontwerpen van de interface met name om het “mappen” (afstemmen) van de velden in het bronsysteem met die in het doelsysteem. Zowel het veldformaat als de inhoud zijn in verschillende systemen vaak niet gelijk.

Met behulp van een interface en conversietabellen is een goede afstemming te bereiken. Het doel van de interface is uiteindelijk om de gegevensoverdracht automatisch te laten verlopen. Daarnaast zijn er aanvullende services beschikbaar voor het verhogen van de data kwaliteit.

Bron

Het bronsysteem is Oracle. Waarom Oracle als bron?

Pepsico heeft verschillende systemen waar informatie opgeslagen is. Voor deze interface is voor Oracle gekozen, aangezien deze de meest relevante productinformatie bevat. In de nabije toekomst gaat de informatie in Oracle over naar SAP. Om reeds vóór de migratie naar SAP ervaring op te doen met het bouwen van een interface naar PDM, heeft Pepsico al een template in handen, zodat de latere interface van SAP naar PDM een relatief kleine aanpassing vergt.

Het is tevens een bewijs, dat BODS niet afhankelijk is van een bepaald ERP systeem, maar met meerdere systemen uit de voeten kan.

Interface

De interface tool die wordt gebruikt is Business Objects Data Services (BODS).

In bijlage 2 van bijlage 1, het Plan van aanpak, staat deze tool beschreven.

Waarom BODS als interface?

SAP Business Objects is reeds bij Pepsico beschikbaar, maar voor Europa slechts in geringe mate.

Met dit project kan ervaring opgedaan worden met BODS als voorbereiding op het SAP project dat binnenkort wordt opgestart door Pepsico Europa.

De kracht van BODS, mede met betrekking tot MDM, zijn o.a. de onderstaande modules:

De Data Services Designer

Biedt een aantal voorgedefinieerde transformaties en functionele objecten, waarmee je ETL flows kunt modelleren, zoals data merging, validatie van records en vergelijking van verschillende tabellen.

De Management Console

Bevat o.a. de volgende functies:

- Data validatie, op basis van de gedefinieerde regels en inschatting en identificatie van de potentiële fouten of inconsistenties van de data bron.

- Impact analyse en Data Lineage Analysis, deze module integreert metadata van de SAP BO platformen van de bron naar het rapport. Deze kan een gegeven rapport element terug herleiden naar het bronsysteem en uitzoeken welke transformaties zijn uitgevoerd om deze waarde te bepalen.

Doelsysteem

Het doelsysteem, de interne datapool, dat Pepsico gebruikt is PDM, Product Data Management (zie bijlage 1 van bijlage 1, het Plan van aanpak). De gegevens in PDM worden binnen Pepsico beheerd en worden elektronisch naar de externe datapool (1Sync) gestuurd.

SRC-PDM is een reeds aanwezige, veelgebruikte tool binnen Pepsico (SRC is de ontwikkelaar).

7.2 Projectafbakening

Voor dit project worden de volgende zaken opgeleverd:

- Mapping van de data van Oracle naar PDM
- Beschrijving van de bouw van de interface, test en implementatie. Indien er voldoende tijd beschikbaar is, wordt deze interface ook gebouwd en geïmplementeerd
- Beschrijving van het support model, alleen een voorstel van hoe te implementeren

In het plan van aanpak (bijlage 1) was ook het interface ontwerp, met toepassing van BODS regels en mogelijkheden opgenomen. Deze is in een later stadium vervallen, omdat de interface volledig ontworpen en gebouwd gaat worden door Pepsico BIS UK. Hierdoor is het niet relevant meer om deze zaken op te nemen.

Systemen

Het gaat hierbij om de volgende systemen:

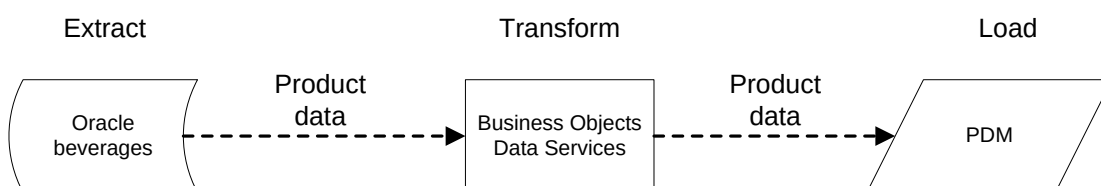
Oracle, BO Data Services, PDM

Data

- Beverage product data, inclusief de klanten waar naartoe gepubliceerd wordt
 - Productnaam
 - Producthierarchy: brand, subbrand, smaak
 - EAN code
 - Units of Measure (UoM): Maten (hoogte, breedte, diepte), Gewicht (Bruto, Netto)
 - Verpakking: Vorm (doos, fles, pak), Aantal CE's per HE (flessen per krat)
 - Palletisering: aantal kratten per laag, aantal lagen per pallet

Proces

- Oracle data extractie:



Wat is niet in scope:

Het volgende wordt niet opgeleverd:

- Support model, implementatie (middelen, set up tool, medewerkers)

Systemen

SAP, AS400, Genxt Masterfiles, Manugistics, 1Sync

Data

Vendors, prijzen, gebruikers en klanten

Processen

- Gegevensuitwisseling met retail (1Sync)
- SAP data extractie:



7.3 Het ontwerp

Het ontwerpproces volgt in grote lijnen de PLM methode, zie bijlage 3 van bijlage 1, het Plan van aanpak.

Voor dit proces zijn de volgende stappen doorlopen:

1. Voorbereiding en planning
Hiervoor is een zogenaamde project charter opgezet, een document waarin beschreven staat wat de bedoeling van het project is, de scope, de bedenkingen en risico's, de betrokkenen, het kostenplaatje en een planning. Deze onderwerpen staan in bijlage 1, het Plan van aanpak beschreven.
2. Overleg: haalbaarheid en akkoord
In dit overleg is aan de hand van de project charter bekeken of het project haalbaar is, ook gezien de tijd. Er is door de sponsors akkoord gegeven om aan het project te beginnen.
3. Bestuderen van het bronsysteem (Oracle) en doelsysteem (PDM)
In Oracle is gekeken naar alle attributen: veldlengtes, naamgeving en karakters.
In PDM is gekeken naar welke velden er gevuld moeten worden (mandatory) en wat de business rules zijn.
4. Experts interviewen
De Oracle en PDM expert zijn geïnterviewd en zij hebben documentatie en diverse bestanden gestuurd om een goed beeld te krijgen, hoe die systemen zijn opgebouwd en welke gegevens er in staan.
5. Mapping maken
Op basis van voorgaande gegevens is een vergelijking gemaakt tussen beide systemen op velden niveau. Voorbeeld "Brand": In Oracle staat daar "LO", maar in PDM moet de volledige merknaam (Looza) komen.
Daarnaast moeten in PDM de producten in diverse categorieën ingedeeld worden, zoals functionele naam en productgroep. Hiervoor zijn dus conversietabellen nodig.

6. Ontwerp maken
Het ontwerp bestaat naast de mapping uit conversietabellen, functionele specificaties en eisen.
7. Overleg: start bouw?
Aan de hand van een voorstel, afgegeven door Pepsico BIS, waarin de benodigde tijd en kosten zijn beschreven, is door de sponsors bepaald dat met de bouw begonnen kan worden. Uiteindelijk is echter besloten om niet met de bouw te beginnen vanwege het komende SAP project. Door dit SAP project zijn er onvoldoende mensen beschikbaar om het BODS project te (bege)leiden.
8. Bouwen en testen
9. Overleg: Go/no Go?
10. Implementeren, nazorg (hypercare) en support

Specificaties en eisen

Aan welke specificaties en eisen de interface moet voldoen staat in bijlage 10 beschreven.

Functioneel ontwerp

In bijlage 11 wordt gedetailleerd beschreven waar gegevens in Oracle te vinden zijn, hoe die geconverteerd moeten worden in BODS en tenslotte geladen worden in PDM.

7.4 Bouwen en testen

Bouwen

Aan de hand van voorgaande gegevens is door Pepsico BIS een voorstel gedaan voor het aantal benodigde dagen en de kosten voor dit project.

Dit komt neer op 38 dagen, ca. 7 weken, en € 7.200,- aan kosten.

De bouw wordt uitgevoerd door Pepsico BIS UK en eventueel door hen gecontracteerde externe partijen.

Testen

De test wordt uitgevoerd door een of twee medewerkers van de Pepsico DMO afdeling.

Na elke testload worden alle gegevens in PDM gecontroleerd op juistheid en consistentie.

7.5 Implementatie en nazorg

Cutover plan

Bij veel hard- en software aanpassingen wordt een cutover plan gemaakt, om de implementatie zo efficiënt en snel mogelijk te laten plaatsvinden, zodat gebruikers er zo weinig mogelijk last van ondervinden. Bij dit project is de aanpassing zodanig, dat het voor de gebruiker nauwelijks van invloed is.

Zowel Oracle als PDM blijven op gelijke wijze functioneren. BODS draait op de achtergrond.

Implementatie

Als de testen goed zijn verlopen, kan de interface in productie genomen worden.

Ook hier worden de eerste loads intensief gecontroleerd.

Nazorg

De eerste twee weken blijft het Pepsico BIS team paraat, om eventueel zich voordoende problemen te verhelpen. Daarna wordt de service overgedragen aan het support team.

7.6 Support model

Het support model gaat over de ondersteuning van de interface voor gebruikers op het moment dat er zich storingen of foutmeldingen voordoen.

Als de BODS interface in productie is genomen en door de gebruiker is goedgekeurd, wordt de applicatie ondersteuning overgedragen aan het support team.

Het support model is gebaseerd op ITIL, Information Technology Infrastructure Library. Dit is een standaardaanpak voor een effectieve en efficiënte levering van IT services.

Dit raamwerk omvat vele functies en processen, maar voor dit relatief kleine project zijn alleen Incident management en Change management uitgewerkt.

Zie bijlage 12 staan de procedures schematisch weergegeven, gevolgd door een beschrijving.

De betrokkenen bij support zijn:

- Gebruiker, welke meestal de melding doet
- Lokale Service Desk (SD), welke de melding registreert en doorstuurt
- BIS of WER (West European Region) solutions team, welke voor de oplossing zorgt
- WER technology team, voor interne software en hardware oplossingen
- Leverancier van software en/of IT oplossingen
- Demand of project manager, voor aansturing van teams
- Sector change management, voor het registreren, controleren en gezamenlijk goedkeuren van releases

7.7 MDM aspecten

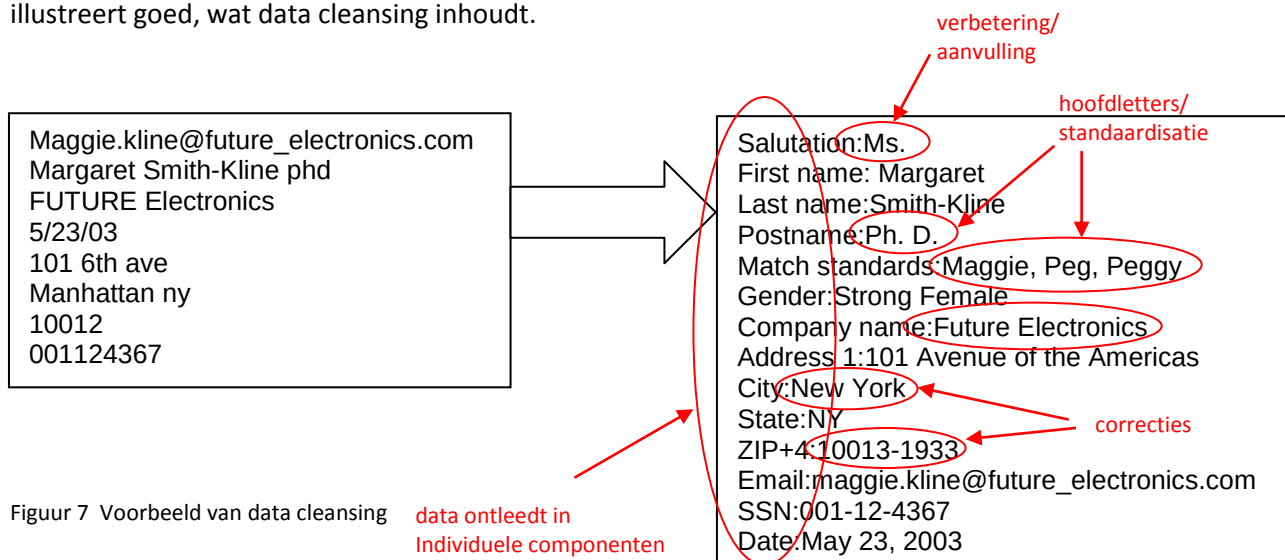
Ook bij het toepassen van deze interface is het van belang dat de gegevens juist zijn.

BODS voorziet daarin met een aantal mogelijkheden, zoals profilering, integratie, kwaliteit, ontwikkeling en metadata management. Daarnaast bevat het uitgebreide ETL- en data integratie functies.

Op het gebied van data kwaliteit zijn er ingebouwde functionaliteiten, zoals data cleansing en matching en consolidatie.

Data cleansing

Bij de interface gaat het om productgegevens, maar onderstaand voorbeeld van klantgegevens illustreert goed, wat data cleansing inhoudt.



Figuur 7 Voorbeeld van data cleansing

In BODS worden standaardisatieregels opgenomen met de voorwaarden, waaraan bepaalde velden moeten voldoen. Als voorbeeld het veld “Name”, zie de mapping in bijlage 7.
Deze moet voldoen aan de volgende structuur:

LO A ACE ORIGINAL 20CL REG

Brand (2), gekoeld (1), subbrand (3), smaak (x), inhoud (x), maat (2) en versie (x).

Naast datastandaarden wordt ook gebruik gemaakt van business rules.

Hierin wordt per veld aangegeven wat de voorwaarden zijn.

Bijvoorbeeld het veld “publishtoGLN1” geeft als resultaat:

if local=NL then publish to data checker GLN 8712345012021

if local=BE then publish to GLN coupon 04051633000000

Tenslotte moet er nog een data kwaliteit controle plaatsvinden.

Dit gebeurt al in PDM, het doelsysteem waarde de data van BODS terecht komt.

Die controle wordt straks door BODS uitgevoerd, voordat het in het doelsysteem terecht komt.

MDM volwassenheidsniveau van dit project

De toepassing van de BODS interface brengt het volwassenheidsniveau naar proactief.

Op het gebied van data governance wordt er al wel gebruik gemaakt van diverse services, maar slechts een beperkt aantal. Bijvoorbeeld wel data standaarden en data cleansing, maar nog geen profilering en integratie, zeker niet organisatiebreed.

Business rules worden geïntegreerd in validatie software en beschikbaar gemaakt op bedrijfsproces niveau, maar zijn nog niet op grote schaal beschikbaar.

Is de interface herbruikbaar voor SAP?

De interface is voor het grootste gedeelte herbruikbaar om later gebruikt te worden tussen SAP en PDM. Alleen het “Extract” gedeelte van de ETL functie moet aangepast worden. De voorwaarden, conversietabellen en koppeling naar PDM zijn dan al aanwezig.

7.8 Deelconclusie

De bedoeling van dit praktijkproject was het ontwerpen van een data interface en daarvoor zijn de volgende producten opgeleverd:

- Een mapping van de data van Oracle naar PDM, zie bijlage 7.
- Beschrijving van de bouw van de interface, test en implementatie.
De bouw en implementatie is uitgesteld, vanwege het komende SAP project. Hierdoor zijn diverse deelnemers aan dit project (BODS interface) niet beschikbaar.
- Beschrijving van het support model, alleen een voorstel van hoe te implementeren

Het ontwerpen van de interface is een goede ervaring om te weten te komen hoe systeemstructuren in elkaar zitten, hoe de inhoud er uit ziet en hoe deze systemen met elkaar kunnen communiceren.

Met dit project wordt ook de rol van MDM duidelijk, gericht op deze praktijkopdracht.

Hoe weet ik dat de aangeboden data juist zijn?

Door gebruik te maken van interne systeemtoepassingen van BODS kan een hoge data kwaliteit bereikt worden. Het gaat dan om data cleansing, consolidatie, het gebruik van business rules en data kwaliteitscontrole (zoals de data validatie functie in de Management Console).

8 Conclusies en aanbevelingen

Tot slot wordt als conclusie een antwoord gegevens op de hoofdvraag:

Wat zijn goede methoden om correcte data aan de klant aan te bieden?

In principe zijn dat alle aspecten die in deelvraag 1 zijn behandeld:

Welke methoden van dataverwerking en controle zijn er?

- 1 Data consolidatie en integratie, het verzamelen en toegankelijk maken van data
- 2 Data synchronisatie, afstemming tussen applicaties en de master data set
- 3 Data governance
 - a Beheer van data standaarden, gestandaardiseerde definities
 - b Beheer van data kwaliteit (data integratie, matching, analyse, controle)
 - c Beheer van metadata (identificeren, kwalificeren)
- 4 Services, automatisering van voorgaande processen

Daarnaast zijn deelvragen beantwoord in de specifieke situatie van Pepsico:

Deelvraag 2 Welke methoden gebruikt Pepsico nu?

Pepsico Nederland gebruikt nog veel handmatige verwerkings- en controlemethoden. Dit zijn:

- Business rules, processen aanwezig
- Metadata, systeemregels, nog niet ten volle benut
- Data validatie (achteraf) op basis van business rules
- Data kwaliteit check, periodiek, nog in ontwikkeling

Deelvraag 3 Welke eisen worden gesteld aan Data Governance?

Hoewel het lastig is om heldere eisen te stellen aan data governance zijn de volgende benoemd:

- Duidelijke identificatie en koppeling van data objecten
- Gebruikers ondersteunen het data governance programma
- Governance processen centraal in de organisatie aanwezig
- Voldoen aan de informatiebehoefte, meerdere data elementen meenemen in de standaardisatie

Deelvraag 4 Welke methoden zijn het meest geschikt voor Pepsico en waarom?

In principe zijn alle methoden essentieel voor MDM binnen Pepsico.

MDM is, zeker voor zo'n grote internationale organisatie met honderden bedrijven en systemen, een belangrijk onderdeel om goed te kunnen functioneren.

- Data governance - beheer van data kwaliteit is een speerpunt
- Omdat controles handmatig worden uitgevoerd, is het een logische vervolgstap om meer van services gebruik te maken.

Deelvraag 5 Hoe weet ik dat de aangeboden data juist zijn?

Door gebruik te maken van bijvoorbeeld interne systeemtoepassingen van BODS, kan een hoge data kwaliteit bereikt worden. Het gaat dan om data cleansing, consolidatie, het gebruik van business rules en data kwaliteitscontrole (zoals de data validatie functie in de Management Console).

Hiermee zijn de doelen van de afstudeeropdracht bereikt, namelijk:

- De afdeling DataManagement beschikt over een beschrijving van de huidige stand van MDM en wat dit kan betekenen voor de afdeling/organisatie. Ook welke onderdelen kunnen worden gebruikt en hoe die kunnen worden toegepast.
- De afdeling IT beschikt over een beschrijving van de te bouwen interface, welke artikelinformatie van Oracle beverages naar PDM stuurt.
- Hopelijk heeft de student zowel voor Pepsico als voor de Hogeschool Utrecht de afstudeeropdracht met goed resultaat afgerond.

9 Evaluatie van het proces

De afstudeeropdracht bestaat uit een deel theorie (bestuderen) en een deel praktijk (interface ontwerpen).

Praktijk

Na akkoord vanuit de HU om met de opdracht te mogen beginnen, ben ik eerst gestart met het praktijkonderdeel van de opdracht.

Ik heb daar voor gekozen, omdat daarvoor diverse zaken geïnitieerd moeten worden, waarbij je tevens afhankelijk bent van anderen. Daar kan vaak veel tijd overheen gaan, vanwege drukke agenda's en andere prioriteiten. Dat bleek een goede inschatting, zeker in een later stadium, omdat toen de meeste betrokkenen (waaronder ik) in een rap tempo aan de voorbereidingen moesten beginnen van een nieuwe project (SAP systeemintegratie).

De meeste gegevens voor het ontwerpen van de interface waren toen al beschikbaar en de afstemming met de interface bouwer waren al in volle gang.

Bouw interface

In het plan van aanpak was ook opgenomen om de toepassing van regels en mogelijkheden in BODS in het interface ontwerp te onderzoeken. Dit is in een later stadium (in de ontwerp fase) vervallen, omdat de interface volledig ontworpen en gebouwd zou gaan worden door Pepsico BIS UK.

Dat was op dat moment een verrassend bericht en het bleek al snel besloten. Dat had ook te maken met de start van het nieuwe SAP project, waarbij de bouw van BODS interfaces voornamelijk intern zal gaan plaatsvinden. Een logisch gevolg is dus, dat dit project daar ook gebruik van gaat maken.

Door deze actie is het niet relevant meer om die regels en mogelijkheden in BODS te bestuderen.

Ook is daarmee het bestuderen van MDM methoden en toepassingen in BODS vervallen.

Dat is aan de ene kant jammer, want dat zou erg leerzaam geweest zijn, maar aan de andere kant niet haalbaar, vanwege tijdsdruk.

Bereikbaarheid van de betrokkenen

Dit bleek in de meeste gevallen goed te verlopen, er werd meestal binnen een of twee dagen gereageerd op verzoeken, doorgaans via mail. Ook afspraken plannen verliep goed.

De benodigde gegevens voor het ontwerpen van de interface kwamen vrij snel beschikbaar.

Begeleiding

De begeleiding voor zowel het praktijk als het theorie gedeelte verliep goed. Er is regelmatig contact geweest en er is goede ondersteuning verleend, daar waar nodig.

Theorie

Het theoretisch gedeelte hield voornamelijk literatuurstudie in.

Helaas zijn boeken en artikelen over MDM vrijwel allemaal in het Engels. Dat heeft veel extra tijd gekost om die stof te vertalen. De specifieke termen en afkortingen maken dat nog iets lastiger.

Bronvermeldingen

Literatuur:

Loshin, D. (2009). *Master Data Management*. 1e dr. Burlington: Morgan Kaufman

Berson, A./ Dubov, L. (2011). *Master Data Management en Data Governance*. 2e dr. McGraw-Hill

Intranet:

KnowledgeNet, International Project Life-cycle Methodology Handbook (2 maart 2012)

Internet:

SAP

Hill, S. (2009). pdf presentatie van SAP BusinessObjects Data Services ,
www.sdn.sap.com/irj/boc/index?rid=/webcontent/uuid/f0b79c51-ff38-2c10-eb8b-8bda91617662
(24 februari 2012)

Gartner

www.gartner.com/technology/research.jsp (26 februari 2012)

InformationAge

www.information-age.com/ (diverse data)

Seminar:

Henderyckx, J. (2011). Master Data Management, Envizion, 11 oktober 2011

Bijlage 1 Plan van aanpak MDM bij Pepsico



PEPSICO

Student: Henk van Beek
ID 1578767

Hogeschool Utrecht, richting BI

Docent begeleider: Jos van Dongen
15 maart 2012

Inhoudsopgave

Inleiding	3
1 Aanleiding	3
2 Context	4
3 Probleemstelling	4
4 Opdrachtoomschrijving	4
5 Projectgrenzen	5
6 Doelstellingen	6
7 Hoofdvraag en deelvragen	6
8 Activiteiten	7
9 Eindproducten en deelproducten	7
10 Methoden en technieken	7
11 Kosten/baten	7
12 Organisatie	8
13 Risico's	9
14 Planning	10
15 Kwaliteitsbewaking	10
Bronnen	11
Bijlagen	
1 PDM	12
2 BusinessObjects Data Services	13
3 PLM, International Project Life-cycle Methodology	14

Inleiding

In de afstudeerfase van mijn studie aan de Hogeschool Utrecht (HU), is het de bedoeling om bij mijn werkgever een opdracht te zoeken met betrekking tot de richting Bedrijfskundige Informatica. Dit plan van aanpak beschrijft de mogelijke toepassing van Master Data Management binnen Pepsico.

Het doel is, om het fenomeen MDM te onderzoeken en mogelijk bruikbare methoden te destilleren voor intern gebruik.

Dit rapport wordt geschreven voor het afstuderen aan de HU en voor kennisvergaring van de afdeling Data Management Organization en IT van Pepsico.

Allereerst wordt de aanleiding tot de opdracht beschreven. Vervolgens komt de plaats in de organisatie aan bod (context) en wordt een probleemstelling geponeerd. De opdracht wordt beschreven voor de twee afdelingen met projectgrenzen. Uit benoemde hoofd- en deelvragen volgen activiteiten en de daaruit voortvloeiende eind- en deelproducten.

Methoden / technieken en kosten / baten worden kort toegelicht, waarna dieper wordt ingegaan op de organisatie, risico's en planning van het project.

1 Aanleiding tot de opdracht

Het aantal artikelgegevens bij Pepsico is explosief gestegen ten opzichte van circa twintig jaar geleden. Dit heeft een aantal oorzaken.

1. De assortimenten worden steeds groter. Een supermarkt heeft al snel een assortiment van vele duizenden producten.
2. De life cycle van producten is veel korter dan voorheen. De meeste fabrikanten leveren hetzelfde product niet langer dan een jaar of zelfs een paar maanden (computer hardware). Daarna komt er een opvolgend product.
3. Het aantal gegevens dat per artikel geregistreerd wordt neemt toe, vooral gedreven door de behoefte van consumenten aan meer informatie. Honderd velden per product is heel normaal. Het soort velden verschilt daarbij weer per type product.
4. De standaardisatie-eisen nemen toe, omdat producten ook op de specificaties vergeleken moeten kunnen worden in product vergelijkers (websites) per productcategorie.
5. Aangezien in veel indirecte verkoopkanalen de marges zwaar onder druk staan, neemt ook de behoefte aan kostenefficiëntie en standaardisatie van catalogus uitwisselingsformaten toe (EDI, Electronic Data Interchange).
6. Toename van het aantal gebruikers, ook vanuit de social media (vragen van consumenten)
7. Introductie van allergenen declaratie (R&D)

Om bovenstaande goed te kunnen beheren is een ondersteunende tool, zoals PDM (zie bijlage 3), een noodzaak.

Supply chain

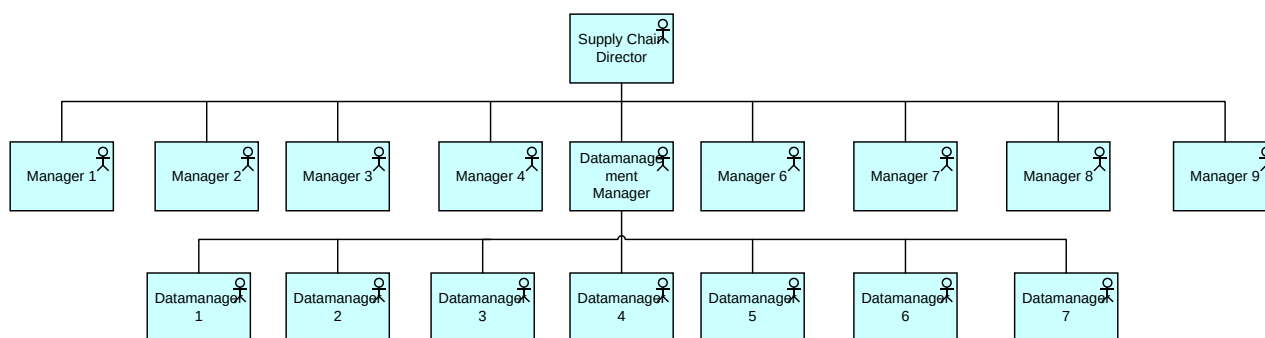
Als het bovenstaande eenmaal ingericht is kan deze informatie, die door de fabrikant beheerd wordt, doorgegeven worden aan de afnemers. Alle gegevens van alle artikelen, die steeds weer wisselen, moeten door de hele supply chain gecommuniceerd worden. Omdat het aantal gegevens zo groot is, is een fout snel gemaakt. Fouten in deze communicatie kosten relatief veel geld. Daarom wordt steeds meer gekozen voor het elektronisch communiceren, bijvoorbeeld met behulp van het PRICAT-bericht (prijs catalogus) dat binnen de EDI standaard gedefinieerd wordt. In Europa wordt PRICAT onder andere toegepast voor levensmiddelen en bruin- en witgoed. Voor de levensmiddelenbranche is Albert Heijn de initiatiefnemer geweest (GS1, Global Standards organisatie, bekend van de streepjescode).

2 Beknopte context waarbinnen het afstudeerproject wordt uitgevoerd

PepsiCo is één van de grootste voedingsmiddelenbedrijven ter wereld met merken in hartige tussendoortjes, ontbijtgranen en dranken en met 285.000 medewerkers met activiteiten in meer dan 200 landen. PepsiCo Nederland maakt onderdeel uit van PepsiCo Europa.

De afdeling DataManagement Organization (DMO) is onderdeel van de supply chain afdeling, welke de gegevens van alle productie eenheden in Nederland en België verzamelt en beheert.

Deze gegevens betreffen informatie over leveranciers, materialen, productie, klanten en prijzen en worden in verschillende systemen opgeslagen en bijgehouden.



Figuur 1 Organogram supply chain afdeling

Positie student binnen de organisatie

Ik werk als één van de zeven Datamanagers binnen de afdeling DMO.

Deze functie omvat het verwerken en beheren van alle supply chain data in de verschillende systemen, zoals SAP en Oracle. Mijn bedrijfsbegeleider Piet Mooij is Data Management manager.

3 Probleemstelling

Vanwege de recente nadruk op naleving van de regelgeving, maar ook door fusies en overnames, is het voor PepsiCo een noodzaak om accurate en complete Master Data te creëren en te onderhouden.

Omdat er nog diverse legacy systemen op diverse locaties actief zijn, met vaak dezelfde gegevens, is het gevaar aanwezig, dat die gegevens niet consistent zijn. Er moeten hiervoor overkoepelende data-onderhoud en governance processen en procedures ontwikkeld worden om “wildgroei” te voorkomen.

4 Opdrachtschrijving

Een middel om goede controle over Master Data te hebben is Master Data Management (MDM).

Een definitie van Master Data Management (MDM) is: de benodigde technologie, tools en processen om consistente en accurate master data te creëren en te onderhouden.

MDM is niet alleen een technologisch probleem. In veel gevallen zijn fundamentele veranderingen van bedrijfsprocessen nodig om zuivere master data te behouden en vaak zijn de lastigste MDM problemen meer van politieke aard dan technisch.

Met deze afstudeeropdracht ga ik onderzoeken, op welke manier MDM van toepassing kan zijn voor twee afdelingen binnen PepsiCo: DataManagement Organization (DMO) en IT.

DMO

Voor DMO gaat het om het binnenhalen van kennis van MDM om te inventariseren welke toepassingen er in de toekomst van belang kunnen zijn.

Hiervoor moet eerst kennis en inzicht worden verkregen van MDM: wat is het, wat houdt het in, wat zijn de huidige best practices?

Vervolgens wordt het huidige volwassenheidsniveau van Pepsico Nederland bepaald op het gebied van Data Governance. Waar staan we nu op het gebied van Data standaarden, Metadata management, Data kwaliteit en Data validatie, de vier deelgebieden van Data Governance.

Later wordt in overleg bepaald naar welk volwassenheidsniveau we willen groeien en op basis daarvan bekeken wat Pepsico moet doen om dat niveau te behalen (Gap analyse).

De criteria worden op basis van voortschrijdend inzicht bepaald.

IT

Ten behoeve van de informatievoorziening naar hun klanten, wil Pepsico productgegevens van de Business Unit “Dranken” aan hen verstrekken.

Deze productinformatie staat in het systeem Oracle, maar is nog niet beschikbaar voor klanten.

De bedoeling is, om die informatie automatisch door te sturen naar een zogenaamde datapool, waarin deze artikelgegevens centraal beschikbaar zijn en beheerd worden.

Het systeem dat Pepsico daarvoor gebruikt heet PDM, Product Data Management, zie bijlage 3.

De gegevens in PDM worden binnen Pepsico beheerd en worden elektronisch naar de externe datapool (1Sync) gestuurd.

Voor deze afstudeeropdracht gaat het om het beschrijven van een pilot om MDM te gebruiken bij het interfacen van data vanuit Oracle naar PDM met behulp van BO Data Services, zie bijlage 4.

Van het IT project ben ik zelf de projectleider en ga ik het project sturen aan de hand van PLM, Project Lifecycle Management, zie bijlage 5.

In de IT wereld zijn er vele Project Management methoden bekend, waarbij Prince2 en PLM de bekendste zijn. Prince2 is van Engelse origine en wordt veel in West-Europa toegepast. PLM is de Amerikaanse variant en wordt veel in Noord-Amerika gebruikt. Pepsico hanteert als Amerikaans bedrijf binnen de organisatie de PLM methode en daarom zal ik deze ook toepassen op dit project.

5 Projectgrenzen

Wat is in scope:

De volgende zaken worden opgeleverd:

- Mapping van de data van Oracle naar PDM
- Interface design, met toepassing van BO data services regels & mogelijkheden
- Beschrijving van de bouw van de interface, test en implementatie. Indien er voldoende tijd beschikbaar is, wordt deze interface ook gebouwd en geïmplementeerd
- Beschrijving van het support model, alleen een voorstel van hoe te implementeren

Systemen:

Oracle, BO Data Services, PDM

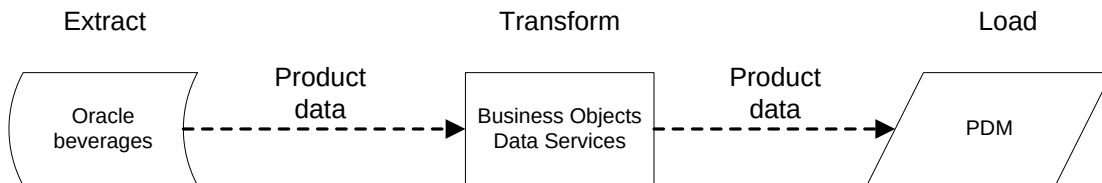
Data:

- Beverage product data, inclusief de klanten waar naartoe gepubliceerd wordt
 - Productnaam
 - Producthierarchy: brand, subbrand, smaak
 - EAN code
 - Units of Measure (UoM): Maten (hoogte, breedte, diepte), Gewicht (Bruto, Netto)

- Verpakking: Vorm (doos, fles, pak), Aantal CE's per HE (flessen per krat)
- Palletisering: aantal kratten per laag, aantal lagen per pallet

Proces:

- Oracle data extractie:



Wat is niet in scope:

Het volgende wordt niet opgeleverd:

- Support model, implementatie (middelen, set up tool, medewerkers)

Systemen:

SAP, AS400, Genxt Masterfiles, Manugistics, 1Sync

Data:

Vendors, prijzen, gebruikers en klanten

Processen:

- Gegevensuitwisseling met retail (1Sync)
- SAP data extractie:



6 Doelstellingen

Na afloop van het afstudeerproject moet het volgende zijn bereikt:

- De afdeling DataManagement beschikt over een beschrijving van de huidige stand van MDM en wat dit kan betekenen voor de afdeling/organisatie. Ook welke onderdelen kunnen worden gebruikt en hoe die kunnen worden toegepast.
- De afdeling IT beschikt over een beschrijving van de te bouwen interface, welke artikelinformatie van Oracle beverages naar PDM stuurt.

7 Hoofdvraag

Wat zijn goede methoden om correcte data aan de klant aan te bieden?

Afgeleide vragen hierbij zijn:

4. Welke methoden van databewerking en controle bestaan er?
5. Welke methoden gebruikt Pepsico nu?
6. Welke eisen worden gesteld aan Data Governance?
7. Welke methoden zijn het meest geschikt voor Pepsico en waarom?
8. Hoe weet ik dat de aangeboden data juist zijn?

8 Activiteiten

Om deze vragen te beantwoorden, zijn de volgende acties nodig:

6. Onderzoeken wat MDM inhoudt en welke onderliggende methoden gebruikt worden.
7. Onderzoeken welke MDM methoden momenteel binnen de afdeling DMO gebruikt worden.
8. Onderzoeken welke eisen worden gesteld aan Data Governance.
9. Een voorstel maken van welke methoden bij Pepsico toe te passen zijn en hoe.
10. Onderzoeken, aan de hand van dit project, hoe MDM methoden tot de juiste data kunnen leiden.

9 Eindproduct en deelproducten

Het eindproduct is een rapport dat bestaat uit twee delen:

DMO

Een beschrijving van de huidige stand van MDM en welke onderliggende methoden bruikbaar zijn voor Pepsico en in het bijzonder de afdeling DMO. Ook wordt ingegaan op hoe die toegepast kunnen worden.

Dit omvat een:

5. Inventarisatie van MDM
6. Inventarisatie van momenteel gebruikte methoden bij de afdeling DMO
7. Eisenpakket van Pepsico
8. Voorstel van mogelijk toe te passen methoden

IT

Een beschrijving van de te bouwen interface, welke artikelinformatie van Oracle beverages naar PDM stuurt. Daarbij wordt onderzocht, welke MDM toepassingen kunnen bijdragen aan juiste data (deelproduct 5).

10 Methoden en technieken

Voor het DMO gedeelte wordt een literatuurstudie uitgevoerd en van internetsites actuele informatie verzameld, zoals white papers en artikelen, zie “bronnen”.

Voor het IT gedeelte wordt Business Objects Data Services gebruikt van SAP, zie bijlage 4. Dit is een conversie / interface tool, welke door Pepsico gebruikt wordt.

Voor project management wordt PLM gebruikt, zie bijlage 5.

11 Kosten en baten

Kosten

Voor het beschrijven van het interface design zijn de volgende kosten verbonden:

De project manager besteedt een dag in de week aan dit project gedurende 17 weken.

Aantal weken	Tijd (uren)	Uurtarief (€)	Kosten (€)
17	8	30	4080

De drie interne experts zullen elk een dag nodig hebben om informatie aan te dragen, $3 * € 300,- = € 900,-$.

In totaal komt dat neer op € 5000,-.

Er moet aan de hand van de mapping onderzocht worden, wat de bouw van de interface zal gaan kosten.

Baten

Een eindrapport, welke enerzijds voldoende is om af te studeren aan de HU en anderzijds Pepsico een goed beeld geeft van de huidige stand van MDM.

Als de interface gebouwd is, is handmatige publicatie niet meer nodig.

Dit resulteert in de volgende kostenbesparing:

Artikelen	Aantal per jaar	Tijd (min)	Uurtarief (€)	Kosten (€)
Nieuw	250	15	30	1875
Wijziging	100	2	30	100
Totaal				1975

Het gaat echter bij dit project niet direct over een financieel voordeel. Het gaat Pepsico er voornamelijk om, dat die productinformatie beschikbaar komt voor de klant als een extra service.

12 Organisatie

Bedrijfs- en persoonsgegevens

Pepsico Nederland
Zonnebaan 35
3542 EB Utrecht
030-2473811

Opdrachtnemer, student HU:
Henk van Beek
030-2473548
henk.vanbeek@pepsico.com

Opdrachtgever, leidinggevende en begeleider:
Piet Mooij
030-2473791
piet.mooij@pepsico.com

Begeleider IT:

Irene uit het Broek
030-2473785
Irene.uitthetbroek@pepsico.com

Docent begeleider HU:
Jos van Dongen
06-24213447
jj.van.dongen@planet.nl

In onderstaande tabel zijn alle betrokkenen in kaart gebracht met hun invloed en bijdrage.

Stakeholder Name	Organization	Role	Degree Impacted	Reaction to Change in the Past	Power & Influence	Level of Interest	Level of Management Needed?	Support Needed?
Piet Mooij	Pepsico	Sponsor, project companion	High	Supporter	8	8	Required attention / Manage closely	Necessary
Raymond van der Meer	Pepsico	Sponsor	High	Neutral	8	8	Required attention / Manage closely	Necessary
Tjalling Dijkstra	Pepsico	Sponsor	High	Neutral	8	8	Required attention / Manage closely	Necessary
Henk van Beek	Pepsico	Project manager, student	High	Supporter	8	10	Required attention / Manage closely	Necessary
Dirk Goffin	Pepsico	Data expert Oracle	Medium	Supporter	6	5	Leverage / Keep satisfied	Necessary
Jan Cumps	Pepsico	Data expert BO Data Services	Medium	Supporter	6	5	Leverage / Keep satisfied	Necessary
Carla Liekens	Pepsico	Data expert PDM	Medium	Supporter	6	5	Leverage / Keep satisfied	Necessary
Irene uit het Broek	Pepsico	Project companion	Medium	Supporter	6	5	Leverage / Keep satisfied	Desirable
Jos van Dongen	Hogeschool Utrecht	Teacher companion	Medium	Supporter	6	8	Required attention / Manage closely	Necessary

Tabel 1 Betrokkenen bij het MDM project

13 Risico's

De risico's zijn in onderstaande tabel opgesomd.

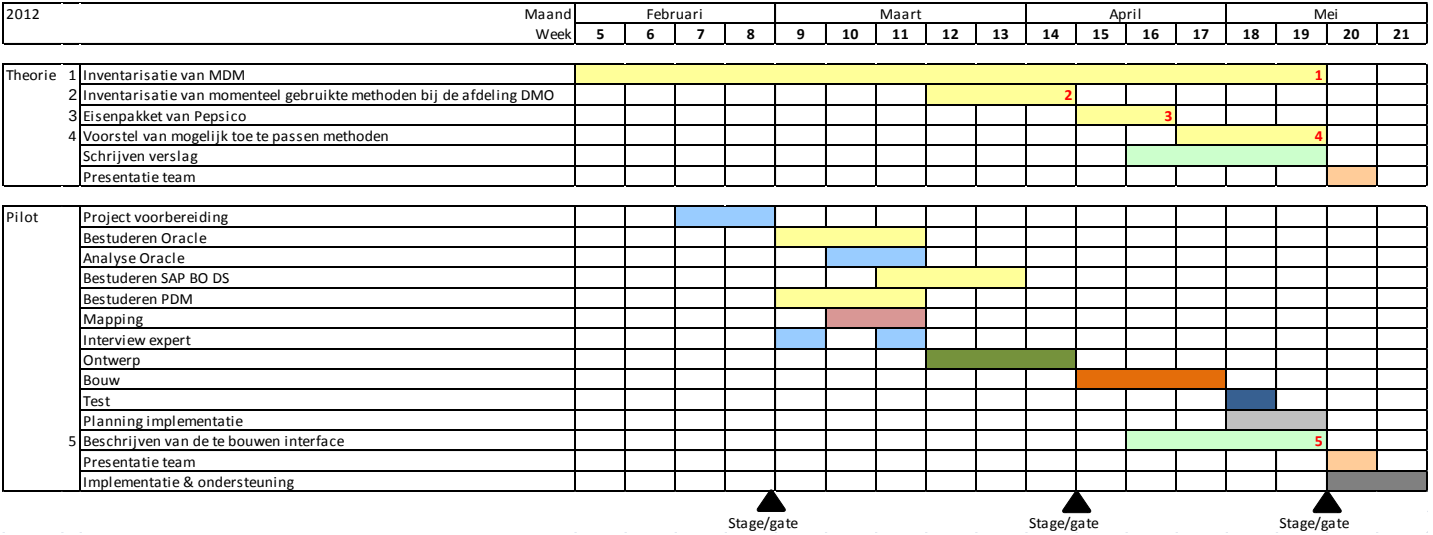
Date Identified	Risk No.	Risk Description	Category	Potential Impact Description	Likelihood of Occurrence	Impact	Likelihood Risk Value	Impact Risk Value	Risk Level	Risk Response
02-16-12	1	No access to applications	Resource Risk	Delay of the project	Low	High	1	3	4	Avoid
02-16-12	2	No support from experts	Resource Risk	Delay of the project	Low	High	1	3	4	Transfer
02-16-12	3	No support from sponsors	Milestone Risk	Delay of the project	Low	Medium	1	2	3	Transfer
02-16-12	4	Limited knowledge and experience	Requirements Risk	Delay of the project	Medium	High	2	3	5	Transfer
02-28-12	5	No budget available?	Budget Risk	Delay of the project	Low	Low	1	1	2	Mitigate

Tabel 2 Mogelijke risico's bij het MDM project

1. Geen toegang tot applicaties: voorkomen door tijdig met de relevante belanghebbenden te communiceren en deze te informeren
2. Geen ondersteuning van experts: doorzetten naar sponsor
3. Geen ondersteuning van sponsors: regelmatig en duidelijk communiceren over de consequenties voor het project
4. Beperkte kennis van en ervaring met BODS: doorzetten naar en/of bespreken met sponsor
5. Geen budget beschikbaar?: duidelijk communiceren over de consequenties voor het project

14 Planning

De projectplanning ziet er als volgt uit.



Tabel 3 Projectplanning

De (rode) cijfers in dit schema geven de deelproducten aan, zoals ook beschreven in hoofdstuk 9.

- 1^e stage/gate: 1 Maart, start Project voorbereiding
- 2^e stage/gate: 9 April, start bouw
- 3^e stage/gate: 14 Mei, Go Live

15 Kwaliteitsbewaking

Elke vrijdagochtend wordt er overleg gepleegd tussen de project manager en de projectbegeleider(s) over de huidige status en bevindingen.

Bronvermeldingen

Literatuur:

Master Data Management, David Loshin, 2009

Master Data Management en Data Governance, Alex Berson, Larry Dubov, 2011

Manuals:

SAP BusinessObjects Data Services, pdf presentatie van Steve Hill, OEM Solutions Engineer, 24 juni 2009

Intranet:

International Project Life-cycle Methodology Handbook (KnowledgeNet)

Internet:

SAP

<http://www.sdn.sap.com/irj/boc/index?rid=/webcontent/uuid/f0b79c51-ff38-2c10-eb8b-8bda91617662>

Gartner

<http://www.gartner.com/technology/research.jsp>

Forrester

http://www.forrester.com/rb/search/results.jsp?N=0+50071&cm_re=Navigation_010710-_-business_data_subnav-_-business_datadriven_research

InformationAge

www.information-age.com/

Bijlage 1 PDM

Product Data Management (PDM) wordt in bedrijfssoftware gebezigd waar het beheer van artikelgegevens centraal staat.

Doel van deze software

Het aantal artikelgegevens is explosief gestegen ten opzichte van, circa, twintig jaar geleden. Dit heeft een aantal oorzaken.

- 1 De assortimenten worden steeds groter. Een supermarkt heeft al snel een assortiment van vele duizenden producten.
- 2 De life cycle van producten is veel korter dan voorheen. De meeste fabrikanten leveren hetzelfde product niet langer dan een jaar of zelfs een paar maanden (computer hardware). Daarna komt er een opvolgend product.
- 3 Het aantal gegevens dat per artikel geregistreerd wordt neemt toe, vooral gedreven door de behoefte van consumenten aan meer informatie. Honderd velden per product is heel normaal. Het soort velden verschilt daarbij weer per type product.
- 4 De standaardisatie-eisen nemen toe, omdat producten ook op de specificaties vergeleken moeten kunnen worden in product vergelijkers (websites) per productcategorie.
- 5 Aangezien in veel indirecte verkoopkanalen de marges zwaar onder druk staan, neemt ook de behoefte aan kostenefficiëntie en standaardisatie van catalogus uitwisselingsformaten toe.

Om bovenstaande goed te kunnen beheren is een ondersteunende tool, zoals bijvoorbeeld PDM, een noodzaak.

Supply chain

Als het bovenstaande eenmaal ingericht is kan deze informatie, die door de fabrikant beheerd wordt, doorgegeven worden aan de afnemers. Alle gegevens van alle artikelen, die steeds weer wisselen, moeten door de hele supply chain gecommuniceerd worden. Omdat het aantal gegevens zo groot is, is een fout snel gemaakt. Fouten in deze communicatie kosten relatief veel geld. Daarom wordt steeds meer gekozen voor het elektronisch communiceren, bijvoorbeeld met behulp van het PRICAT-bericht dat binnen de [EDI](#)-standaard gedefinieerd wordt. In Europa wordt PRICAT onder andere toegepast voor levensmiddelen en bruin- en witgoed. Voor de levensmiddelen is Albert Heijn de initiatiefnemer geweest.

Bijlage 2 BusinessObjects Data Services

BusinessObjects Data Services (BODS) is een complex platform voor het data integratie proces en maakt deel uit van de laatste versie van de Business Objects applicatie.

Het SAP BODS platform heeft een modulaire structuur en bestaat uit de volgende drie tools:
een Data Services Designer, een Management Console en een centrale opslag.

Data Services Designer

Het is een functionele, onafhankelijke “client” applicatie die het mogelijk maakt om data workflows te monitoren, welke zijn geclassificeerd in taken. Data Designer helpt verbindingen te definiëren met de diverse data bronnen. De BODS applicatie biedt een aantal voorgedefinieerde transformaties en functionele objecten, waarmee je ETL* flows kunt modelleren.

Voorbeelden van objecten zijn: data merging , kolom mapping , SQL queries, validatie van records en vergelijking van verschillende tabellen.

Het is ook mogelijk om speciale functies te programmeren met behulp van een script taal.

Elke data flow die al gedefinieerd en geladen is in de database kan herhaaldelijk en continu gebruikt worden. Het gebruik van Data Designer geeft de mogelijkheid om de executies van de data flows te ontwerpen en te monitoren.

Management Console

Deze applicatie is beschikbaar in de internet browser (Internet Explorer en Firefox) en bestaat uit de volgende functionele modules:

Administrator

Deze module is ontworpen voor het ETL platform, het applicatie repository management en voor inplannen en monitoren van ETL taken. Elke taak, die door de ETL server wordt verwerkt, genereert een log file welke voor een bepaalde periode beschikbaar is. De administrator maakt het mogelijk om de getransformeerde taken te benaderen en de details te controleren, zoals de start en stop tijd van de taak, executie tijd en de status van de data.

Het monitoring proces kan onjuistheid van sommige taken (op log-niveau) laten zien, wat specifieke rapporten van deze fouten oplevert (van de ETL server of database). Er is ook de mogelijkheid om je eigen rapporten te definiëren m.b.v. de script taal.

Auto Documentatie

Deze module genereert de ETL projectdocumenten (rft/pdf) met het vereiste niveau van specificaties.

Data validatie

Dit omvat data validatie op basis van de gedefinieerde regels en inschatting en identificatie van de potentiële fouten of inconsistenties van de data bron.

Impact analyse en Data Lineage Analysis

Deze module integreert metadata van de SAP BO platformen van de bron naar het rapport. Deze kan een gegeven rapport element terug herleiden naar het bronsysteem en uitzoeken welke transformaties zijn uitgevoerd om deze waarde te bepalen.

Operationele Dashboard

Deze module bevat statistieken en historie van de ETL server taken als functioneel dashboard.

Data Kwaliteit Rapporten

Met deze module kun je de voorgedefinieerde set van rapporten beoordelen.

Centrale opslag

BODS oplossingen zijn gebouwd, gebaseerd op een centrale database, welke onafhankelijk is van de lokale databases van de ETL ontwikkelaars.

***ETL**

ETL is de afkorting voor Extraction, Transformation and Load. De term ETL staat voor een groep technologieën die veelal gebruikt worden bij de koppeling tussen systemen, waarbij er gestreefd wordt naar een minimale technische en semantische koppeling tussen de systemen. Het is een batchproces dat regelmatig gebruikt wordt.

- Extract = leest data van een bron en pakt het gewenste pakket uit
- Transform = Zet opgenomen data om, gebruik makende van regels, opzoektabelen of maakt combinaties van data van verschillende bronnen
- Load = schrijft de data naar een gewenst doel

Bijlage 3 PLM, International Project Life-cycle Methodology

International Project Life-Cycle Methodology is the standard project development and implementation methodology for PepsiCo International (PI). The methodology will reside on KnowledgeNet-International folder. It will help you plan the phases of the project life cycle and properly align with the various teams within the organization to deliver a successful, high quality product to the business.

As a project manager or project team member, you will need to tailor the methodology for the requirements of each project. However, all projects will follow the general outline described here, using a consistent approach and common terminology that is independent of the project, and following prescribed Stage Gates to move from one phase to the next. Clear communications from the project manager are critical to a project's success. Clear communications help keep the customer's expectations under control and ensure that the project team members understand their responsibilities. KnowledgeNet-International folder will be your source for project templates, guidelines, techniques, checklists, examples, and other accelerators that will help you create the deliverables required in each phase. The methodology also discusses the tools and methods to use. Project roles and responsibilities are clearly defined for key project teams. Those teams will help manage deliverables and obtain commitments from the various teams involved.

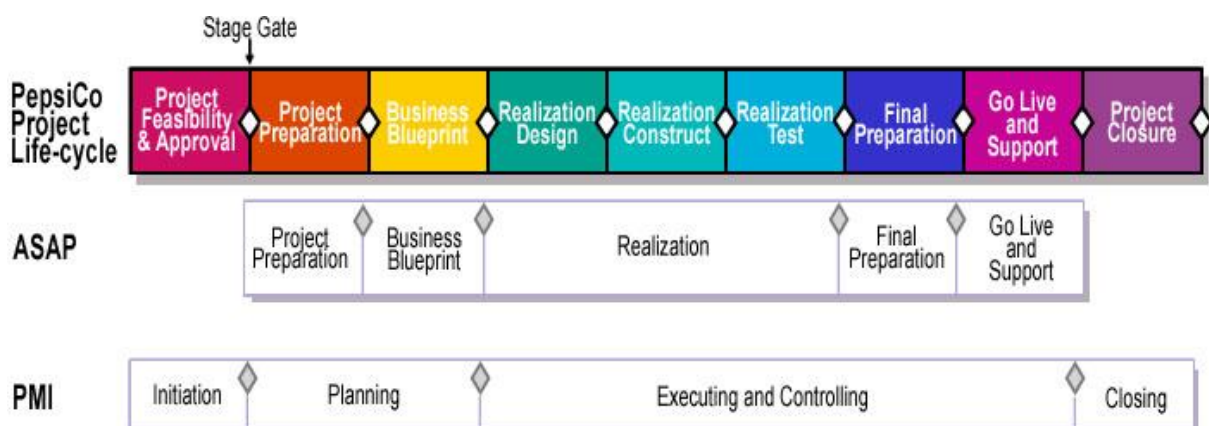
The handbook incorporates the Project Management Institute (PMI®) standards and includes techniques and many of the current best approaches identified by project managers and team members. The intent is to provide the processes necessary for a project manager to manage any project, regardless of size. The specific section of the handbook for each phase identifies the key deliverables required for different classifications of projects. Project classifications are determined by analyzing multiple project characteristics, such as duration of the project, budget, risk, and business value for the project.

Project Life-cycle Phases

The International Project Life-Cycle Methodology organizes the approach to executing a project into nine phases:

- Project Feasibility and Approval
- Project Preparation
- Business Blueprint
- Realization Design
- Realization Construct
- Realization Test
- Final Preparation
- Go Live and Support
- Project Closure

The following diagram shows each of the phases in the order in which they are executed:



The nine phases encompass the entire process of delivering a system/project, from the moment a need is identified until the project is completed and the solution is implemented. Each phase is associated with the current state of a project's evolution. The project team must complete the deliverables associated with each phase before the project can progress to the next one. Once completed, there is a review of the required activities, deliverables, and exit criteria. They must be

reviewed, accepted, and approved before the phase is considered complete. The approval is required for authorization to proceed to the next phase. This review and approval process is referred to as a Stage Gate; to move to the next stage, the project team must meet the specified criteria to start work on the next phase.

A Stage Gate meeting marks the transition from one phase to the next. This meeting helps ensure that the project meets the documented criteria and that the solution can function effectively in the BIS architecture. A Stage Gate meeting is required at the end of each project phase unless otherwise decided with the Steering Committee during Project Preparation.

In early phases, most activities involve planning, defining requirements, and system design. The project team captures these business needs, requirements, and specifications in various documents that are deliverable at the end of each phase. The definitions progressively evolve from high-level concepts to precise statements describing those needs, requirements, and specifications. The accelerators provided in KnowledgeNet help ensure consistent documentation that meets the criteria for “acceptable deliverables” during the Stage Gate meeting.

In later phases, activities focus on building and testing the system and establishing the production environment.

Deliverables from early phases provide the requirements and design specifications. Quality is assured by writing good, complete, and clear specifications, developing quality programs in a solid technical environment, and by complete testing on the system until it is ready for a production environment. The specification, development, and system testing are also “deliverables” during a Stage Gate meeting.

Bijlage 2 Evaluatie van het eigen functioneren

Praktijk

In het praktijk gedeelte van deze opdracht had ik de rol van projectleider.

Dat houdt dus in, dat ik voor het ontwerpen van de BODS interface de initiator en aanspreekpunt ben geweest. Ik heb daarvoor veel contact gehad met systeemexperts, met de interface bouwer, maar ook met de sponsors en begeleiders van dit project.

Dit verliep over het algemeen prima. Er is regelmatig contact geweest met de betrokkenen, persoonlijk, via mail, telefoon en conference call.

Leermoment

In het beginstadium van het project zou de interface gebouwd worden door een externe partij.

Daarvoor zou ik een afspraak plannen met dat bedrijf en diverse interne deelnemers.

Ik heb toen te lang vastgehouden aan een voorgestelde datum van dat bedrijf (extern) en gezocht naar mogelijkheden om de interne deelnemers daar naar te schikken.

Dat bleek de verkeerde volgorde. Het is erg lastig om agenda's af te stemmen voor een bepaalde datum en tijd, dus in het vervolg zal ik eerst intern een datum en tijd plannen om vervolgens de externe partij te benaderen.

Theorie

Op het gebied van rapportgeving over de status van het rapport, heb ik vrij weinig contactmomenten gehad met de bedrijfsbegeleider en docent begeleider. Dat had te maken met de lange aanloop van het bestuderen van de literatuur. Allereerst was ik begonnen om het praktijkproject van de grond te tillen, om daarna de starten met de literatuurstudie. Vervolgens bleek het vertalen van de uitgebreide theorie van MDM erg lastig. Zodoende duurde het vrij lang, voordat ik iets kon presenteren. Ik wilde overigens pas iets tonen als het al enige vorm aan had genomen.

Leermoment

Het blijkt handiger om tijdens een project regelmatig tussenstanden te bespreken, ook al zijn het kleine brokken. Het is dan beter mogelijk om gaandeweg bij te sturen en de betrokkenen hebben een beter beeld van de huidige status en vorderingen.

Bijlage 3 Wat is MDM?

In deze bijlage wordt in meer detail ingegaan op Master Data Management (MDM) en de aspecten die daarbij een rol spelen.

Deze bijlage is opgebouwd uit een veertiental paragrafen. Hieronder staat een opsomming van de onderwerpen die daarin aan bod komen.

Paragraaf:

- 1 Inleiding
Wat is master data en MDM?
- 2 Coördinatie: belanghebbenden, eisen en planning
Wie hebben belang bij MDM en wat is hun deelname daarin?
- 3 MDM componenten en het volwassenheidsmodel
Op basis van de huidige status en de gewenste status van volwassenheid van de organisatie op het gebied van verspreiden van geconsolideerde informatie, kan een roadmap opgesteld worden om de doelstellingen te halen.
- 4 Data governance
Hoe wordt bedrijfsbeleid gevormd uit informatie richtlijnen en hoe dragen regels bij aan conformiteit aan die richtlijnen?
- 5 Data kwaliteit
Geeft een globaal overzicht van data kwaliteitscomponenten en -methoden voor data integratie.
- 6 Metadata management
Dit gaat over data standaarden en hoe metadata wordt beheerd.
- 7 Identificeren van Master data en Metadata
Het proces van identificatie en vinden van data, welke de bron vormen voor master data.
- 8 Data modellering
Aspecten m.b.t. het ontwikkelen van modellen.
- 9 Architectuur
Dit gaat over bestaande applicatie- en informatie architectuur.
- 10 Consolidatie en integratie
Verzamelen van informatie uit de hele organisatie en omvormen naar geïntegreerde master data.
- 11 Synchronisatie
Behoeften en benaderingen voor het synchroniseren van data terug naar de bestaande applicaties (datastuwende bronnen).
- 12 Functionele service laag
Het gebruik van een service laag, welke meerdere applicatie benaderingen toevoegt, geeft meerwaarde aan bestaande en toekomstige applicaties.
- 13 Management richtlijn
Enkele richtlijnen voor het ontwikkelen van een business case.
- 14 Trends en richtingen
Trends in de markt en in technische mogelijkheden

1 Inleiding

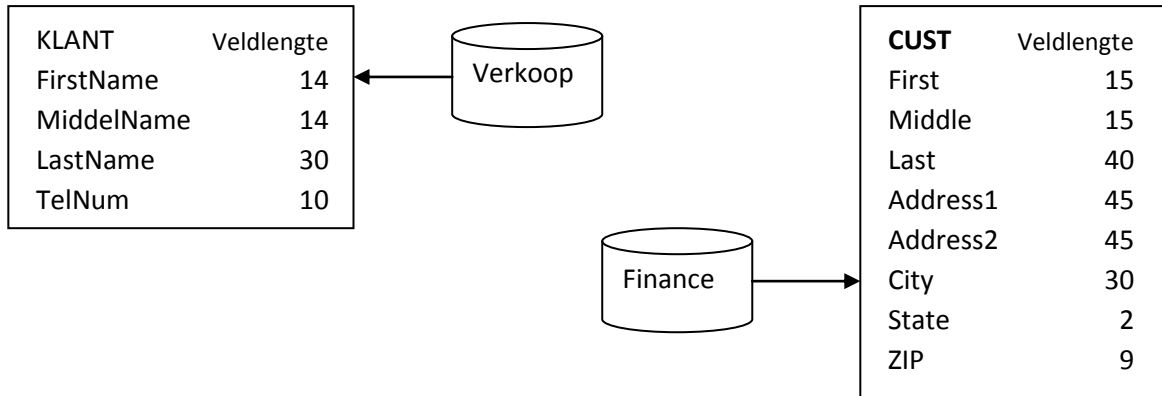
In deze bijlage wordt er voor het begrip “gegevens” steeds de term “data” gehanteerd, aangezien (master) data management een internationaal gebruikte term is en voor vele bedrijven als een algemeen bekende term wordt gehanteerd.

Onze wereld wordt op verschillende manieren steeds meer samenhangend en de hoeveelheid informatie die we in handen krijgen lijkt geëxplodeerd.

Flexibele organisaties die zich kunnen aanpassen aan een vloed van redundante en soms conflicterende gegevens, merken dat de combinatie van informatie uitwisseling en operationele samenwerking een onderscheidende factor is voor organisatorisch succes.

Aangezien partnerbedrijven meer informatie delen en uitwisselen, neemt verbondenheid ook toe. Het zijn niet alleen systemen die beter samenwerken, maar ook mensen die de systemen beheren vormen betere werkrelaties, hetgeen leidt tot effectievere bedrijfsleiding en uiteindelijk tot concurrentievoordeel.

In het verleden werden applicaties ontwikkeld om in operationele behoeften te voorzien op een specifiek gebied. De applicatie architectuur bood ondersteuning in die deelgebieden. Dat gold ook voor de afgeleide IT ondersteuning met data definities, tabellen, structuren, functionaliteiten etc.. Verschillende applicaties kunnen echter refereren naar meerdere datasets met hetzelfde bedrijfsconcept, maar dezelfde namen kunnen ook refereren naar verschillende concepten, zie figuur 1. Om tot een organisatiebrede gegevensuitwisseling te komen, zal eerst gewerkt moeten worden aan een consistent beeld van alle data uit de verschillende databronnen.



Figuur 1 Verschillende applicaties kunnen refereren naar dezelfde bedrijfsconcepten

Om dit te realiseren moet een organisatie in staat zijn om de bedrijfsconcepten helder te definiëren, de verschillende manieren te identificeren waarop data sets een bedrijfsconcept vertegenwoordigen, de data integreren in een consistent beeld en vervolgens het consistente beeld beschikbaar maken binnen de gehele organisatie.

Dit is een enorme uitdaging voor integratie van bedrijfsinformatie, management en uitwisseling om een echt geïntegreerde organisatie te vormen.

De essentie van de MDM uitdaging is om een bedrijfsbreed overzicht te organiseren van de belangrijkste informatieobjecten en om de kwaliteit, het gebruik en synchronisatie te waarborgen, daarmee het gebruik van informatie te optimaliseren en zo de operationele en strategische doelstellingen van de organisatie te bereiken.

De toename van zowel kracht als functionaliteit van computersystemen heeft een fijnmaziger dataverspreiding veroorzaakt. Dit staat een grotere vrijheid in beschrijving en modellering van bedrijfsinformatie toe. Of het nu in database servers zit of in Excel bestanden, er ontstaat een warboel aan concepten, naast creatieve manieren om die concepten toe te passen.

Dit is de eerste drijfveer voor MDM: de mogelijkheid tot rationaliseren van definities en betekenissen van veel gebruikte bedrijfsconcepten en de mogelijkheid om verschillende bedrijfstermen te differentiëren als ze niet refereren naar hetzelfde concept.

1.1 Wat wordt verstaan onder Master data?

Het verkrijgen van “één versie van de waarheid” en het beheren van de kwaliteit van gegevens, is een probleem voor bijna elke organisatie.

Master data zijn die entiteiten, relaties en attributen, welke kritisch zijn voor een onderneming en fundamenteel voor de belangrijkste bedrijfsprocessen en applicaties.

Master data wordt gedefinieerd door master data objecten.

Master data objecten zijn de kern bedrijfsobjecten, welke in verschillende applicaties in de organisatie gebruikt worden met de bijbehorende metadata, attributen, definities, rollen, verbindingen en taxonomieën (manier waarop bedrijfsconcepten worden beschreven, gekarakteriseerd en geclassificeerd). In figuur 2 staat een voorbeeld van master data objecten.

Master data objecten zijn de sleutel objecten die er het meest toe doen, de zaken die worden vastgelegd in transactionele systemen, gemeten en gerapporteerd in rapportagesystemen en geanalyseerd in analytische systemen. Het gaat dus om klanten, medewerkers, leveranciers, onderdelen, producten, locaties, contracten, etc.

Master Data Object	Waarde
Klant	Jumbo
Product	Cruesli 12x500g
Prijs (doos)	€ 25,00
Locatie	Utrecht

Figuur 2 Master Data Elementen

Een master data systeem, dat bestaat uit een master data set, is een (potentieel virtuele) opslag of index van uniek geïdentificeerde entiteiten met hun kritische data attributen, gesynchroniseerd vanuit de originele databronnen en beschikbaar gemaakt voor bedrijfsbreed gebruik.

Met geschikte controle en overzicht, kunnen de data in het master data systeem (of opslag) gekwalificeerd worden als een eendrachtig en coherente databron, waar alle applicaties op kunnen vertrouwen voor consistente informatie van hoge kwaliteit.

1.2 Wat is MDM?

MDM is een verzameling van methoden, welke belanghebbenden, deelnemers en interne klanten leiden in het opnemen van bedrijfsapplicaties, informatie management methoden en Data Management tools om het beleid, procedures, service en infrastructuur uit te voeren ter ondersteuning van het laden, integreren en gedeeld gebruik van accurate, tijdige, consistente en complete master data.

Met andere woorden, een MDM programma is bedoeld ter ondersteuning van bedrijfsbehoeften door het leveren van toegang tot consistente overzichten van uniek te identificeren master data entiteiten door de gehele operationele applicatie infrastructuur.

MDM leidt de methoden, tools, informatie en service om onderstaande punten te bewerkstelligen:

- Beoordeel het gebruik van veel gebruikte informatie objecten, verzamelingen van valide datawaarden en expliciet en impliciet business rules in de reeks applicaties binnen het bedrijf.
- Identificeer kern informatie objecten, welke relevant zijn voor het succes van de onderneming en in verschillende applicatie datasets worden gebruikt, welke kunnen profiteren van centralisatie.
- Bekrachtig een gestandaardiseerd model voor integratie en beheer van sleutel informatie objecten.
- Beheer verzamelde en ontdekte metadata als een toegankelijk, doorzoekbare bron en gebruik metadata om consolidatie te vergemakkelijken.
- Verzamel data van kandidaat databronnen, evalueer hoe verschillende data refereren naar dezelfde dagelijkse entiteiten en creëer een uniek, geconsolideerd overzicht van elk.
- Verstrek methodes voor transparant toegang tot het gezamenlijk overzicht van dagelijkse objecten voor zowel bestaande als nieuw ontwikkelde bedrijfsapplicaties
- Installeer geschikt databeheer en managementbeleid en -procedures op management en operationeel niveau om verzekerd te zijn van hoge kwaliteit master data.

1.3 Voor- en nadelen van MDM

De voornaamste voordelen van MDM zijn:

- Uitgebreide kennis van klanten door consolidatie van verschillende bronnen
- Verbeterde klantenservice door beschikbaarheid en juistheid van data
- Consistente rapportage
- Verbeterde concurrentiekracht door gereduceerde complexiteit en verhoogde flexibiliteit
- Beter risico management door betrouwbare en consistente informatie
- Verbeterde operationele efficiëntie en kostenreductie door minder onderhoud
- Verbeterde besluitvorming
- Betere analyse en planning
- Naleving van regelgeving door eenvoudige controles op informatie
- Verbeterde informatiekwaliteit
- Snellere resultaten
- Verbeterde bedrijfsproductiviteit
- Vereenvoudigde applicatie ontwikkeling door geconsolideerde applicatie functionaliteit

Enkele nadelen van MDM zijn:

- MDM vereist instemming van alle deelnemers om succesvol te zijn. Als die ontbreekt, ligt onvolledige implementatie op de loer.
- Bedrijfsapplicaties zijn afhankelijk van één informatiebron. Als deze fouten bevatten, werkt dat organisatiebreed door.

Wat MDM onderscheid van vorige integratie inspanningen is, dat de focus niet alleen op technologie ligt, maar vooral op de bedrijfsvoering, geconcentreerd op het proces van entiteit identificatie en validatie met de interne gebruikers. Waar MDM de technologie en tools gebruikt, is het de combinatie van die technologie met de juiste bedrijfs- en data management praktijken die het doorslaand succes levert.

1.4 Organisatorische uitdagingen

MDM moet beschouwd worden als een “programma”, niet als een project of een applicatie. Een succesvol MDM programma onderscheidt zich door:

1. Organisatorische bereidheid
2. Data governance
3. Metadata management
4. Technologie integratie
5. Anticipatie op verandering

Data kwaliteit

Data kwaliteit is kritisch voor het succes van MDM en het is van invloed op bijna alle aspecten van het MDM programma.

Technologie

MDM is geen technologisch project, maar het ontwikkelen van een MDM programma is onmogelijk zonder tools en technologie. Het gaat in deze bijlage dus niet om de tools, maar meer hoe de tools worden gebruikt om een geconsolideerd overzicht van uniek te identificeren master dat objecten te krijgen.

2 Coördinatie: belanghebbenden, eisen en planning

Het beheren van de master data is een kritische component van de bedrijfsinformatie management strategie. Dit vergt veel energie met betrekking tot coördinatie: het identificeren van belanghebbenden, hun steun krijgen, onderlinge samenwerking bewerkstelligen, eisen opstellen en rollen en verantwoordelijkheden aan de juiste mensen toekennen.

Alvorens een MDM wordt gebouwd en geconfigureerd, moeten er door de teamleden een aantal taken uitgevoerd worden:

- Identificeer de personen in de organisatie die profijt hebben van MDM
- Omschrijf de toegevoegde waarde van MDM en ontwikkel de rechtvaardiging ervan
- Verzamel en prioriteer de data eisen, welke het ontwerp en de ontwikkeling van de onderliggende infrastructuur bepalen
- Ontwerp een plan voor bedrijfsbrede data integratie
- Ontwerp een migratieplan voor de betreffende applicaties

2.1 Wie zijn de belanghebbenden?

Er zijn een achttal belanghebbenden te onderscheiden:

1. Hoger management
Deelname is een absolute voorwaarde, zij zorgen voor nieuwe initiatieven, betrokkenheid en gedragsveranderingen van de deelnemers ten gevolge van gewijzigde verantwoordelijkheden en beloningen
2. Interne gebruikers
Ondersteuning omvat een aantal specifieke acties en verantwoordelijkheden:

- Het MDM programma team moet verwachtingen van de gebruikers omtrent data en service levels boven water halen en documenteren en hen verzekeren dat die verwachtingen worden waargemaakt.
 - Bepaal welke data objecten worden gebruikt door de bedrijfsapplicaties en hoe die worden gebruikt (om het globale plaatje te begrijpen voor het technisch team).
3. Eigenaars van applicaties
Zij moeten er zeker van zijn dat de bedrijfsapplicaties voorspelbaar opereren
 4. Informatie architecten
Zij zorgen voor de informatie modellen in de bedrijfsarchitectuur (huidige behoeften en toekomstige eisen)
 5. Data Governance en Data kwaliteit
De databeheerder en data eigenaar bewaken de beperkingen waarmee gebruikers data creëren, benaderen, gebruiken, veranderen en verwijderen
 6. Metadata analisten
Metadata management moet sterk gekoppeld worden aan informatie, applicatie architectuur en data governance
 7. Systeemontwikkelaars
Aspecten van prestaties en opslag als data in het master data systemen worden gekopieerd
 8. Operationele medewerkers
Zij moeten wellicht hun gebruiken veranderen als data worden overgezet naar de master omgeving (als er nog gewerkt wordt met “bypasses”)

Ontwikkel een project handvest

Om een duidelijk beeld te schetsen en voor acceptatie van het MDM programma door het hoger management, moet het MDM team een handvest opstellen.

In een project handvest staat, wat er bereikt gaat worden met het MDM programma en hoe de diverse mijlpalen bereikt worden.

Het project handvest beschrijft de business case, het doel en de waarde van MDM en de belangrijkste personen in het project. Verder staan er de doelstellingen in, de scope, de acceptatie- en succescriteria en potentiële risico's.

2.2 Coördinatie van deelnemers

Een team bestaat veelal uit een diversiteit aan deelnemers en hun verschillende eisen en verwachtingen. Om de individuele behoeften af te stemmen op de bedrijfsdoelen voor data integratie zijn er een aantal technieken voor de initiële en voortdurende coördinatie van deelnemers:

- Stel processen en procedures op voor samenwerking voor de kick-off en ontwikkel basisregels voor deelname.
Individen uit verschillende afdelingen en van verschillende applicaties zullen verschil van mening hebben over namen, definities, bronnen en gebruik. Het doel is om te kijken waar de definities en betekenissen overeenkomen en waar de verschillen zitten. De processen en procedures voor samenwerking moeten gaan over metadata, over harmonisatie en standaardisatie van gebruik waar mogelijk en scheiding van gebruik waar dit niet mogelijk is.
- Gebruik een RACI model voor overzicht
RACI staat voor Responsible, Accountable, Consulted, Informed.
Om er voor te zorgen dat de behoefte van elke deelnemer zijn bepaald en dat hun taken goed zijn uitgevoerd, moet er een beschrijving zijn van de rollen en verantwoordelijkheden. Het RACI model is een voorbeeld hiervan en is een tweedimensionale matrix, waarbij de taken in de rijen staan en de rollen in de kolommen.

- **Gebruik Business Process Modelling, BPM**
Als een organisatie data consolideert in een master omgeving, ontstaat de mogelijkheid om de bedrijfsprocessen werkelijk te begrijpen en hoe die de verschillende data objecten gebruiken.
Het afzonderlijk modelleren van bedrijfsprocessen van de huidige toepassing, openbaart kennis van toe te passen applicaties.
- **Verstrek master metadata**
Een doelstelling van MDM is het streven naar overeenstemming over wat de master data objecten zijn en hoe die worden gedefinieerd en gebruikt. Door het verstrekken van een set van processen voor overeenstemming en het opstellen van een metadata catalogus vormt het management platform voor het coördineren van gedeelde data objecten en data elementen.
- **Stel data governance op**
Een data governance programma omvat data kwaliteit, standaarden, verantwoordelijkheid en audit en controle van rapportage voor het nakomen van beleid . Data governance is voor MDM een van de belangrijkste peilers.

2.3 Haalbaarheidsstudie

Om de haalbaarheid van MDM vast te stellen, worden de eisen die aan data gesteld worden geanalyseerd.

Een haalbaarheidsstudie legt de nadruk op het verzamelen van eisen, de evaluatie en definitie van het juiste detailniveau van data- en procesmodellen. Dit ter voorbereiding van het toepassen van master data services. Een haalbaarheidsstudie omvat de volgende stappen:

- **Identificeer de context**
Het doel van dit proces is het karakteriseren en documenteren van de omgevingscontext en bedrijfsimpact en -beperkingen, gerelateerd aan verzamelen en verspreiden van master data.
- **Voer interviews met belanghebbenden**
Om de eisen van belanghebbenden te begrijpen is het van belang de eisen van operationele data te verzamelen en deze om te vormen naar een bedrijfsbreed overzicht.
- **Stel eisen op**
Beoordelen, samenvatten en ordenen van de verkregen informatie uit de interviews en deze informatie samenvoegen met de bedrijfscontext. Hierbij wordt gekeken naar overeenkomsten van de data objecten welke het vaakst zijn geïdentificeerd tijdens de interviews en de objecten die in de bedrijfsprocessen naar voren zijn gekomen.
- **Bepaal haalbaarheid**
Het proces voor het opstellen van eisen aan data moet bepalen of er gebruikelijke data objecten zijn die op gelijke wijze worden gebruikt door de applicatie infrastructuur en of er kandidaat data sets zijn die effectief gebruikt kunnen worden voor het master overzicht. Dit vormt samen met de rechtvaardiging van de bedrijfswaarde de haalbaarheid.

3 MDM componenten en het volwassenheidsmodel

Voordat aan een MDM programma begonnen wordt, is het zaak om eerst de essentiële bekwaamheden te begrijpen om MDM succesvol uit te rollen en de volwassenheid van die bekwaamheden om MDM uitvoerbaar te maken.

Er is geen enkele lijst die alle benodigde services bevat, welke een bedrijf nodig heeft voor zijn master data set, maar het is waardevol om high level de MDM bekwaamheden te inventariseren. In deze paragraaf worden een aantal technische MDM componenten besproken en wordt het volwassenheidsniveau van elk van die componenten bekeken. Aan de hand van de huidige situatie en vervolgens de gewenste situatie van een onderneming, kan een MDM programma manager een implementatie roadmap ontwerpen met de te nemen stappen in het MDM programma.

Om een MDM oplossing in te zetten, moet de onderneming voorbereid zijn op de technische, operationele en management uitdagingen, waarbij de ontwikkeling in meerdere iteraties zal verlopen. Een volwassen MDM oplossing zal de bekwaamheden en services, benoemd in onderstaande figuur 3, bevatten.

Architectuur	Service laag (controle)			
	Masterdata model (structuur)		Systeemarchitectuur (kracht)	
Governance				
	Data standaarden	Metadata management	Data kwaliteit	Databeheer
Management	Migratieplan			
	Hierarchie		Identiteit	
	Administratie en configuratie			
Identificatie	Zoek en bepaal unieke omschrijving			
	Koppeling records		Samenvoegen en versterken/completeren	
Integratie	Applicatie integratie en synchronisatie laag			
	MDM component service laag			
Business Process Management	Business Process Integratie			
	Business rules			
	MDM business component laag			

Figuur 3 MDM componenten en service model (service lagen)

Hieronder wordt elke laag van deze componentstapel besproken.

3.1 Componenten

De MDM stapel bestaat uit zes componenten:

1 Architectuur

Dit omvat de volgende aspecten:

Master data model (structuur)	Er moet een geconsolideerd model voor master data zijn en modellen voor data extractie en data uitwisseling.
Systeem architectuur (kracht)	De architectuur is gebaseerd op een service georiënteerd raamwerk waarin de functionaliteit bestaat uit de levenscyclus activiteiten (creëer, benader, update en verwijderen) gerelateerd aan het master object type.
Architectuur service laag (controle)	Een servicelaag voor componenten en applicaties. Aangezien applicaties steeds afhankelijker worden van de MDM omgeving, worden er ook meer eisen gesteld aan data object services gerelateerd aan het serviceniveau t.b.v. gebruik van applicaties, zoals synchronisatie, toegangscontrole, integratie en consolidatie. Verschillende typen applicaties kunnen verschillende serviceniveaus eisen, zodat het wellicht waardevol is om die componenten te scheiden.

2 Governance

Dit richt zich op het integreren van technisch datamanagement met overzicht, waarmee organisatorisch toezicht gehouden wordt op het gedefinieerde informatiebeleid.

De volgende gebieden zijn daarbij van belang:

Data standaarden	Als sommige data elementen met dezelfde naam niet dezelfde betekenis hebben, of met verschillende namen welke juist wel behoren tot hetzelfde concept, dan vormt dat een gevaar voor het integratie- en consolidatieproces. Gestandaardiseerde definities moeten de naamgeving van de master data object catalogus sturen en controleren en hoe die tot uniek overzicht komen.
Metadata management	De metadata database van de onderneming wordt ook wel gezien als de controlekamer, welke de bedrijfsapplicaties stuurt en beheert. Daarom is een bedrijf metadata management systeem een kritische component om het juiste controleniveau te behalen. Het is van groot belang om het complete plaatje te krijgen van het gespreid gebruik van master data objecten door heel de organisatie en het bereiken van consensus op bedrijfs metadata.
Data kwaliteit	Het besef, dat prestaties van de onderneming en operationele productiviteit op zowel organisatie niveau als persoonlijk niveau, afhankelijk is van data van hoge kwaliteit is een kern competentie van elk MDM programma.
Databeheer	De operationele aspecten van data governance, geformuleerd als databeheer activiteiten, vullen eigenaarschap modellen en overzichtsmodellen aan, voor actief beheer en informatiekwiteit.

3 Management

Dit gaat om de volgende aspecten:

Identiteitsmanagement	Elk voorbeeld van een master data object type moet een uniek object vertegenwoordigen, waarbij er slecht één uniek identificeerbare record is van een klant, product, medewerker, etc. Identiteitsmanagement beheert de bepaling van de te identificeren attributen, noodzakelijk voor unieke identificatie, samen met de zoek en match technieken om zowel exacte overeenkomsten als “bijna” overeenkomsten op te sporen. Het resultaat van dit concept is de input voor de “identiteit zoektocht en vaststelling” component van het volgende niveau.
Hiërarchiemanagement	Dit richt zich op het bewerken van meerdere records naar een enkele representatie, het terugkoppelen van data van de master index naar de originele bronbestanden en op onderliggende verbindingen van master objecten over meerdere systemen.
Migratiemanagement	Dit richt zich op applicatiemigratie en -transitie bij modernisering.
Administratie en configuratie	Het MDM technisch team heeft tools en processen nodig om het dagelijks beheer van diverse aspecten van het onderliggende MDM raamwerk te configureren en te verzorgen.

4 Identificatie en consolidatie

Om één enkele versie van een entiteit te verkrijgen, worden de volgende aspecten bekeken:

Identiteit zoektocht en vaststelling	Identiteitsvaststelling gaat om het vaststellen dat twee of meer data samengevoegd kunnen worden tot één representatie van een uniek object, zoals productnamen of productcodes. Deze vaststelling begint met het “ontdekken” en combineert data profilering activiteiten met handmatige beoordeling van data. Daarna volgt gelijkenis aantekenen en matchen om zo records te koppelen.
Koppeling records	De algoritmes uit het vorige punt worden, al of niet geautomatiseerd, toegepast op een grotere hoeveelheid records vanuit verschillende bronnen. Deze component onderhoud zowel identiteitsmanagement als de processen voor samenvoegen en consolideren.
Samenvoegen en consolideren	Gelijke records worden onderworpen aan algoritmes om de waarden te kwalificeren binnen elk data attribuut.

5 Integratie

Dit omvat de volgende aspecten:

Applicatie en synchronisatie service laag	Informatie moet eenvoudig en geconsolideerd worden in de master en moet eenvoudig toegankelijk worden door bedrijfsapplicaties. Het MDM raamwerk moet afgestemd worden op bestaande applicatie infrastructuren, deze minimaal verstoren en voorzien in een gestandaardiseerde transitieroute naar de gesynchroniseerde master.
MDM component service laag	Gestandaardiseerde master data objecten reduceren de behoefte van applicatie architecten om zich te richten op traditionele data georiënteerde problemen en kan in plaats daarvan meer abstracte functionaliteiten gebruiken. Deze mogelijkheid om applicatiefunctiealiteit te consolideren door het gebruik van service lagen, geeft toegevoegde waarde voor zoowel bestaande als toekomstige applicaties door vereenvoudigde incrementele ontwikkeling.

6 Business Proces Management

Op het hoogste abstractieniveau worden de eisen belicht voor het maken van applicatie ontwerp beslissingen. Een belangrijk doel binnen MDM systeemontwerp is dat het systeem op businessniveau gestuurd wordt, niet op technisch niveau. Op dit niveau wordt systeemarchitectuur ingelijfd door bedrijfsproces modellering.

Integratie van bedrijfsprocessen	Het model toont dat de output van de ene activiteit is gekoppeld aan getriggerde gebeurtenissen van een andere activiteit, welk het gedrag van betreffende applicaties beïnvloeden. De beschrijvingen van het bedrijfsprocesmodel worden aangevuld met de master data objecten, wat noodzakelijk is om de procedure te voltooien. Dit integreert het bedrijfsproces effectief met MDM, waarbij de data afhankelijkheid zichtbaar wordt en de identificatie en selectie van master data objecten valideert.
Business rules	Binnen elk bedrijfsproces model combineert de logica, die wordt gebruikt voor het uitvoeren van een bepaalde operatie, de evaluatie van waarden van gedeelde data objecten met de waarden van gedefinieerde controles. Het beoordelen van het bedrijfsproces model stelt de applicatie ontwerpers in staat om specifieke regels te identificeren, alsook het naar buiten brengen van de volledige set van condities waarop gericht moet worden tijdens het bedrijfsproces. Dit beoordelingsproces leidt tot een meer compleet model van het systeem en de overeenkomstige master data afhankelijkheden.
MDM bedrijfscomponent laag	De bedrijfscomponent laag ligt ten grondslag aan de definities en eisen welke naar voren zijn gebracht door BPM, integratie en de toepassing van business rules. In deze laag gaan meer geraffineerde, herbruikbare bedrijfsservices gevormd worden, zoals voorspellende analyses geïntegreerd in operationele applicaties.

3.2 Het model

Het volwassenheidsniveau beschrijft hoe reeds ingezette componenten en services kunnen worden benut ten behoeve van een MDM programma en geeft aan welke ontbrekende bekwaamheden moet worden verkregen, zodat een applicatie op een betrouwbare manier gebruik kan maken van master data. Er zijn een vijftal volwassenheidsniveaus te onderscheiden:

1 Initieel

Op dit niveau zijn er slechts beperkte mogelijkheden om gebruik te maken van master data, maar er is een bepaalde mate van herkenning dat er meerder kopieën zijn van data sets die relevant zijn voor meerdere applicaties. Er is enigszins bereidheid tot consolidatie van data voor analytische doelen.

2 Reactief

Op dit niveau heeft er nog geen analyse plaatsgevonden van de echte bedrijfsbehoeften, terwijl het technische team wel tools koopt. Dit bevredigd enkele operationele applicatie behoeften, maar leereffecten worden niet gedeeld. Er worden pogingen ondernomen voor het consolideren van metadata vanuit verschillende applicaties.

3 Beheerst

Hier wordt het hoge management betrokken wat leidt tot uitgebreidere bedrijfsmodellering activiteiten. Het gebruik van master data wordt een herhaalbaar proces en kan uitgebreid worden om zowel nieuwe als bestaande applicaties in te lijven. Consolidatie en synchronisatie services zijn beschikbaar.

4 Proactief

De applicaties zijn over het algemeen geïntegreerd door de service laag met de master data omgeving. Synchronisatie, identiteit vaststelling, hiërarchie management en identiteitsmanagement zijn ingebed in de component service laag. Data governance is bedrijfsbreed aanwezig.

5 Strategisch

MDM, gekoppeld aan een service orientated architecture (SOA), maakt snelle ontwikkeling mogelijk van applicaties van hoge kwaliteit, dat zowel de operationele als analytische eisen aan bedrijfsapplicaties ondersteunen.

Analytische resultaten, verenigd met business intelligence processen, worden beheerd als master objecten. Dit maakt effectievere en consistente, voorspellende analyses mogelijk, welke ingebed worden in klantgerichte applicaties.

In bijlage 4 staat een voorbeeld van een Maturity Matrix, waarin een onderneming het huidige en het gewenste volwassenheidsniveau kan aangeven.

3.3 Ontwikkel en implementeer een roadmap

Hoewel elk bedrijf uiteindelijk een strategische volwassenheid wil bereiken, is het raadzaam om met behulp van een roadmap door de verschillende volwassenheidsniveaus te lopen, om vervolgens de verwachte waarde voor het bedrijf bij elk bereikt niveau te beschrijven.

Dit proces heeft drie voordelen, namelijk:

- Door leereffecten uit “big bang” projecten, zoeken project managers meer naar korte termijn tactische bereikte waarden.
- Door een beeld te hebben van het eindstadium en daar naartoe te werken geeft een goed beeld van het ontwerp, ontwikkeling en migratieprocessen tot het punt dat applicaties en informatiearchitecturen kunnen vertrouwen op de master data.

- De strategische kijk maakt heldere positionering, communicatie en socialisatie mogelijk van de organisatorische veranderingen, wat nodig is voor een fatsoenlijke acceptatie van de transitie.

Het doel van het toepassen van een roadmap is om de bedrijfsdoelen helder te krijgen, weten welke bekwaamheden er nodig zijn om die doelen te bereiken, bepalen welke bekwaamheden al in de organisatie aanwezig zijn, welke kloven zijn er, bepaal waar de organisatie moet zijn en beslis hoe daar te komen.

De implementatie van een roadmap verloopt als volgt:

- Evaluer de bedrijfsdoelen
- Evaluer de bedrijfsbehoeften
- Bepaal de huidige status
- Bepaal de initiële kloof
- Bepaal de gewenste situatie
- Analyseer de kloof met betrekking tot mogelijkheden
- Breng de mogelijkheden in kaart
- Plan het project

Het volwassenheidsmodel wordt gebruikt als de maatstaf waarmee de organisatie zijn MDM bekwaamheid en gereedheid worden gemeten. Het wijst een eindstadium aan, welke de bedrijfsanalisten helpt de uitvoering van het plan vorm te geven naar de huidige applicatie architectuur en bekijk de beste manier om naar het eindstadium te komen.

4 Data governance

Data governance is het beheren en bewaken van de risico's die zich voordoen in een bedrijfsinformatie portfolio en het reduceren van impact als gevolg van het ontbreken van overzicht. Van data governance wordt verwacht dat de data voldoen aan alle bedrijfsdoelen in de context van databeheer, eigenaarschap, compliance (nakomen van het bedrijfsbeleid), privacy, veiligheid, data risico's, gevoeligheid van data, metadata management en MDM.

Elk van deze problemen moeten worden geïntegreerd met controle voor management overzicht samen met verificatie van organisatorisch toezicht en informatiebeleid.

Als een bedrijf meerdere informatiesystemen heeft, is het verstandig dat de directie / het management uitspraken doet over de wijze waarop daarmee omgegaan moet worden.

Dit noemen we informatiebeleid.

Met andere woorden, elk aspect van data governance is gerelateerd aan de specificatie van informatiebeleid dat de bedrijfsbehoefte en verwachtingen reflecteert, samen met de processen voor het monitoren van conformiteit naar dat informatiebeleid.

Deze paragraaf gaat over de sturende factoren voor het opzetten van een data governance programma in relatie tot master data management, het bepalen van kritische data elementen en hun relatie met sleutel entiteiten van het bedrijf, te nemen stappen voor implementatie en rollen en verantwoordelijkheden.

4.1 Afstemming informatie doelstellingen met de bedrijfsstrategie

Elke organisatie heeft een bedrijfsstrategie dat de bedrijfsdoelstellingen reflecteert, evenals de risico management doelstellingen en de compliance management doelstellingen. Uiteindelijk hangt het succes van de onderneming af van het beheer over hoe alle processen conform de bedrijfsstrategie

verlopen. De uitdaging ligt in de bekwaamheid van het management om de bedrijfsstrategie effectief te communiceren.

Als voorbereiding voor het opzetten van een gedegen data governance niveau, zouden de volgende stappen genomen moeten worden:

Duidelijkheid geven over de informatie architectuur

Dit proces omvat identificatie van data sets voor elke applicatie en nummering van data attributen binnen elke data set. Dit proces vergt echter enorme inspanningen, zodat het verstandig is om zich te richten op die data elementen die een uitgebreide specifieke relevantie voor het bedrijf heeft. Een team van experts en data analisten moeten kijken naar de fysieke componenten en deze herschikken naar een logisch overzicht welke consistent is binnen het gehele bedrijf. Het harmonisatie proces onderzoekt de data elementen naar overeenkomsten of afwijkingen in betekenis. Hoewel deze activiteit de technische aspecten van master data object analyse overlapt, zit het belang van toezicht in de identificatie en registratie van de kritische data elementen, de samengestelde informatie structuren en de applicatiefuncties gekoppeld aan de data element levenscyclus.

Mapping van informatie functies naar bedrijfsdoelstellingen

Bij het controleren van de informatie architectuur, moet elke informatie functie gedocumenteerd worden en hoe deze gekoppeld is om bedrijfsdoelstellingen te bereiken.

In een omgeving met veel gelijke data zijn ook veel gelijke functionaliteiten en als de data in een master data hiërarchie worden gezet, dan worden de functionele mogelijkheden ook geconsolideerd naar een service laag, welke het bedrijf ondersteund.

Een proces raamwerk opzetten voor informatiebeleid

Het doel van een assessment is het begrijpen van hoe de informatie architectuur en de bijbehorende functionaliteit de implementatie van informatiebeleid ondersteund.

In werkelijkheid moet het proces andersom zijn, informatiebeleid moet eerst gedefinieerd zijn en daarna de dataobjecten en de bijbehorende services worden ontworpen voor implementatie en documentatie van compliance met dat beleid.

Data governance en data kwaliteit

Een data kwaliteit en data governance assessment verklaart hoe de informatiearchitectuur wordt gebruikt ter ondersteuning van compliance met gedefinieerde informatiebeleid. Het geeft aan, dat data kwaliteit en data standaard management een onderdeel zijn van een veel groter geheel met betrekking tot overzicht van bedrijfsinformatie.

Voor dit overzicht is nodig:

- Beoordeling van data kwaliteit dimensies in relatie tot de data elementen
- Bepalen hoe kritisch deze zijn voor bedrijfsprocessen
- Gegevensregels bepalen, welke compliance beïnvloeden
- Kwantitatieve metingen definiëren voor conformiteit aan informatiebeleid
- Manieren bepalen om het voorgaande te integreren in een data governance raamwerk

4.2 Risico's

Een belangrijke peiler voor de noodzaak van data governance is risico, zowel bedrijfsrisico als compliance risico. Het is goed om te kijken naar enkele risicogebieden in relatie tot master data, welke informatie management vereisen.

Risicogebieden

Enkele risicogebieden zijn:

- Bedrijf en financieel
bijv. onvoldoende controle op financiële rapporten heeft negatief effect op het bedrijf en kan leiden tot waardevermindering, onvoldoende overzicht over betalingen kan leiden tot boekhoudkundige afwijkingen.
- Rapportage
t.b.v. voorschriften zoals SOx, onjuiste data kunnen leiden tot waardevermindering van het bedrijf, rechtszaken, etc.
- Kennis van entiteiten
Kennis van partijen waar de onderneming zaken mee doet is kritisch voor reductie van risico's (kunnen zij hun beloftes nakomen?)
- Bescherming, privacy en beveiligde informatie
- Beperking van gebruik, toegangsrechten

Risico's van MDM

MDM vraagt overeenstemming van deelnemers om het programma succesvol te laten verlopen. Dit leidt tot unieke uitdagingen voor bedrijven die aan MDM beginnen.

- Consensus voor coördinatie en samenwerking
Alle deelnemers delen hun data, leveren positieve input naar verbetering en vertrouwen de uiteindelijk geconsolideerde versies.
- Eigenaarschap van data
Beheer en overzicht wordt vereenvoudigd, omdat het aantal data objecten gereduceerd is en overeenkomstig ook de hoeveelheid ondersteuning reduceert (gecentraliseerd eigenaarschap). Een nadeel hiervan is, dat het verantwoording wegneemt van individuen, waarmee sommigen zich bedreigd voelen.
- Semantiek: vorm, functie en betekenis
De variatie van betekenissen tussen applicaties vormen diverse risico's tijdens de consolidatie van elk bedrijfsdata object naar een enkele weergave. Bijvoorbeeld verschillende betekenissen van het begrip "klant" wordt misschien getolereerd binnen een applicatie, maar in een geconsolideerde omgeving is er veel minder tolerantie.

Risico's beheersen

Risico's beheersen door middel van gemeten conformiteit en informatiebeleid:

Om blootstelling aan risico's te voorkomen, moet er een beleid gevormd worden dat de grenzen van risico's aangeeft en een overkoepelend proces om ervoor te zorgen dat het beleid nagekomen wordt. Elk beleidsstuk kan verschillen afhankelijk van het risico, maar de problemen vertonen overeenkomsten:

- Vereniging
Alle bestuursgroepen moeten zich verenigen om hun afdelingsgrenzen te laten vervagen voor bedrijfsbreed compliance management

- Gedefinieerd beleid
Richtlijnen voor risicobeperking, vertaald naar regels
- Behoeft aan controles
Meten van conformiteit aan het beleid
- Transparant overzicht
Beleid conformiteit moet aan iemand of aan instantie gerapporteerd worden
- Te verifiëren
Moet een audit “spoor” vormen, welk te beoordelen is en moet tevens een transparant proces hebben

Wat de doelstelling ook is, gecentraliseerd management verspreid data governance door de organisatie. Dit gebeurt over het algemeen met twee technieken:

- Data governance wordt gedefinieerd door middel van een verzameling van informatiebeleidstukken, waarbij elke wordt gemapt naar een set regels welke opgelegd worden over de levenscyclus van kritische data elementen.
- Databeheerders krijgen de verantwoordelijkheid voor zowel de kwaliteit van de data elementen en de zorg voor conformiteit aan het informatiebeleid.

Deze vormen de technische middelen voor het monitoren van data governance en leveren de context voor effectief beheer van data governance.

Sleutel data entiteiten

Dit zijn de fundamentele informatie objecten die worden gebruikt door de bedrijfsapplicaties om de operatie van het bedrijf uit te voeren en het legt tevens de basis voor het analyseren van de prestaties. In een MDM omgeving zijn de data objecten meestal ook de sleutel data entiteiten. Omdat deze entiteiten worden samengevoegd in één bedrijfsbron is het belangrijk om een beleid te hebben, welke voorschrijft hoe verantwoordelijkheden van eigenaarschap worden toegekend.

Kritische data elementen

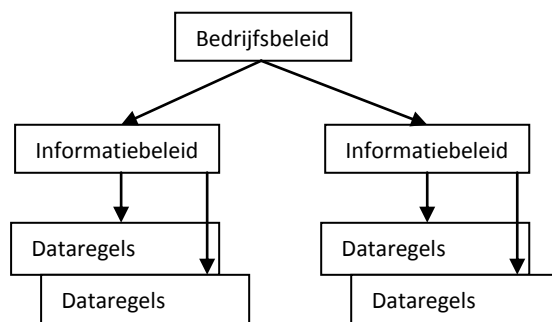
*Dit zijn de elementen, welke van vitaal belang zijn voor de succesvolle operatie van de organisatie. (bijvoorbeeld identificerende informatie van data objecten zoals leveranciers, klanten)
Het beste proces om uit alle data elementen het kritische element naar boven te halen is het werken van voren naar achter bij het onderzoeken van informatie artefacten (rapporten, analyses of operaties). Kijk naar elk geïdentificeerd data element (bijv. totale verkoop) en kijk vervolgens naar de data elementen waarvan de voorgaande afhankelijk is (verkoop per divisie). Dit kan stapsgewijs uiteengegrafeld worden, waarbij steeds kritische data elementen benoemd worden.*

Definitie van informatiebeleid

Informatiebeleid vertegenwoordigt de specificaties van management doelstellingen verbonden aan data governance, of ze nu gerelateerd zijn aan risicobeheer of aan algemeen dataoverzicht. In feite is het informatiebeleid die een brug slaat tussen bedrijfsbeleid en de karakterisering van de kwaliteit van kritische data elementen.

Metten en monitoren

Metingen welke de conformiteit met het informatiebeleid reflecteren kan gezien worden als een samenvatting van de verschillende dataregels die vanuit het beleid uiteen zijn gerafeld. Zolang elke regel meetbaar is, kan een hiërarchie van metingen gecreëerd worden die uiteindelijk gecombineerd kunnen worden tot KPI's ten behoeve van data governance.



Figuur 4 Een hiërarchie van dataregels worden afgeleid van het bedrijfsbeleid

Om monitoring effectief te laten verlopen, moeten de metingen direct getoond worden aan de verantwoordelijke. Het is dan aan hem om continu de conformiteit aan het beleid te monitoren en bij problemen een proces van “afgeleiden” toe te passen om te bepalen welke punten gefaald hebben en vervolgens een verbeterproces te initiëren.

4.3 Raamwerk voor verantwoordelijkheid

Een van de grootste problemen die in het verleden bij data governance optraden was het gebrek aan opvolging. Er ontbrak vaak de onderliggende organisatorische structuur om het ook uitvoerbaar te maken. Dit vergt twee dingen:

- De definitie van de management structuur om de uitvoering van het governance raamwerk te kunnen volgen
- Een beloningsmodel voor de uitvoering

Er bestaan verschillende rollen binnen een data governance raamwerk:

- Data governance bestuurder, dagelijkse leiding, communiceert met en rapporteert aan deelnemers
- Data governance raad, afstemming van processen en houdt toezicht op databeheerders
- Databeheerder, verantwoordelijk voor kwaliteit en standaardisatie van data

5 Data kwaliteit

Datakwaliteit is, zoals hierboven genoemd, een belangrijk onderdeel van data governance. De uitdagingen voor het monitoren van en zorg dragen over data kwaliteit binnen een MDM omgeving zijn:

- het identificeren van kritische data elementen
- het bepalen van welke data elementen bijdragen aan master data
- lokaliseren en isoleren van master data objecten binnen het bedrijf
- beoordelen en wegnemen van variaties tussen de verschillende representaties

om zo tot een geconsolideerde weergave te komen.

Zelfs na een initiële integratie van data naar een master data omgeving, is er nog steeds behoefte aan data inspectie, monitoring en controles ter identificatie van potentiële gevaren voor datakwaliteit.

5.1 Dimensies

Dimensies van datakwaliteit zijn afgestemd op de te meten bedrijfsprocessen. Dimensies geassocieerd met datawaarden lenen zich goed voor systeemautomatisering, maken zich geschikt voor dataregels binnen datakwaliteit tools voor data validatie.

Dit zijn o.a. de volgende dimensies:

- Uniceit
Er is geen entiteit in de MDM omgeving die meer dan één keer voorkomt en deze kan ook uniek benaderd worden (sleutel) in een data set
- Accuraatheid
De mate waarin een gegeven op juiste wijze een “dagelijks” object vertegenwoordigen
- Consistentie
Twee datawaarden uit verschillende data sets mogen niet met elkaar conflicteren
- Compleetheid
Specifieke data elementen mogen niet de waarde nul hebben
- Tijdigheid
De verwachte tijd voor benadering en beschikbaarheid van informatie
- Actualiteit
De mate waarin informatie actueel is

Daarnaast wordt er gekeken naar data formaten en patronen en referentiele integriteit.

Data entiteiten die als master data objecten worden beheerd krijgen een uniek label.

Er moet geverifieerd worden of elke gelabelde entiteit refereert naar een bestaande entiteit binnen de omgeving.

5.2 Data integratie tools

Data kwaliteit en integratie tools zijn geëvolueerd van eenvoudige standaardisatie en patroonvergelijking naar toolsets voor complexe automatisering van data analyse, standaardisatie, matching en aggregatie. Bijvoorbeeld data profilering heeft zich ontwikkeld tot een set van complex geautomatiseerde analysetechnieken, welke gebruikt kunnen worden voor identificeren, isoleren, monitoren, auditen en het helpen vinden van afwijkingen welke de waarde van een bedrijfsinformatie master degraderen.

5.3 Data profilering

Data profilering is ontstaan als een set algoritmes voor statistische analyse en beoordeling van datawaarde kwaliteit in een data set, maar ook om relaties te ontdekken die bestaan tussen een verzamelingen van waarden binnen en buiten data sets.

Profilering voor metadata oplossing

Een voorwaarde om master data te kunnen beheren is het beheren van metadata.

De beste techniek voor het verzamelen van de meest uitgebreide en consistente metadata uit alle bedrijfsbronnen is het toepassen van zowel statistische als analytische algoritmes, verstrekt door profilering tools. Dit resulteert in een empirische beoordeling van structuur en formaat metadata en tegelijk gekoppelde data modellen en afhankelijkheden.

Profilering voor data kwaliteit beoordeling

Voor het migreren van databronnen naar de master database is het noodzakelijk om de kwaliteit te beoordelen van elke bron en bepalen in welke mate die bron overeenkomt met de wensen van het bedrijf. Hiervoor is profilering nuttig. Kolom profilering geeft statistische informatie van verdeling van datawaarden en geassocieerde patronen, welke zijn toegewezen aan elk data attribuut, inclusief reeks analyse, verspreiding, formaat en patroon evaluatie, analyse van kardinaliteit en uniciteit, ontbrekende waarden, etc.

Data cleansing

Het automatiseringsproces voor MDM in de service laag is in staat om de functionaliteit van databaseer te integreren. In plaats van batch verwerking van cleansing kunnen services dit nu automatisch uitvoeren .

Data controles

Data kwaliteitscontroles zijn bedoeld voor het beoordelen van afwijkingen in data, het vaststellen van de oorzaak en deze, indien mogelijk wegnemen.

Ondanks alle moeite voor het bereiken van data kwaliteit, zullen er altijd afwijkingen zijn die aandacht vragen en waarvoor een oplossing gezocht moet worden. Hiervoor moeten protocollen opgesteld worden om fouten zo snel mogelijk op te sporen, naar de juiste persoon te sturen en binnen een redelijk tijd op te lossen.

Hierbij zij twee aspecten van belang:

- Controles, op proces en op data
- Service level agreements (SLA's), belangrijke component voor het controle raamwerk

Data kwaliteitscontrole

Dagelijks proces ter reductie van het aantal fouten tot een redelijk en beheersbaar niveau en tevens het zoeken naar een oplossing binnen een bepaalde tijd.

Data validatie

Het proces ter beoordeling en meting van conformiteit van data aan een set van gedefinieerde business rules.

6 Metadata management

Alle aspecten om de noodzaak van MDM te bepalen, vereisen een helder beeld van de informatie over data, welke in de organisatie gebruikt worden, de metadata.

Vaak wordt metadata omschreven als “gegevens over gegevens”. Het is echter meer dan alleen maten en types van elke data element, bijvoorbeeld ook betekenissen en domeinen.

Metadata kunnen worden ingedeeld in zeven niveaus, welke kritisch zijn voor MDM, zie figuur 5.

Bedrijfsdefinities			
Concepten	Bedrijfstermen	Definities	Semantiek
Referentie metadata			
Conceptuele domeinen Provincie	Waarde domeinen Utrecht	Referentie tabellen Onderlinge relaties van waardedomeinen	Mappings
Data elementen			
Kritische data elementen attributen en velden	Data element definities	Data formaten	Aliassen en synoniemen
Informatie architectuur			
Entiteitsmodellen	Relationele tabellen	Master object directory tracking object types bij data migratie	
Data Governance			
Informatiegebruik	Informatiekwaliteit	Datakwaliteit SLA's	Toegangbeheer
Services			
Service directory	Service gebruikers	Interfaces	
Bedrijfsmetadata			
Bedrijfsbeleid	Informatiebeleid	Bedrijfsregels	

Figuur 5 MDM Metadata stapel

De uitdaging is om te kijken naar de wijze waarop deze niveaus op elkaar in werken als onderdeel van een algehele metadata management strategie. In deze paragraaf wordt elke laag bekeken en beoordeeld op de relevantie voor het metadata management raamwerk.

Bedrijfsdefinities

Deze laag initieert high level de begrippen geassocieerd met bedrijfsactiviteiten, bedrijfsconcepten en termen die normaal gebruikt worden.

Concepten

Dit is een kernbegrip, zoals “product” of “klant” en is geassocieerd aan een kern karakteristiek, zoals “fabriek” of “waarde”, welk op hun beurt ook weer een concept kunnen zijn.

Termen

Welke verschillende termen worden er gebruikt voor elk concept?

Maak een mapping tussen de termen en concepten (bijv. klant en account).

Definities

Zoek een heldere definitie voor elke term in relatie tot het concept (wat is een klant?).

Semantiek

De semantiek beschrijft de onderlinge verbindingen van de concepten, termen en definities.

Referentie metadata

Referentie data zijn de verzamelingen van waarden, welke gebruikt worden om de bestaande applicatie gegevensopslag te vullen, evenals het master data model.

Conceptuele domeinen

Een land is een conceptueel domein, bijvoorbeeld "Nederland"

Deze bestaan uit provincies, welke zelf ook een conceptueel domein zijn.

Waarde domeinen

Een waardedomein is een verzameling van waarden in een conceptueel domein, bijvoorbeeld "Groningen", "Drenthe" of "Gr", "Dr".

Referentie tabellen

Voor elk conceptueel domein moet er een verbinding beschreven staan met een specifiek waarde domein en hoe die waarden in het waardedomein refereren naar de objecten in het conceptueel domein. Referentietabellen bevatten deze informatie met een directe een-op-een mapping en worden ook wel "code tabellen" of "lookup tabellen" genoemd.

Mappings

Voordat records worden geïntegreerd vanuit verschillende data sets, moet men weten wanneer er verschillende waarde domeinen worden gebruikt die refereren naar hetzelfde conceptueel domein. Hiervoor kunnen dus twee of meerdere referentietabellen bij elkaar gevoegd worden.

Data elementen

Een data element is bouwsteen voor data modellen en elk data element is gespecificeerd in definities, een naam, een representatie en een set van valide waarden.

Kritische data elementen

De data elementen die kritische informatie bevatten over producten, leveranciers, personen, etc.

Data element definitie

De definitie moet de essentiële betekenis van het element weergeven, precies en ondubbelzinnig.

Data formaten

Elk data element krijgt een data type (integer, varchar) of maat of lengte toegewezen.

Synoniemen

Twee data elementen refereren naar hetzelfde concept, ook al lijken ze totaal verschillend.

Informatie architectuur

Logische en fysieke modellen worden opgebouwd uit data elementen. Op dit niveau zien we een samensmelting van beschrijving van master data object types en hun structuur, meer concreet gemaakt met gedefinieerde data modellen.

Elk logisch model reflecteert een master object, refererend naar data elementen.

De master object directory traceert hoe master object types zijn gemapped naar de applicaties die ze gebruiken.

Data governance

Metadata ondersteund Data Governance op een aantal manieren:

- Het beoordelen van de data kwaliteit met behulp van een set van data regels en richtlijnen, welke kwantificeerbare metingen van informatie kwaliteit geven.
- Diverse data kwaliteit SLA attributen worden opgeslagen als metadata in de database, wat het voor de databeheerder makkelijker maakt om die overeenkomsten te monitoren.
- Toegangsrechten worden toegekend aan gebruikers voor bescherming van informatie.

Metadata services

Als meerdere master objecten worden samengevoegd in een enkele master, dan kan de afgeleide functionaliteit van de levenscyclus van master objecten ook geconsolideerd worden.

Er zijn drie verzamelingen van services:

- de essentiële conceptuele services
- de manieren waarop die services worden ingezet in de huidige omgeving.
- de gebruikte interfaces om de services aan te roepen

Bedrijf metadata

Vanuit de bedrijfsbeleid wordt het informatie beleid afgeleid en van daaruit worden business rules afgeleid. Het vastleggen van die business rules in een metadata database zorgt ervoor dat de toepassing van die regels geautomatiseerd kan worden in een geassocieerd softwaretoepassing voor die regels.

7 Identificeren van Master data en Meta data

Het identificatieproces bestaat uit twee activiteiten, een top-down proces voor beoordeling van bedrijfs datamodellen en procesmodellen om te identificeren welke data objecten kritisch zijn binnen de bedrijfsapplicaties. De andere activiteit is een bottom-up proces welke de bedrijfsgegevens evalueert voor het vinden van applicaties die data structuren gebruiken ter identificatie van master data objecten en deze samenbrengt naar een master data omgeving. Naast de omschrijving van master data objecten in paragraaf 1.1, kunnen deze ook nog als volgt gekarakteriseerd worden:

- dienen als referentie in meerdere bedrijfsgebieden en processen
- dienen als referentie in zowel transactionele als analytische systeemrecords
- kunnen geclassificeerd worden binnen een hiërarchie met verschillende niveaus van classificatie, attributen en specialisatie (bijv. “partij” kan betekenen “persoon”, “organisatie”, “klant”, etc.)

Top-down benadering, bedrijfsproces modellering

Organisaties moeten de relaties onderhouden tussen bedrijfsstrategie en de componenten van de applicatiemodellen. Door het beoordelen van bedrijfsprocesmodellen en vervolgens identificeren welke data objecten kritisch zijn in de processtromen van meerdere applicaties. De objecten die gedeeld worden met diverse bedrijfsactiviteiten zullen analisten gaan beheren als master data.

Bottom-up benadering, evaluatie van gegevens

Deze benadering identificeert master data objecten die reeds gebruikt worden binnen de organisatie. Het zoekt naar bedrijfsgegevens in applicaties welke in feite kopieën zijn van master objecten en gaat deze samenbrengen in een master data omgeving.

Het gaat daarbij om de volgende drie stappen:

1 Identificeren en centraliseren van semantische metadata

Om master data te beheren, moet men eerst in staat zijn om master metadata te beheren.

Hiervoor moeten dan eerst de volgende zaken op orde zijn:

- Formaat op element niveau: de juiste syntax
- Structuur op record niveau: dezelfde informatie over het data element
- Semantiek op elk niveau: dezelfde betekenis of definitie

De drie fasen van voor het werken naar een master data oplossing zijn:

- Verzamelen en analyseren master metadata zodat er een uitgebreide catalogus van bestaande data elementen is. De uitdaging bij metadata is dat het in hoge mate schaars gedocumenteerd is. Om metadata op te sporen kan dan gebruik gemaakt worden van data profilering, zie § 5.3.
- Zorgen voor gelijke structuur voor applicatie interactie. De structuur voor bijna elke klantentabel bevat een naam, adres en telefoonnummer. Een data set van een leverancier bevat dezelfde attributen. Deze overeenkomst suggereert een onderliggend concept “partij”.
- Begrijpen en unificeren van semantiek om er voor te zorgen, dat objecten met verschillende onderliggende data types niet abusievelijk geconsolideerd worden.

2 Unificeren van data object semantiek

Dit omvat het begrijpen wat de gegevens betekenen, hoe die betekenis wordt overgebracht, hoe die betekenis data sets binnen de organisatie koppelt, hoe de gegevens gebruikt worden en benaderingen voor het omschrijven van semantiek als een attribuut voor het metadata raamwerk.

3 Identificeren en kwalificeren van master data

Door het toevoegen van karakteriseringinformatie voor elke metadata profiel van een data set, wordt er meer kennis toegevoegd aan het proces voor het bepalen van bron data sets, die geschikt zijn voor toevoegen aan de master data database, hetgeen de taak van de analist verlicht.

Het gebruik van een eenvoudig classificatieschema reduceert de complexiteit van gegevensvariatie en verlicht de beperkingen bij het vergelijken van meerdere metadata records.

De technieken en tools die gebruikt kunnen worden voor het bepalen van de bronnen van master data objecten zijn in essentie dezelfde type tools die later gebruikt worden om data te consolideren naar een master database. Het gebruik van data profilering, ontleding, standaardisatie en matching kan bijdragen aan het proces van het identificeren welke data sets (tabellen, bestanden, spreadsheets, etc) refereren naar welke master data objecten.

8 Data modellering

Hierbij wordt gekeken naar aspecten met betrekking tot het ontwikkelen van modellen voor master data, van extractie, consolidatie, vastleggen en verdelen.

Om elke master entiteit te kunnen onderscheiden moeten hun identiteit attributen gemodelleerd worden.

Identificatie attributen zijn die data elementen met hun waarden die gebruikt worden om één entiteit voorval te onderscheiden van een andere entiteit voorval.

Dat kan bijvoorbeeld door het toevoegen van een datum (timestamp) of een ordernummer bij een overeenkomst.

Onafhankelijk van welke architectuurbenadering wordt toegepast, de naamgeving is een set data attributen voor het formuleren van een “gesleutelde” master index, welke gebruikt kan worden om unieke attributen te vormen. De master index bevat een aantal gebruikelijke data attributen. De minimale set van attributen is de set van identificerende attributen.

Informatie delen en uitwisselen

Hierbij wordt gekeken naar de conceptuele aspecten van het modelleren van extractie en uitwisselingsmodel, als ook beheersbare gecentraliseerde opslag, welke gebruikt kan worden voor consolidatie, identiteitbeheer, publicatie en toegangsbeheer.

Er zijn daarvoor drie modellen:

- Extractie, gegevensuitwisseling en deling
- Consolidatie
- Vastlegging

9 Architectuur

In deze paragraaf wordt gekeken naar master data gebruikscenario's en architectuurvormen.

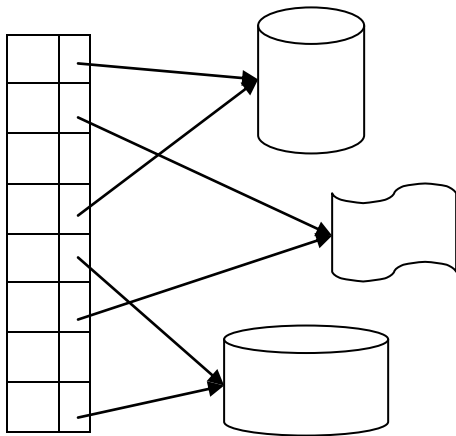
Deze beschrijven de criteria die gebruikt kunnen worden om te bepalen welk architectuur raamwerk het best bij de onderneming past.

Er zijn drie basis architecturen voor het beheren van master data.

- Register, een virtuele master data index
- Transactie Hub, een volledig geconsolideerde master data set voor gesynchroniseerde toegang
- Hybride / gecentraliseerde master, een combinatie van voorgaande benaderingen

1 Register

In deze architectuur wordt een “dunne” master data index gebruikt, welke de minimale set van identificeerbare attribuutwaarden pakt. Elke geïdentificeerde waarde van een gepresenteerde record wordt geïndexeerd met een identiteit zoektocht en match routine. De master index in het register bevat “lijnen” van elke unieke geïdentificeerde entiteit naar het bronsysteem.

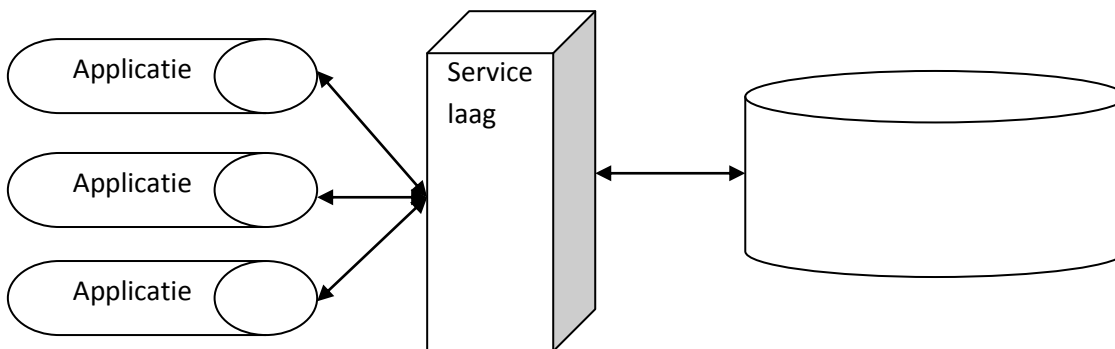


Figuur 6 Een register architectuur heeft een index, welke data mapt door het hele bedrijf.

2 Transactie Hub

Dit is een enkele opslagplaats die alle aspecten van master data beheert. Geen enkel gegeven wordt naar applicatiesystemen gepubliceerd en omdat er maar één kopie is, worden alle applicaties aangepast om direct gegevens uit te wisselen met de hub (letterlijk “naaf”, een kern waaraan spaken zitten die de verbindingen voorstellen naar de diverse applicaties).

Applicaties roepen toegangsservices aan om gegevens uit te wisselen met het data systeem, beheerd in de hub, en alle data levenscyclus acties worden gefaciliteerd via services, inclusief creatie, modificatie, toegang en afvloeiing.

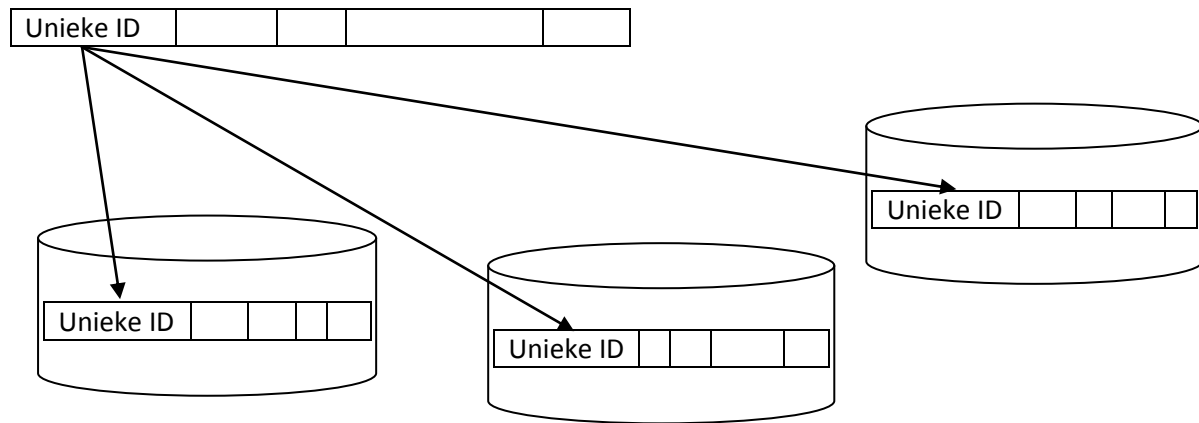


Figuur 7 Alle toegangen tot een transactie hub worden gefaciliteerd door een service laag.

3 Hybride / gecentraliseerde master

Dit is een middenweg ten opzichte van de voorgaande twee benaderingen.

Hier wordt een set van kern attributen, geassocieerd met elk master data model, gedefinieerd en beheerd in een enkel master systeem. De gecentraliseerde master opslagplaats is de bron voor het beheeren van de kern master data objecten, welke vervolgens gepubliceerd worden naar de applicatiesystemen. De gecentraliseerde master zorgt voor de standaard representatie van elke entiteit. Binnen elk afhankelijk systeem worden applicatie specifieke attributen lokaal beheerd, maar zijn teruggekoppeld naar de master via een gedeelde algemene primaire sleutel (ID).



Figuur 8 Gebruikelijke attributen worden samen beheerd in een gemapte master

Als de bedrijfsapplicaties blijven opereren op hun eigen lokale gegevenskopieën, die ook in de master worden beheerd, is er geen zekerheid dat de waarden in de master net zo actueel zijn als die in de applicatie data sets. Daarom is de hybride benadering goed voor omgevingen met een geharmoniseerde weergave van unieke master objecten, maar die geen hoge mate van synchronisatie vereisen.

Om een goede selectie te kunnen maken, welke benadering het meest geschikt is voor een onderneming, zijn er een zestal variabelen van belang:

- Aantal attributen
- Mate van consolidatie
- Mate van synchronisatie
- Toegang
- Service complexiteit
- Prestaties

Grofweg kan gehanteerd worden dat hoe groter het aantal attributen beheerd moet worden en hoe hoger de mate van consolidatie (dus hoe strakker de koppeling), hoe meer de architectuur wijst richting de transactie hub. Het tegenovergesteld geldt voor het register, terwijl de hybride er tussenin zit.

10 Consolidatie en integratie

Zodra het master data model helder is, moet bekeken worden hoe de gegevens van de geïdentificeerde bron naar een geconsolideerde master omgeving worden gebracht. Het gaat daarbij om de initiële migratie van data vanuit de bron en om de “lopende” integratie van data voorvallen die worden gecreëerd, veranderd of verwijderd van de master omgeving.

Data integratie omvat de processen voor het verzamelen van data uit verschillende bronnen en deze data toegankelijk maken voor verschillende applicaties.

10.1 Informatie delen

De benadering voor het delen van informatie is afhankelijk van de bedrijfseisen aan applicaties. Informatie wordt gedeeld op drie verschillende manieren met behulp van data integratie tools:

4. Extractie en consolidatie, met data integratie producten:
 - Transformatie, op basis van regels naar een juist formaat
 - Monitoring, conformiteit aan bedrijfseisen, validatie
 - Consolidatie, m.b.v. data kwaliteit technologieënStandaardisatie en publicatieservices aan de bron i.p.v. in de master vereenvoudigd het voorgaande.
5. Data verenigen, door het gebruik van een register met een index in de master database, kunnen gegevens bij de bron blijven en op verzoek getoond worden.
6. Data verspreiden, vanuit de master terug naar de bronnen.

10.2 Informatie identificeren

Onafhankelijk van welke architectuur wordt gekozen voor het opbouwen van een MDM systeem, er moet altijd een identificatie service zijn die wordt opgevraagd, telkens als een applicatie een master record wil benaderen.

De identificatie service wordt ondersteund door het gebruik van een register van entiteiten die geïndexeerd is met behulp van de waarden in de geïdentificeerde attributen.

De master identificatie service moet de “identiteit herkenning” techniek integreren, die sterke overeenkomsten kan vinden in waarden, wanneer een exacte match niet te vinden is. Deze techniek zit vaak in data kwaliteit tools.

10.3 Consolidatie technieken

Identiteit herkenning

Identiteit herkenning gaat over de bepaling dat twee of meer data representaties samengebracht kunnen worden tot één representatie van een uniek object.

Het gaat daarbij vaak over namen en adressen, maar ook veel over productnamen, productcodes, objectomschrijvingen, etc.

Dit proces omvat twee fases:

- Ontdekking en combineert data profilering activiteiten met manuele beoordeling van data.
- Overeenkomsten bekijken en matchen voor het koppelen van records.

De volgende data kwaliteit tool functies zijn van groot belang voor het master data integratie proces:

4. Ontleding en standaardisatie
De eerste stap in identiteit herkenning is het scheiden van relevante componenten van de te identificeren attribuutwaarden en deze te ordenen, zodat ze op een voorspelbare manier gemanipuleerd kunnen worden. Bijv. een naam splitsen in tokens: first_name, middel_name en last_name. Deze ontleding kan met behulp van speciale tools. Deze tools zorgen ervoor dat de data analist patronen kan definiëren en deze in software integreert. Daarna worden de componenten weer bij elkaar gevoegd tot een enkele representatie, welke andere data services weer kunnen manipuleren.

5. Data transformatie

Is gebaseerd op regels gestuurd door mappings van data waarden of met behulp van normalisatietabellen.

6. Matchen en koppelen

Een van de meest voorkomende data kwaliteitsproblemen gaat over:

- Als er meerdere records zijn die refereren naar dezelfde entiteit
- Als er een perceptie is dat een record niet bestaat voor een entiteit, maar er feitelijk wel is

Deze problemen (vals negatieven) kunnen opgelost worden met gelijkennis analyse, vaak gebaseerd op gewogen "matching bij benadering". Zo worden kleine afwijkingen herkend, gegevenswaarden gekoppeld en vervolgens ook het record gekoppeld.

10.4 Classificatie

Master data objecten kunnen vaak geclassificeerd worden in een logische hiërarchie of taxonomie. Classificatie heeft te maken met de applicatie eisen en hoe de entiteiten gebruikt worden. Het verbetert het consolidatieproces door de "zoekruimte" te reduceren bij het lokaliseren van overeenkomsten voor een specifieke zoekopdracht.

10.5 Consolidatie

Dit is het resultaat van de toegepaste taken van data integratie.

- Identificerende waarden zijn ontleed, gestandaardiseerd en genormaliseerd over alle data domeinen
- Deze zijn onderworpen aan classificatie en blokkeringschema's
- Deze zijn voorgelegd aan de unieke identificatie service voor het analyseren van duplicaten, kijken naar hiërarchische groepering, lokaliseren van een bestaand record of bepalen dat die niet bestaat.

Overeenkomstdrempels

Als er "matching bij benadering" wordt uitgevoerd, moeten de informatie architect en gebruikers definiëren wanneer twee waarden als gelijk moeten worden beschouwd, gespecificeerd in een drempelwaarde.

Overleving

Bij operationele samenvoeging kan het voorkomen, dat twee of meer records refereren naar dezelfde entiteit, maar conflicterende informatie bevatten. Om dan te bepalen, welke record wordt samengevoegd, wordt het "overlevingsproces" toegepast.

Dit proces moet gegevens of business rules inlijven in het consolidatieproces en deze regels moeten de kwaliteit van de data karakteriseren.

Fouten

Er zijn twee typen fouten te herkennen tijdens data integratie, vals positieven en vals negatieven.

- Vals positieven treden op als er twee records, die refereren naar twee verschillende entiteiten, foutief worden toegewezen naar dezelfde entiteit en per ongeluk worden samengevoegd in de enkele master.
- Vals negatieven treden op als er twee records refererend naar dezelfde entiteit geen match hebben, waarmee die records mogelijk dubbel in de master komen.

Ondanks alle inspanningen, ligt het voor de hand, dat deze fouten optreden tijdens de initiële datamigratie. Het is belangrijk om hierop voorbereid te zijn.

Bepaal de risico's en impact en communiceer deze naar belanghebbenden en/of stel een impact analyse op met mogelijke oplossingen.

Als er een fout opgestreden is, moet deze teruggedraaid kunnen worden. Daarvoor is het belangrijk om een historie op de bouwen. Elke keer als er een wijziging is geweest in de master record, moet het systeem deze wijziging in een logboek registreren evenals de bron, datum en tijd.

11 Synchronisatie

De beschikbaarheid van master data hangt af van de mate van synchronisatie tussen applicaties en de master data set evenals tussen de componenten van de master data set zelf. Hierbij zijn specifieke aspecten, de synchronisatie dimensies, van belang.

Synchronisatie dimensies:

- Tijdigheid
- Actualiteit
- Vertraging
- Consistentie (overall hetzelfde)
- Coherentie (samenhangend)
- Reproduceerbaarheid

Elk van de karakteristieken beïnvloed het onderhoud aan een consistente weergave van master data door de applicatie infrastructuur en elk van de verschillende MDM architecturen zullen voldoen aan de eisen binnen die verschillende dimensies op verschillende manieren.

Er zijn nog twee aspecten, waar rekening mee gehouden moet worden bij data synchronisatie, te weten:

1 Data afhankelijkheid

De volgorde van transacties hebben invloed op de uitkomst; elke transactie is afhankelijk van het resultaat van een vorige transactie. Bijvoorbeeld een prijs die wordt berekend aan een klant moet overeenkomen met de gestuurde offerte.

2 Seriële transacties

Elke transactie in operationele systemen bestaat vaak uit subtransacties en elke wordt uitgevoerd in een bepaalde volgorde. Op hetzelfde moment kunnen andere applicaties simultaan transacties uitvoeren, welke betrekking hebben op dezelfde data sets. Hoe houden we dit onder controle? Dit kan bijvoorbeeld door bij elke subtransactie de data set tijdelijk te blokkeren, totdat het systeem kan garanderen, dat de transactie consistent is verlopen.

Het blokkeren van een data object betekent, dat geen andere transactie dat object kan wijzigen en elke uitlezing ziet de waarde van voor die transactie.

Conceptuele modellen voor het delen van data

In paragraaf 9 zijn de drie verschillende MDM architecturen, register, transactie hub en hybride, reeds besproken.

Als we nu de synchronisatie dimensies bekijken van elke architectuur, wordt zichtbaar op welke vlakken de architecturen zich van elkaar onderscheiden.

	Register	Transactie hub	Hybride
Tijdigheid	+	+	?
Actualiteit	+	+	?
Consistentie	-	+	?
Coherentie	-	+	?
Reproduceerbaarheid	-	+	?

Tabel 1 Synchronisatie karakteristieken van verschillende modellen

Bij de hybride architectuur staan alleen vraagtekens, omdat deze tussenvorm van gegevensuitwisseling meerdere kopieën bevat van unieke gegevens. Hierbij sturen de applicaties de volgorde van transacties en daarnaast doorkruisen de gegevens meerdere onderliggende database platformen, hetgeen de complexiteit van recordblokkering verhoogt. Hierdoor is het moeilijk om de niveaus van de synchronisatie dimensies te bepalen.

12 Functionele service laag

Onder services wordt verstaan: software tools, interfaces of applicaties die bepaalde acties of bewerkingen in een systeem automatiseren. Elk proces heeft zijn eigen service en al die services komen samen in een service raamwerk.

Master data kunnen op twee manieren geïntegreerd worden in een service raamwerk:

Tactisch, een servicelaag om de transitie van applicaties naar de master opslag te faciliteren.

Strategisch, een hiërarchische set van informatie services ter ondersteuning van snelle en efficiënte ontwikkeling van geautomatiseerde applicaties.

Beide noodzakelijke aspecten worden opgelost door een SOA, Service Orientated Architecture.

Een belangrijk aspect van een service georiënteerde benadering is dat de eisen aan een bedrijfsapplicatie de behoefte stuurt om "mogelijkheden" te identificeren en consolideren, welke door het bedrijf heen beheerd worden, op een manier dat het gedeeld en hergebruikt kan worden binnen de grenzen van een bestuurlijke omgeving.

Een organisatie die SOA toepast, is er een die zich aanpast van levering van applicaties naar een van levering van bedrijfsoplossingen.

Identificatie van services

Het identificeren van applicatie functionaliteit en categoriseren hoe die functionaliteit te abstraheren in een service laag, toont hoe de applicaties de data objecten raken die gerelateerd zijn aan master object typen.

Dit proces wordt gebruikt om die functionaliteit te standaardiseren voor vereenvoudiging en hergebruik.

Er zijn applicaties die gebruik maken van dezelfde set van kern mogelijkheden. Het doel is om deze mogelijkheden te identificeren in relatie tot hoe ze gebruik maken van master data objecten en deze te consolideren naar een set van services.

Met een service georiënteerde benadering is er een grotere snelheid en flexibiliteit in de mogelijkheid om te reageren op veranderingen in de bedrijfsomgeving. Het integreren van nieuwe mogelijkheden in de bedrijf service laag, zorgt er voor dat alle applicaties daar profijt van hebben, hetgeen ontwikkelingskosten reduceert en dat nieuwe mogelijkheden sneller operationeel zijn.

Enkele voorbeelden van kern services zijn:

- Levenscyclus services (creëren, lezen, wijzigen, afvloeien)
- Governance (audit, veiligheid, legitimatie, toegangscontrole)
- Integratie (integratie, data kwaliteit, metadata management)
- Consolidatie (modellering, match/samenvoegen, identiteit en hiërarchie management)
- Workflow/regels (zoeken en aanpassen, regels, workflow, invoeren en synchroniseren)

Er is een symbiotische relatie tussen MDM en SOA: een service georiënteerde benadering voor het ontwikkelen van bedrijfsoplossingen steunt op het bestaan van een master data opslag en het succes van een MDM programma hangt af van het gebruik van master object toegang en manipulatie services.

Als het programma is gelanceerd en ontwikkeld, vereenvoudigd het gebruik van een multiple service laag de toepassing van een data object levenscyclus evenals de transitie van een diffuus , verdeelde applicatie architectuur naar een geconsolideerde master applicatie infrastructuur.

De flexibiliteit, geleverd door een SOA, zorgt voor de fundamentele organisatorische veranderingen die nodig zijn voor gedeelde informatie op de meest effectieve wijze.

13 Management richtlijnen

Deze paragraaf behandelt een high-level overzicht van activiteiten voor het ontwikkelen van een MDM programma. Een van de moeilijkste delen van het totale proces is om te bepalen waar te beginnen. Het is vaak effectiever om externe experts in te schakelen tijdens de initiële fasen van het project, om daarmee zaken te bespoedigen.

Om een MDM programma uit te voeren, moeten de volgende stappen genomen worden:

- 1 Rechtvaardiging verkrijgen voor MDM van het hoger management
Stel een business case op voor alle betrokkenen in de organisatie.
- 2 Ontwikkel een MDM roadmap en uitrol plan
 - 2a De implementatie van een complex programma vereist strategische planning, samen met details voor de gefaseerde uitvoering. Het opstellen van een roadmap voor het ontwerp en ontwikkeling van de infrastructuur en processen ter ondersteuning van MDM, schept duidelijkheid aan het management.
De verschillende stappen zijn in paragraaf 3 al aan de orde geweest.
 - 2b Het uitrol plan wordt ontworpen om de stappen uit de roadmap uit te voeren.
Dit kan de volgende stappen omvatten:
 - Voorbereiding, bepaal of een MDM programma nodig is
 - Bepaal haalbaarheid en de doel architectuur
 - Initiële inzet, ontwikkel data modellen die voldoen aan de behoeften van de organisatie
 - Initiële uitvoering, beleid en protocollen zijn opgesteld
 - Transitie, toepassen en integreren van services
 - Operationeel, ondersteuning en onderhoud
- 3 Beschrijf rollen en verantwoordelijkheden
Dit is in paragraaf 2 beschreven.

4 Project planning

De volgende activiteiten zijn reeds in voorgaande paragrafen aan bod gekomen.
Voor gedetailleerdere informatie kan de betreffende paragraaf er op nageslagen worden.
Bij dit project plan staan steeds de specifieke mijlpalen en opleveringen opgesomd.

Het project plan omvat de volgende activiteiten:

1. Business Proces Modellen en gebruik scenario's

Opleveringen:

- Een set van bedrijfsprocesmodellen voor elke activiteit die bedoeld zijn voor opname in het MDM programma
- Een set van scenario's voor de bedrijfsapplicaties voor de master data

2. Identificeer initiële data sets voor master integratie

Opleveringen:

- Een lijst van applicaties die deelnemen in het MDM programma en de master data object types, welke kandidaten zijn voor mastering.

3. Data governance

Opleveringen, o.a.:

- Management handleiding voor data governance
- Processen voor definiëren, toepassen en verzamelen van kwantificeerbare metingen
- Definities van acceptabele drempelwaarden voor monitoren van metingen
- Processen voor operationele data governance en beheer
- Selectie en aankoop van tools

4. Metadata

Opleveringen:

- Gestelde eisen aan metadata management tools
- Evaluatie en selectie van metadata tools
- Processen om consensus te verkrijgen over bedrijfsdata definities
- Processen voor het documenteren van het gebruik van data elementen en service

5. Object analyse

Opleveringen:

- Cross-system data analyse profielen
- Metadata profielen
- Business rules
- Bron-doel mappings
- Extractie en transformatie richtlijnen

6. Object modellering

Opleveringen:

- Een model voor representaties van data geëxtraheerd uit originele bronnen
- Een model voor het data consolidatieproces
- Een model voor opslag van geselecteerde master data attributen

7. Data kwaliteit management

Opleveringen:

- Bepaling van bedrijfseisen voor data kwaliteit tools

- Formalisatie en toepassing van operationele data kwaliteit technieken
 - Formalisatie en toepassing van operationele data governance procedures
 - Aankoop en uitrollen van data kwaliteit tools
8. Data extractie, delen, consolideren en invoeren
Opleveringen, o.a.:
- Identificatie van bedrijfseisen voor data integratie en consolidatie tools
 - Aankoop van data integratie en consolidatie tools
 - Ontwikkeling van regels voor consolidatie en “overleving”
 - Toepassing van consolidatie services
9. Architectuur
Opleveringen:
- Evaluatie van master data gebruik scenario’s
 - Toetsing van synchronisatie eisen
 - Evaluatie van vendor data modellen
 - Selectie van een vendor oplossing
10. Master data services
Opleveringen:
- Kwalificatie van master service lagen
 - Ontwerp, implementatie en ontwikkeling van master data services
 - Testen van applicaties
11. Transitieplan
Opleveringen:
- Ontwikkeling van een migratie/transitie proces
12. Onderhoud
Opleveringen:
- Stel de minimale niveaus van service op voor data kwaliteit, applicatie prestaties en synchronisatie
 - Pas monitoring en rapportage toe van het service niveau

14 Trends en richtingen

MDM oplossingen zijn nog relatief nieuw en de trendanalyses in de markt geven diverse richtingen aan. Er wordt onderscheid gemaakt trends in de markt en trends op het gebied van technische mogelijkheden.

Trends in de markt

- Er blijven twee MDM domeinen dominant, namelijk partij/klant- en productdomein
- Meer initiatieven op het gebied van het ophalen van externe data en dit intern onderhouden
- Externe leveranciers van referentie data gaan een prominentere rol spelen
- Zowel real-time als batch processen zullen naast elkaar blijven bestaan
- Grote MDM leveranciers zullen steeds meer kleinere opkopen
- De mogelijkheden van data governance zullen verder groeien en deels worden geïntegreerd in data hub systemen, inclusief metingen. Organisaties die deze metingen hanteren zullen een data governance groep vormen als een business unit met gedefinieerde prestatiedoelen.

Trends in technische mogelijkheden

- MDM krijgt steeds meer mogelijkheden, door ontwikkeling van kern functionaliteiten, nieuwe technologieën (geavanceerde data kwaliteit en matching), virtualisatie en cloud computing
- MDM komt steeds meer beschikbaar als een open source oplossing, wat leidt tot betere, snellere en goedkopere ontwikkeling van oplossingen en modellen, omdat alles op basis van brede samenwerking is ontstaan.
- MDM zal ook steeds meer ongestructureerde en semi-gestructureerde data ondersteunen
- Er komen steeds meer geïntegreerde MDM oplossingen (totaalpakket)

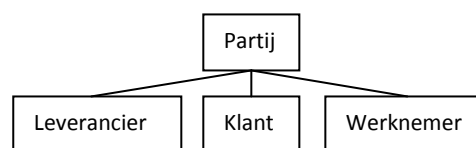
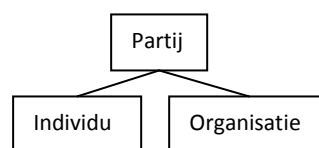
Bijlage 4 MDM Maturity Matrix

Technische component		Initieel	Reactief	Beheerst	Proactief	Strategisch
Architectuur	Masterdata model (structuur) De data attributen van het kernmodel is een superset van alle belangrijke attributen van elk applicatiemodel. Ook het formaat en de structuur van die attributen moeten alle formaten en structuren over de diverse modellen ondersteunen.	Geen master data modellen Een verzameling van databestanden in diverse vormen Beperkte bedrijfsconsolidatie van representatieve modellen	Applicatie architecturen zijn gedefinieerd voor elke bedrijfsapplicatie Pogingen tot het verzamelen van databestanden in een enkele centrale opslag	Gedefinieerd kern datamodel Fundamentele architectuur voor gedeeld master data raamwerk	Master modellen zijn opgesteld	
	Systeemarchitectuur (kracht) Focus op data levenscyclus aspecten en ondersteunt generiek een hoger niveau servicelaag		Beoordeling van opties voor het delen van informatie	Operationeel raamwerk voor low-level masterdata levenscyclus activiteiten	Capabel om van index raamwerk naar transactie gebaseerd MDM raamwerk te gaan Gecentraliseerd beheer van bedrijfs metadata	Vervolmaak transactie integratie beschikbaar voor interne applicaties
	Service laag (controle) Hogere eisen aan data object services t.b.v. synchronisatie, toegang, integratie, etc.		Eerste onderzoek naar low-level applicatie services	Services gedefinieerd voor integratie met masterdata centrale database	SOA ingevoerd voor applicatie architectuur	Interfaces maken recht toe bewerking mogelijk m.b.t. master data
Governance overzicht verkrijgen	Algemeen	Bewust van een benodigd overzicht		Business gerelateerd databeheer	Enterprise data governance programma aanwezig	Data governance door de hele organisatie verzekert hoge kwaliteit van informatie uitwisseling
	Data standards			Processen aanwezig	Bedrijfsbreed aanwezig	
	Metadata management		Gebruik van externe applicaties om metadata te beheren		Bedrijfsbreed aanwezig	
	Data kwaliteit	Beperkte data cleansing door applicaties, voor specifieke doelen (bijv. adres standaardisatie)	Introductie van data quality management voor ontleding, standaardisatie en consolidatie	Beleid en procedures voor data quality management, data quality tools aanwezig, data quality issues tracking	Proactieve monitoring voor data kwaliteits controles in het governance programma	
	Gegevensbeheer Operationele aspecten	Geen gedefinieerde eigenaarschap en beheerdersmodellen		Line of business niveau		
Management (operations)	Identiteit Unieke sleutel entiteiten. Identity management omvat de beschrijving van de identificerende attributen benodigd voor unieke identificatie, samen met zoek en match mogelijkheden	Identity management m.b.v. applicaties, indien nodig		Identity management wordt bij veel applicaties gebruikt Identity management gecentraliseerd in master index	Unieke identificatie van alle master objecten Identity management volledig en bedrijfsbreed geïntegreerd	Naadloze identity management van alle data objecten gesynchroniseerd naar zowel interne als externe representaties
	Hierarchie Gaat om afkomst van data (herleiden), het samenvoegen van meerdere records in een enkele representatie en de onderlinge verbindingen van objecten tussen verschillende systemen			Geïdentificeerde hiërarchien worden gebruikt bij analytische applicaties	Volledige cyclus hiërarchie management ondersteund zowel analytische als operationele activiteiten Hiërarchie management maakt roll-back van vals positieve consolidatiefouten mogelijk	
	Migratieplan			Migratieplan aanwezig voor geselecteerde applicaties		Migratie van legacy applicaties compleet
	Administratie en configuratie Technisch team heeft tools en processen nodig voor configuratie en vastlegging	Enige applicatie configuratie, maar niet gecoördineerd door gecentraliseerd management	Opleiding voor bedrijfsbrede uitrol van tools en technologie, centrale administratie van metadata, middelen zijn toegewezen voor het beheer van nieuwe tool sets	Geavanceerde configuratie en administratie van applicatie gebruik van masterdata		
Identificatie	Zoek en bepaal unieke omschrijving Ontdekken (combinatie van data profilering en beoordeling), overeenkomsten vinden en matchen	Beperkt gebruik van identity management door business	Identiteits zoektocht en match ter reductie van duplicaten en voor rudimentaire record koppeling	Identiteits zoektocht en match service beschikbaar voor alle applicaties	Services voor data levenscyclus integreren Identiteit zoektocht, match en samenvoeging	Identity resolution diensten worden zichtbaar gemaakt in de organisatie
	Koppeling records Refereren naar dezelfde entiteit, toegepast bij grotere hoeveelheid records uit verschillende systemen			Record koppeling geïntegreerd in de MDM service laag	Alle data levenscyclus handelingen zijn gestructureerd bovenop de samenvoeging en consolidatie services	Bedrijfsprestatie direct gekoppeld aan master dimensies
	Samenvoegen en versterken/completeren Consolidatie	Pogingen tot consolidatie van klantgegevens, benodigd voor applicaties		Regels voor samenvoeging en consolidatie zijn gestandaardiseerd en beheerd onder gecentraliseerde controle, processen zijn vastgesteld en herhaalbaar	Consolidatie gebeurt op de achtergrond	
Integratie	Applicatie integratie en synchronisatie laag Gegevens moeten eenvoudig samengevoegd kunnen worden, maar ook eenvoudig beschikbaar zijn in het masterbestand	Verschillende kopieën van referentie data	Initieel onderzoek naar dataconsolidatie voor nieuw ontwikkelde analytische applicaties	Processen voor terugwaarde integratie naar datastuwende applicaties	Synchronisatie is volledig ingebed in de levenscyclus services	Alle applicatie ontwikkeling wordt gedreven door BPM en hun interactie met kern master object modellen
	MDM component laag Ter vereenvoudiging van incrementele ontwikkeling van applicaties, samenvoegen van applicatiefunctionaliteit d.m.v. servicelaag als aanvulling op de "meervoudige applicatie benadering"	Beperkt hergebruik van data Geen hergebruik van applicatie services	Datawarehouse voor centrale opslag van masterdata Beperkte of geen integratie met datastuwende applicaties	Beschrijving van beschikbare component services voor applicatie integratie Services voor synchronisatie tussen applicaties en masterdata services	Component laag ondersteund applicatie integratie op master object niveau SOA stuurt bedrijfs applicatie integratie	
Business Process Management Hoogste abstractieniveau	Business Process Integratie Het Business process model communiceert de juiste details van de bedrijfsprocessen (invoer, controles, triggers, uitvoer)	Beperkte of geen betrokkenheid vanuit de business, behalve op het hoogste niveau van definitie van eisen	Conceptuele BPM beschreven, analytische applicatie integratie van geconsolideerde data	Fundamentele verbinding tussen bedrijfsapplicaties en kern data objecten	Geautomatiseerde bedrijfsprocessen, bedrijfslogica wordt hergebruikt	Bedrijven sturen volledig applicatie ontwerp en ontwikkeling
	Business rules Combineert de evaluatie van (gedeelde) data object waarden met de beschreven waarden in gedefinieerde controles		Initieel gebruik van business rules toegepast in applicaties	Integratie van business rules met masterdata operations	Business rules worden geïntegreerd in validatiesoftware en beschikbaar gemaakt op bedrijfsproces niveau	Applicaties integreren validatiesoftware op grote schaal
	MDM business component laag Toepassing van slimme, herbruikbare business services (zoals voorspellende analyses)			Bedrijfsproces analisten nemen deel aan het opstellen van master data engineering eisen	Bedrijfsanalisten integraal naar applicatie ontwikkeling Persoonlijke klantrelaties	MDM maakt geïntegreerde voorspellende analyses mogelijk Data profielen (klant of leverancier profielen) worden beheerd binnen master data centrale database

Bijlage 5 Lijst met afkortingen en begrippen

AD	Active Directory
BIS	Business + Information Solutions
BODS	Business Objects Data Server
CAB	Change Advisory Board
CI	Configurable Item, zoals hardware, software, server
ETL	Extract, Transform, Load
ITIL	Information Technology Infrastructure Library
OVSD	Open View Service Desk (HP applicatie)
PIRT	Pepsico International Reference Standard
PDM	Product Data Management
PLM	Project Lifecycle Management
RFC	Request For Change (Verzoek)
SD	Service Desk
SOA	Service Orientated Architecture
SR	Service Request

Taxonomieën manier waarop bedrijfsconcepten worden beschreven, gekarakteriseerd en geclassificeerd, hiërarchische ordering



WER West European Region

Bijlage 6 Informatie theorie

Informatie theorie

Een voorbeeld van een wetenschappelijke benadering van data kwaliteit voor MDM is gebaseerd op Informatie theorie.

Stel, dat er een benchmark is ontwikkeld met een “gouden”record set voor “klant”, A.

Er zijn diverse gegevensbronnen binnen het bedrijf die klantattributen bevatten, één daarvan is a.

De data kwaliteit van a wordt gekwantificeerd in relatie tot A.

Het doel is om metingen te definiëren die de brongegevens a vergelijken met de benchmark set A.

Voorbeeld:

Referentie “A”			Record “a”		
ID	Naam	Plaats	ID	Naam	Plaats
1	Jansen	Utrecht	1	Janssen	Utrecht
2	de Bond	Tiel	2	de Bond	Vlissingen
3	Bartels	Zwolle	3	Bartels	Zwolle

Een voorbeeld van bovenstaande is de kans dat de achternaam “Jansen” in referentie A en “Janssen” in bron a allebei refereren naar dezelfde klant en zijn gekoppeld aan een enkele klant identificatie. De waarden van a en A hoeven niet gelijk te zijn, maar wel consistent en herhaalbaar. Dat is waar het bij data kwaliteit om draait!

Informatie theorie is gebaseerd op kansberekening, waarbij informatie entropie een belangrijke term is.

Informatie entropie is een maat voor onzekerheid geassocieerd met de voorspelbare waarde van informatie inhoud. De hoogste informatie entropie is als de dubbelzinnigheid of onzekerheid van de uitkomst het grootst is.

In het algemeen kan gezegd worden, dat de entropie een maat is voor voorspelbaarheid en onzekerheid van het gedrag van een systeem. Een lage entropie wil zeggen, dat een systeem een eenvoudige voorspelbaarheid heeft en ordelijk is, terwijl een hoge entropie staat voor een moeilijk te voorspellen systeem, welke chaotisch is.

Bij het gooien van een munt, zijn er twee uitkomsten mogelijk, 0 of 1. De kans voor beide uitkomsten is 0,5. In dit geval is de maximale entropie = 1.

Bij twee munten is de kans 0,25 en is de maximale entropie=2

$E(A)$ is de entropie van de referentie en $E(a)$ die van de bron.

Een nieuwe entropie (E_2) kwalificeert deze functie als een goede kandidaat voor het kwantificeren van de data kwaliteit van a ten opzichte van A.

De data kwaliteit (DQ) kan nu gedefinieerd worden als $DQ = E_2/E(A) \cdot 100\%$.

Het voordeel van deze meting is dat het niet alleen kwantificeerbaar is, maar ook dat het gebruikt kan worden voor het bepalen van de data kwaliteit van delen van attributen (tokens), een functie die nieuwe inzichten verschaft in de oorzaak van de meest significante data kwaliteitsproblemen.

Dit is slechts een introductie van de informatie theorie. Meer informatie is te vinden in het boek van Berson, Dubov (2011) p. 412.

Bijlage 7 Mapping Oracle to PDM

DBASE	TAB	Field Name	Length	Unit	Field	Example	Meaning	EACH (CONSUMER UNIT)	NO-EACH (TRADE UNIT)	Match	Source
central	identification 1	Target Market Country Code						depending on w hich market is selected -NL -BE (including LUX)	depending on w hich market is selected -NL -BE (including LUX)		Data Collectsheet
central	identification 1	information provider	13,0		infoprovider			8710398000002	8710398000002		Fixed value
central	identification 1	name of information provider	35A		infoproviderartname			PepsiCo Nederland BV	PepsiCo Nederland BV		Fixed value
central	identification 1	product group	100A		productgroup	1701440483	Juice/nectar 100% - Ready to Drink	see tab 'productgroup'	see tab 'productgroup'		Bus rules
central	identification 1	brand	250A		ClassName	LOOZA		Brand as printed on sku: see tab "Brand"	Brand as printed on sku: see tab "Brand"	X	Oracle
central	identification 1	unique ID	14,0	CU	ArticleNr			UPC EAN13 14digits w ith leading 0		X	Oracle
			14,0	TU	ArticleNr	05410188563528			UPC EAN13 14digits w ith leading 0	X	Oracle
central	identification 1	GTIN number	14,0	CU	EANNr			UPC EAN13 14digits w ith leading 0			copy
			14,0	TU	EANNr	05410188563528			UPC EAN13 14digits w ith leading 0		copy
central	identification 1	article name(GTIN-name L1)	140A	CU	Name	LO A ACE ORIGINAL 20CL REG		UPC description		x	Oracle
			140A	TU	Name	LO A ACE ORIGINAL 24X20CL REG			UPC description	X	Oracle
central	identification 2	product type	2A	CU	producttype			EA-each			Bus rules
			2A	TU	producttype	CA	Case		CA-case (also for a display) PK-package (for midlevel)		Data Collectsheet
central	identification 2	Language 1	VS		language			dutch,flemish	dutch,flemish		Fixed value
central	identification 2	Language 2	VS		languageL2			english	english		Fixed value
central	identification 2	Language 3	VS		languageL3			french	french		Fixed value
central	identification 2	Language 4	VS		languageL4			german	german		Fixed value
central	identification 2	GTIN-name L2	140A		GTINNameL2	copy from GTIN-name L1		copy from GTIN-name L1	copy from GTIN-name L1		copy
central	identification 2	GTIN-name L3	140A		GTINNameL3	copy from GTIN-name L1		copy from GTIN-name L1	copy from GTIN-name L1		copy
central	identification 2	GTIN-name L4	140A		GTINNameL4	copy from GTIN-name L1		copy from GTIN-name L1	copy from GTIN-name L1		copy
central	global	Brand owner GLN	13,0		owningorg			5410188000016	5410188000016		Fixed value
central	global	Brand owner name	35A		owningorganisationname			Tropicana Looza Benelux	Tropicana Looza Benelux		Fixed value
central	global	Pack	6,0	CU	pack			1			Fixed value
			6,0	TU	pack	24	number of CU's		number of consumer units	X	Oracle
central	global	Base unit indicator	B	CU	baseunitindicator			yes			Fixed value
			B	TU	baseunitindicator				no		Fixed value
central	global	Consumer unit indicator	B	CU	istradeitemCU			yes			Fixed value
			B	TU	istradeitemCU				no, if the product type is a PK and is sold to consumers the yes		Data Collectsheet
central	global	Own label/private label	50A		additionalU01			no	no		Fixed value
central	global	Variable weight trade item	B		variableweighttradeitem			no	no		Fixed value
central	global	Is Trade Item a Display Unit?	V	CU	istradeitemDU			ONWAAR			Fixed value
			V	TU	istradeitemDU				FALSE, When "packaging logistical unit" = "DIS" then TRUE	X	Oracle
central	global	net content 1	30,15	CU	sizemetric			net content		X	Oracle
			30,15	TU	sizemetric	4800			net content		Calc
central	global	Net content UOM	2A		sizemetricUOM	ML	Milliliter	ML (milliliter)	ML (milliliter)		Fixed value
central	global	Packaging type 1	3A	CU	packagingtype			see tab 'packag,type+packag,mat model'		X	Oracle

DBASE	TAB	Field Name	Length	Unit	Field	Example	Meaning	EACH (CONSUMER UNIT)	NO-EACH (TRADE UNIT)	Match	Source
				TU					yes		Fixed value
central	general	Dispatch unit indicator	B	CU	dispatchunitindicator			no			Fixed value
				TU					yes		Fixed value
central	general	Alternate item classification code 1	29A		alternateitemcode	10000220	Juice Ready To Drink (Shelf Stable)	see tab 'GPC'	see tab 'GPC'		Bus rules
central	general	Alternate item classification code 2	VS		alternateitemscheme			GPC	GPC		Fixed value
central	general	Alternate item number	18A	CU	alternateitemnumber			EAN13			copy
			18A	TU	alternateitemnumber				EAN13		copy
central	general	Alternate item number UOM	VS		alternateitemidagency			90-assigned by manufacturer	90-assigned by manufacturer		Fixed value
central	measures	Depth	30,15	CU	depth			measurement as defined see tab 'depth examples'			Data Collectsheet
			30,15	TU	depth	387			measurement as defined see tab 'depth examples'		Data Collectsheet
central	measures	Depth UOM	2A		depthUOM	MM	Millimeter	Mm-millimeter	Mm-millimeter		Fixed value
central	measures	Height	30,15	CU	height			measurement as defined			Data Collectsheet
			30,15	TU	height	193			measurement as defined		Data Collectsheet
central	measures	Height UOM	2A		heightUOM	MM	Millimeter	measurement as defined	measurement as defined		Fixed value
central	measures	Width	30,15	CU	w idth			measurement as defined			Data Collectsheet
			30,15	TU	w idth	256			measurement as defined		Data Collectsheet
central	measures	Width UOM	2A		w idthUOM	MM	Millimeter	Mm-millimeter	Mm-millimeter		Fixed value
central	measures	Diameter	30,15		diameter			if packaging type=BOT / CAN then measurement as defined	blank		Data Collectsheet
central	measures	Diameter UOM	VS		diameterUOM			Mm-millimeter	blank		Fixed value
central	measures	volume	30,15	CU	volume			calc : (Depth X Height X Width) /1000			Data Collectsheet
			30,15	TU	volume	19120,896			calc : (Depth X Height X Width) /1000		Data Collectsheet
central	measures	volume UOM	2A		volumeUOM	CC	Centiliter	Cc-cubic centimeters	Cc-cubic centimeters		Fixed value
central	measures	Gross weight	30,15	CU	grossw eight			measurement as defined			Data Collectsheet
			30,15	TU	grossw eight	11300			measurement as defined	X	Oracle
central	measures	Gross weight UOM	2A		grossw eightUOM	GR	Gram	Gr-gram	Gr-gram	X	Fixed value
central	measures	Net weight	30,15	CU	netw eight			nett w eight			Data Collectsheet
			30,15	TU	netw eight	4800			nett w eight (#cons unitsXnett w eight cons unit)		Data Collectsheet
central	measures	Net weight UOM	2A		netw eightUOM	GR	Gram	Gr-gram	Gr-gram		Fixed value
central	Product description	Functional Name L1	70A		productname	JUICE - READY TO DRINK		Should not contain Brand or Subbrand -> see tab "Funct name + subbrand"	Should not contain Brand or Subbrand -> see tab "Funct name + subbrand"		Bus rules
central	Product description	Functional Name L2	70A		productnameL2	JUICE - READY TO DRINK	copy from Functional name L1	copy from Functional name L1	copy from Functional name L1		copy
central	Product description	Functional Name L3	70A		productnameL3			copy from Functional name L1	copy from Functional name L1		copy
central	Product description	Functional Name L4	70A		productnameL4			copy from Functional name L1	copy from Functional name L1		copy
central	Product description	Product description L1	140A		productdescription	LOOZA Sinaas-		UPC-description + additional (fi	UPC-description + additional (fi		Data Collectsheet

DBASE	TAB	Field Name	Len	er	Field	Example	Meaning	EACH (CONSUMER UNIT)	NO-EACH (TRADE UNIT)	Ma	
central	Product description	Short descr/POS desript L1	140A		productdescription			copy from GTIN-name L1 (first 35 characters)	copy from GTIN-name L1 (first 35 characters)		copy
central	Product description	Short descr/POS desript L2	140A		productdescriptionL2			copy from GTIN-name L1 (first 35 characters)	copy from GTIN-name L1 (first 35 characters)		copy
central	Product description	Short descr/POS desript L3	140A		productdescriptionL3			copy from GTIN-name L1 (first 35 characters)	copy from GTIN-name L1 (first 35 characters)		copy
central	Product description	Short descr/POS desript L4	140A		productdescriptionL4			copy from GTIN-name L1 (first 35 characters)	copy from GTIN-name L1 (first 35 characters)		copy
central	Product description	Variant L1	70A		variant	ACE ORIGINAL	flavour	Flavour : Should not contain Brand or Subbrand - Guideline is description on packaging->Dutch	Flavour : Should not contain Brand or Subbrand - Guideline is description on packaging->Dutch	X	Oracle
central	Product description	Variant L2	70A		variantL2	ACE ORIGINAL		Flavour : Should not contain Brand or Subbrand - Guideline is description on packaging->English	Flavour : Should not contain Brand or Subbrand - Guideline is description on packaging->English		copy
central	Product description	Variant L3	70A		variantL3			Flavour : Should not contain Brand or Subbrand - Guideline is description on packaging->French	Flavour : Should not contain Brand or Subbrand - Guideline is description on packaging->French		copy
central	Product description	Variant L4	70A		variantL4			Flavour : Should not contain Brand or Subbrand - Guideline is description on packaging->German	Flavour : Should not contain Brand or Subbrand - Guideline is description on packaging->German		copy
central	Packaging	Pricing on product	B		pricingonproduct			no	no		Fixed value
central	Packaging	Packaging marked returnable	B		packagingmarkedreturnable			no	no		Fixed value
central	Packaging	Product marked reXclable	B		productmarkedrecyclable			no	no		Fixed value
central	Packaging	Product marked with ingredients	B	CU	productmarkedingredients			yes			Fixed value
			B	TU	productmarkedingredients				no		Fixed value
central	Packaging	Packaging marked expiration date type	VS		additionalTMX001	Expiry date		"Expiry date", if support then blank	"Expiry date", if support then blank		Data Collectsheet
central	Packaging	net content declared	B		netcontentdeclarationindic			yes, if support product then no	yes, if support product then no		Data Collectsheet
central	Packaging	Packaging material code L1	2,0	CU	packagingmaterialcode	79	Plastic	see tab 'package type + mat code'			Data Collectsheet
			2,0	TU	packagingmaterialcode				see tab 'package type + mat code'		Data Collectsheet
central	Packaging	Material code agency 1	3A		materialcodeagency			E	E		Fixed value
central	Packaging	Packaging material composition 1	30,15	CU	packagingw eight	251		source is data collect			Data Collectsheet
			30,15	TU	packagingw eight				calculate gross w eight-net w eight		Data Collectsheet
central	Packaging	Packaging material composition UOM	2A		packagingw eightUOM	GR	Gram	GR-gram	Gr-gram		Fixed value
central	Handling and delivery	import classification value	70A		additionalHDL007	20099059		Intrastat code (8-digit code w ithout leading 0)	Intrastat code (8-digit code w ithout leading 0)	X	Oracle
central	Handling and delivery	import classification type	VS		additionalHDL033			intrastat	intrastat		Fixed value
central	Handling and delivery	storage/handling temperature min	30,15		handlingtempmin			if Name contains "C" then 2, otherw ise leave blank	if Name contains "C" then 2, otherw ise leave blank		Calc
central	Handling and delivery	storage/handling temperature max	30,15		handlingtempmax			if Name contains "C" then 7, otherw ise leave blank	if Name contains "C" then 7, otherw ise leave blank		Calc
central	Handling and delivery	storage/handling temperature UOM	VS		handlingtempUOM			if min/max temperature <> blank then CELSIUS	if min/max temperature <> blank then CELSIUS		Calc
central	Subarticles	unique ID	50A		ArticleNr			blank	GTIN of subarticle		copy
central	Subarticles	amount	50A	CU	subarticle			blank			Fixed value
			50A	TU	subarticle				how many?		copy
local	dates	Start availbilty date	10,0		startavaililtydate	2011-02-21	YYYY-MM-DD	current date	current date		Fixed value
local	dates	first order date	D		additionalUCC02			blank	copy of "start availability date"		copy
local	dates	first delivery date	D		additionalCRE029			blank	copy of "start availability date"		copy
local	dates	End availabilty date	10,0		endavailabilitydate	2012-03-08	YYYY-MM-DD	blank	blank		Fixed value
local	dates	Discontinued date	10,0		discontinueddate	2012-03-08	YYYY-MM-DD	blank	blank		Fixed value
local	Handling and delivery	Pallet code	VS	CU	palletcode			blank			Fixed value

DBASE	TAB	Field Name	Len	er	Field	Example	Meaning	EACH (CONSUMER UNIT)	NO-EACH (TRADE UNIT)	Match	Source
local	Handling and delivery	nonGTIN pallet gross weight	30,15	CU	nonGTINpalletgrossweight			blank			Fixed value
				TU	nonGTINpalletgrossweight	1109800			nonGTIN pallet-number of items per pallet X gross weight + 25000gram(euro BE) of 28000gram(chep NL)		Calc
local	Handling and delivery	nonGTIN pallet gross weight	2A	CU	nonGTINpalletgrossweightU			blank			Fixed value
				TU	nonGTINpalletgrossweightUOM	GR	Gram		Gr-gram		Fixed value
local	Handling and delivery	Stacking factor	10,0	CU	stackingfactor			blank			Fixed value
				TU	stackingfactor				if palletheight <130cm then 2 else 1		Data Collectsheet
local	Handling and delivery	minimum product lifespan from production	3,0		minproductlifespan	392	days	Data Collect Sheet	Data Collect Sheet	X	Oracle
local	Handling and delivery	Minimum product life span from arrival	3,0		additionalWRE23	252	days	Data Collect Sheet	Data Collect Sheet		Data Collectsheet
local	Ordering and selling	order quantity multiple	10,0		orderqtymultiple			blank	blank		Fixed value
local	Price and tax	Tax rate	30,15		taxrate	6		%VAT, beverage = 6	%VAT, beverage = 6		Fixed value
local	Price and tax	Tax rate tax type code	14A		taxtype	VAT		VAT	VAT		Fixed value
local	Price and tax	(Tax Rate) Tax Type Agency	3A		additionalCRE080	BE		if local = BE -> BE / if local = NL ->NL	if local = BE -> BE / if local = NL ->NL		Bus rules
local	Price and tax	Tax Rate/Tax Applicability	VS		additionalCRE156	Low		Low	Low		Fixed value
local	Publish to	Publish to GLN 1	13,0	CU	publishtoGLN1			blank			Fixed value
			13,0	TU	publishtoGLN1				enter GLN code from Trading partner(s)		Manual, no interface
		Publish to GLN x	13,0	CU	publishtoGLNx			blank			Fixed value
			13,0	TU	publishtoGLNx				enter GLN code from Trading partner(s)		Manual, no interface
local	Publish to	Publish Date	D	CU	publishdate1			blank			Fixed value
			D	TU	publishdate1				today		system date
local	Publish to	Publish to GLN: GLN 8712345012021 GS1 Datakwaliteit Checker	E	CU	publishtoGLN1			blank			Fixed value
			E	TU	publishtoGLN1				if local=NL then publish to data checker GLN 8712345012021		Bus rules
local	Publish to	Publish to GLN: GLN 4051633000000 GLN Coupon	E	CU	publishtoGLN1			blank			Fixed value
			E	TU	publishtoGLN1				if local=BE then publish to GLN coupon 04051633000000		Bus rules
local-Extensions	Contact GLN	Contact GLN	13,0		additionalU420001			BE + NL -> 5410188000016	BE + NL -> 5410188000016		Fixed value
local-Extensions	GS1DAS	Check GS1DAS rules	B		checkGS1DASrules			if local=NL then yes	if local=NL then yes		Bus rules
local-Extensions	Agentrics	Check Agentrics rules	B		checkWWRErules			yes	yes		Fixed value

Afkortingen:

- A Alpha numeric
- B Boolean (Yes or No)
- D Date
- E GLN/GTIN code
- VS Codelijst, doorzoekbaar

Bijlage 8 Package type and material code

[illegible]

Bijlage 9 Bepaling PDM veldwaarden voor “Brand, Subbrand, Functional name, Product group and GPC code”

Abbr	Brand	Subbrand	Functional Name	ProductGroup	Udex (PDM)	GPC (PDM)
TR	Tropicana	PURE PREMIUM	Juice - Ready to Drink	1701440484	Juice/nectar drinks - Ready To Drink	10000222 Juice Drinks Ready To Drink (Perishable)
TR	Tropicana	ESSENTIALS	Juice - Ready to Drink	1701440484	Juice/nectar drinks - Ready To Drink	10000222 Juice Drinks Ready To Drink (Perishable)
TR	Tropicana	SMOOTHIE	Juice - Ready to Drink	1701440484	Juice/nectar drinks - Ready To Drink	10000222 Juice Drinks Ready To Drink (Perishable)
AL	Alvalle	GAZPACHO	Soups - Prepared	1701440484	Juice/nectar drinks - Ready To Drink	10000261 Soups - Prepared (Perishable)
AL	Alvalle	APERITIVO DE VEGETALES	Soups - Prepared	1701440484	Juice/nectar drinks - Ready To Drink	10000261 Soups - Prepared (Perishable)
AL	Alvalle	COLD SOUP	Soups - Prepared	1701440484	Juice/nectar drinks - Ready To Drink	10000261 Soups - Prepared (Perishable)
LO	Looza		Juice - Ready to Drink	1701440483	Juice/nectar 100% - Ready to Drink	10000220 Juice Ready To Drink (Shelf Stable)
7U	7-Up		Drinks Flavoured - Ready To Drink	1701440359	Flavored drinks other - Ready to Drink	10000201 Drinks Flavoured - Ready To Drink
PE	Pepsi	MAX	Drinks Flavoured - Ready To Drink	1701440212	Cola - Ready to Drink	10000201 Drinks Flavoured - Ready To Drink
DO	Dole		Juice - Ready to Drink	1701440483	Juice/nectar 100% - Ready to Drink	10000220 Juice Ready To Drink (Shelf Stable)
GA	Gatorade		Sportsdrinks - Ready to Drink	1701440359	Flavored drinks other - Ready to Drink	10000265 Sportsdrinks - Rehydration - Ready to drink
GA	Gatorade	POWDER	Sportsdrinks - Not Ready to Drink	1701430724	Powder concentrates Other - Non Alcoholic Beverages	10000264 Sportsdrinks - Rehydration - Not ready to drink
MD	Mountain Dew		Drinks Flavoured - Ready To Drink	1701440359	Flavored drinks other - Ready to Drink	10000201 Drinks Flavoured - Ready To Drink
MI	Mirinda		Drinks Flavoured - Ready To Drink	1701440359	Flavored drinks other - Ready to Drink	10000201 Drinks Flavoured - Ready To Drink
RS	Rockstar	ENERGY DRINK	Energy Drinks - Ready to Drink	1701440359	Flavored drinks other - Ready to Drink	10000266 Stimulants - Energy Drinks - Ready to drink
RS	Rockstar	JUICED	Energy Drinks - Ready to Drink	1701440359	Flavored drinks other - Ready to Drink	10000266 Stimulants - Energy Drinks - Ready to drink
VA	Val Tea		Drinks Flavoured - Ready To Drink	1701440359	Flavored drinks other - Ready to Drink	10000201 Drinks Flavoured - Ready To Drink

Bijlage 10 Specificaties en eisen van de BODS interface

Hier volgen detailbeschrijvingen met betrekking tot de interface.

Algemene informatie BODS

- BODS is een SAP interface tool
- Deze tool wordt voornamelijk gebruikt binnen BIS Europa
- De gebruikers zijn alleen BODS key users
- Support – WER BIS Solutions team (Tjalling Dijkstra)

Functionele specificaties:

Waar moet de interface aan voldoen?

BODS moet:

- dagelijks de beschikbaar gestelde bestanden van Oracle vanuit de database kunnen inlezen
- de gegevens eenduidig converteren naar een geschikt PDM formaat
- de geconverteerde gegevens beschikbaar stellen voor PDM door het bestand op de server te plaatsen
- een foutmelding genereren en versturen als een load, conversie of dump niet is geslaagd

Eisen:

- Gegevens worden opgehaald met een ODBC
- Update frequentie: dagelijks, elke avond om 20:00 uur
- Gemiddelde load: 1 product per dag
- Er zijn geen eisen voor veiligheid (interne gegevensuitwisseling)
- Bestanden worden in een “ANY” formaat weggeschreven
- Na elke load moet er een melding verzonden worden naar:
DMANAGM@pepsico.com (Pepsico DataManagement, NEBU mailbox)
- Als de interface faalt, moet een foutbericht gestuurd worden naar:
DMANAGM@pepsico.com (Pepsico DataManagement, NEBU mailbox) en naar
sfg.applications@pepsico.com (applications team).

Technische gegevens:

Met welke systemen wordt er gewerkt?

	Versies:
Oracle Business Suite	10.204
SAP BODS	XI 3.2
SRC-PDM	4.4 GDS BMS 2.7

Deze servers zijn momenteel in gebruik:

Productie: PWU0156NL1 & PWU0167NL1
Test: PWU0050BE1

Bijlage 11 Functioneel ontwerp van de BODS interface

In deze bijlage wordt gedetailleerd beschreven waar gegevens in Oracle te vinden zijn, hoe die geconverteerd moeten worden in BODS en tenslotte geladen worden in PDM.

Oracle

De gegevens worden uit de Oracle database geëxtraheerd door middel van SQL*Plus binnen de Oracle client. Voor een ODBC koppeling is de Oracle client software nodig, inclusief een ODBC driver.

Een Open DataBase Connectivity (afgekort ODBC) is een standaard database toegankelijkheidsmethode, om elk programma met een database te kunnen laten communiceren en dit onafhankelijk van het type database.

Daarna moet alleen nog de verbinding met de database geconfigureerd worden met “tnsnames.ora”

BODS

Het hart van het functionele ontwerp is de mapping van de gegevensvelden uit Oracle met de doelvelden in PDM. In bijlage 7 staat de volledige mapping.

In de tabel is voor beide systemen per veld gekeken naar:

- Technische veldnaam
- Veldlengte of vorm, aantal karakters, datum, waarde uit een keuzelijst of ja/nee, 0/1.
- Voorwaarde waaraan een veld moet voldoen (business rules)
- Voorbeeld van een veldwaarde, voor beeldvorming
- Betekenis van een veldwaarde, voor interpretatie

Naast de mapping zijn er ook velden in het doelsysteem, die niet direct uit de bron zijn te halen. Hiervoor zijn conversietabellen nodig. In bijlage 8 staat de conversietabel van “package type” en “material code”.

In bijlage 9 staat de bepaling PDM veldwaarden voor “Brand, Subbrand, Functional name, Product group and GPC code”. Deze waarden worden afgeleid uit de merk afkorting (Abr) en subbrand uit Oracle.

Niet alle velden van PDM zijn uit Oracle op te halen. Er is besloten, op basis van scope en tijd, om die velden handmatig in PDM te gaan vullen.

SRC-PDM

Gegevens formaat

De geconverteerde data wordt opgeslagen in een bestand, welke vervolgens wordt geladen naar PDM door het gebruik van een “.ini” bestand.

SRC-PDM accepteert verschillende soorten gegevens formaten. Bij gebruik van de meest gangbare formaten, ANY of XML, verloopt het uploaden automatisch.

Bij het gebruik van andere formaten, zoals Excel of CSV, zal er nog een extra conversie moeten plaatsvinden en gebeurt het laden handmatig.

CU en TU

SRC-PDM gebruikt voor de Consumer Unit (CU, fles) en voor de Trade Unit (TU, krat of doos) twee verschillende records. Oracle heeft beide gegevens in één record staan.

De sleutel voor deze records is de GTIN (EAN code). In het ANY bestand zijn dit twee “article” records met eventueel hun “property” records. De TU moet een verwijzing hebben naar de CU om deze aan elkaar te koppelen. Dat gebeurt op de TU met behulp van het attribuut met code “subarticle”. Deze krijgt als waarde de GTIN van de CU. Met het attribuut met code “subaantal” wordt het aantal CU’s in de TU beschreven.

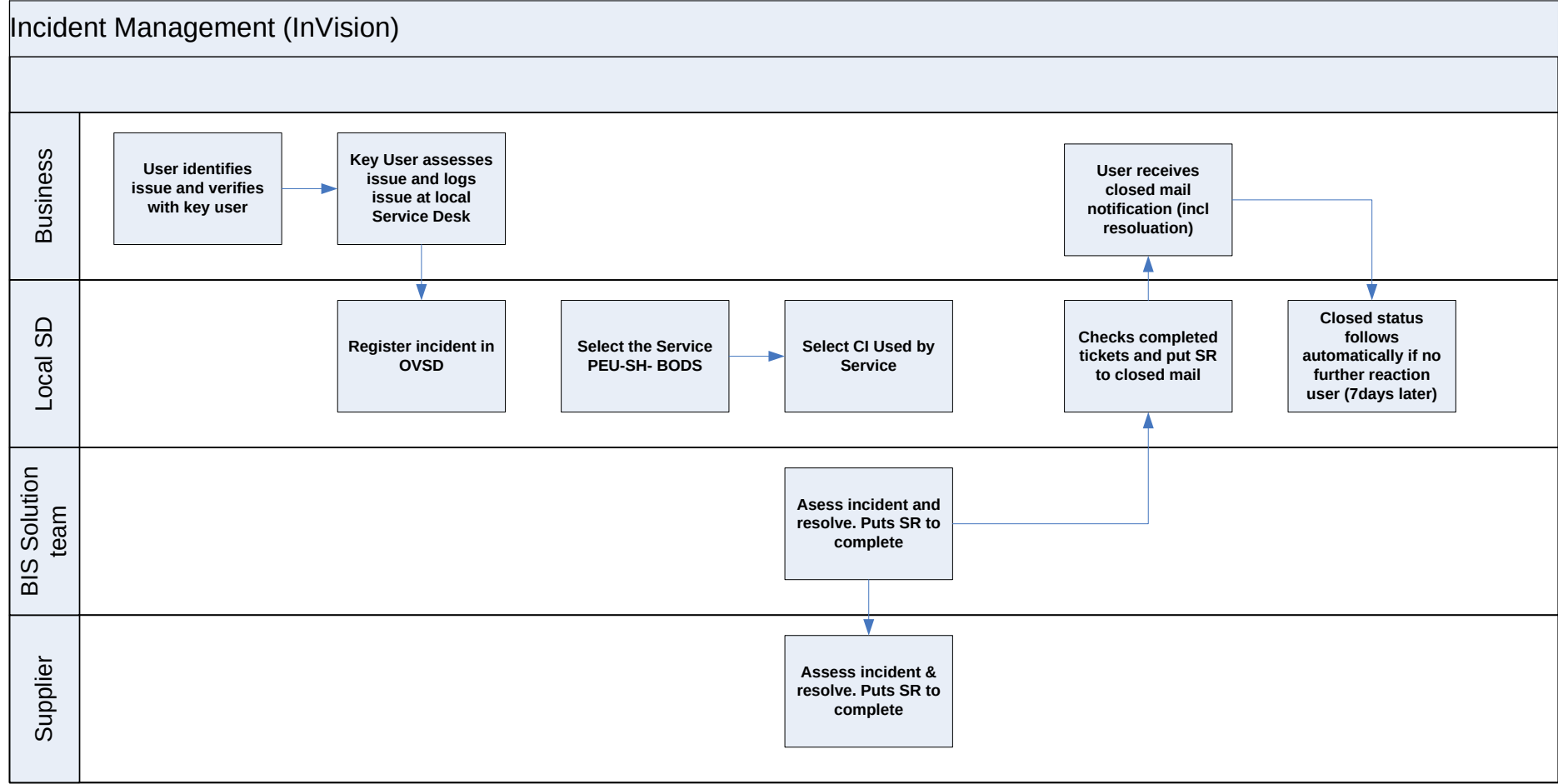
Voor een display artikel dat meerdere verschillende soorten CU’s bevat worden meerdere “subarticle” en “subaantal” attributen gebruikt. Bijvoorbeeld “subarticle1” met “subaantal1” en “subarticle2” met “subaantal2”. Dit kan oplopen tot nr 100.

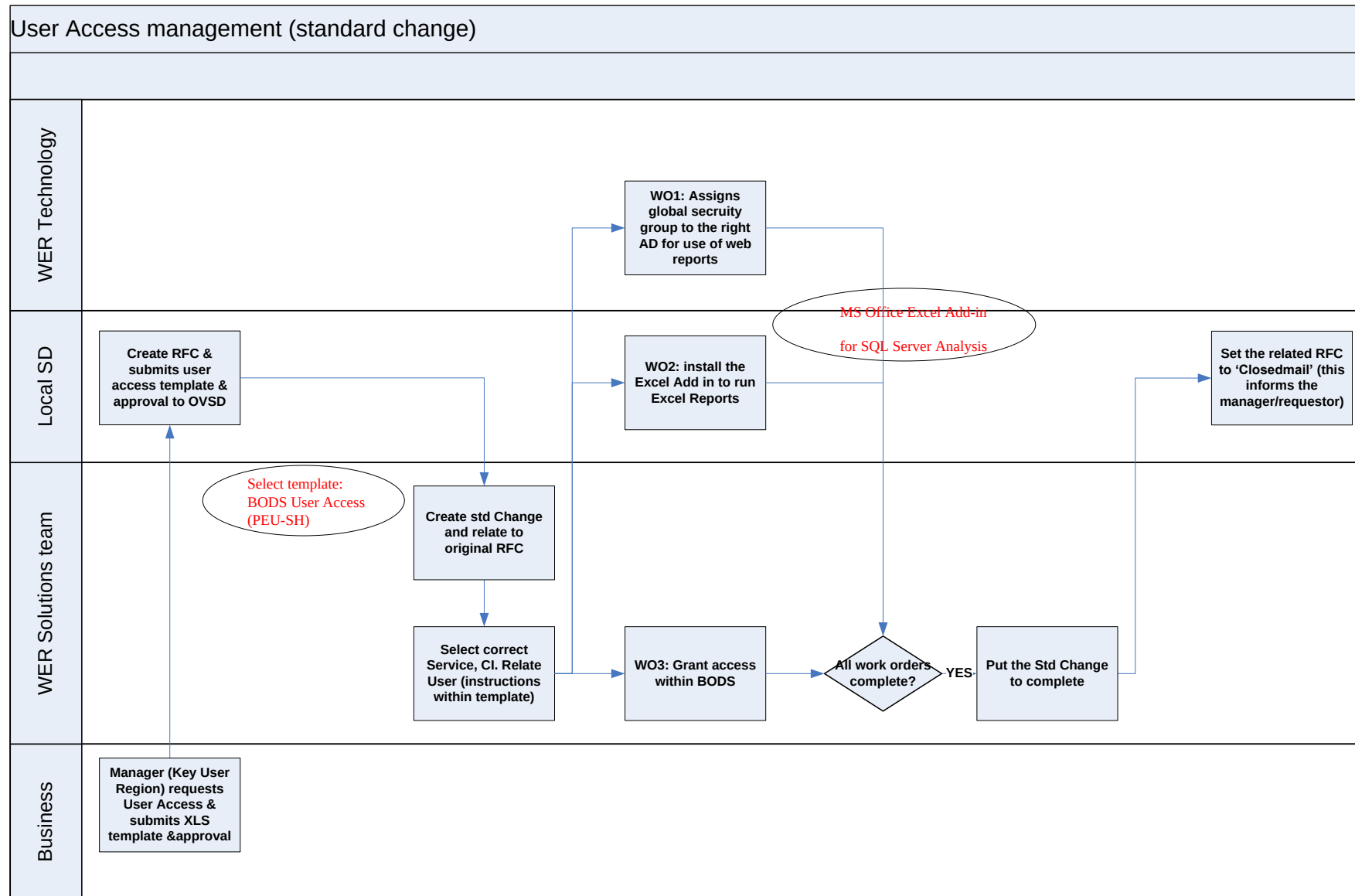
Het ANY bestand kan een onbeperkt aantal artikelen bevatten. Tevens kan een artikel meerdere keren voorkomen in het bestand. De informatie wordt bij elkaar gevoegd door SRC-PDM. Indien ook dezelfde attributen worden aangeboden dan wordt de informatie telkens in SRC-PDM overschreven en daarbij blijft de informatie die het laatst in het bestand staat aanwezig in SRC-PDM.

De bestanden mogen op een willekeurige plaats op de server gezet worden. Echter beter is in de SRC-PDM folders, zodat alles bij elkaar staat. De SRC-PDM bestanden staan in een folder “\SRC-PDM\” op de server.

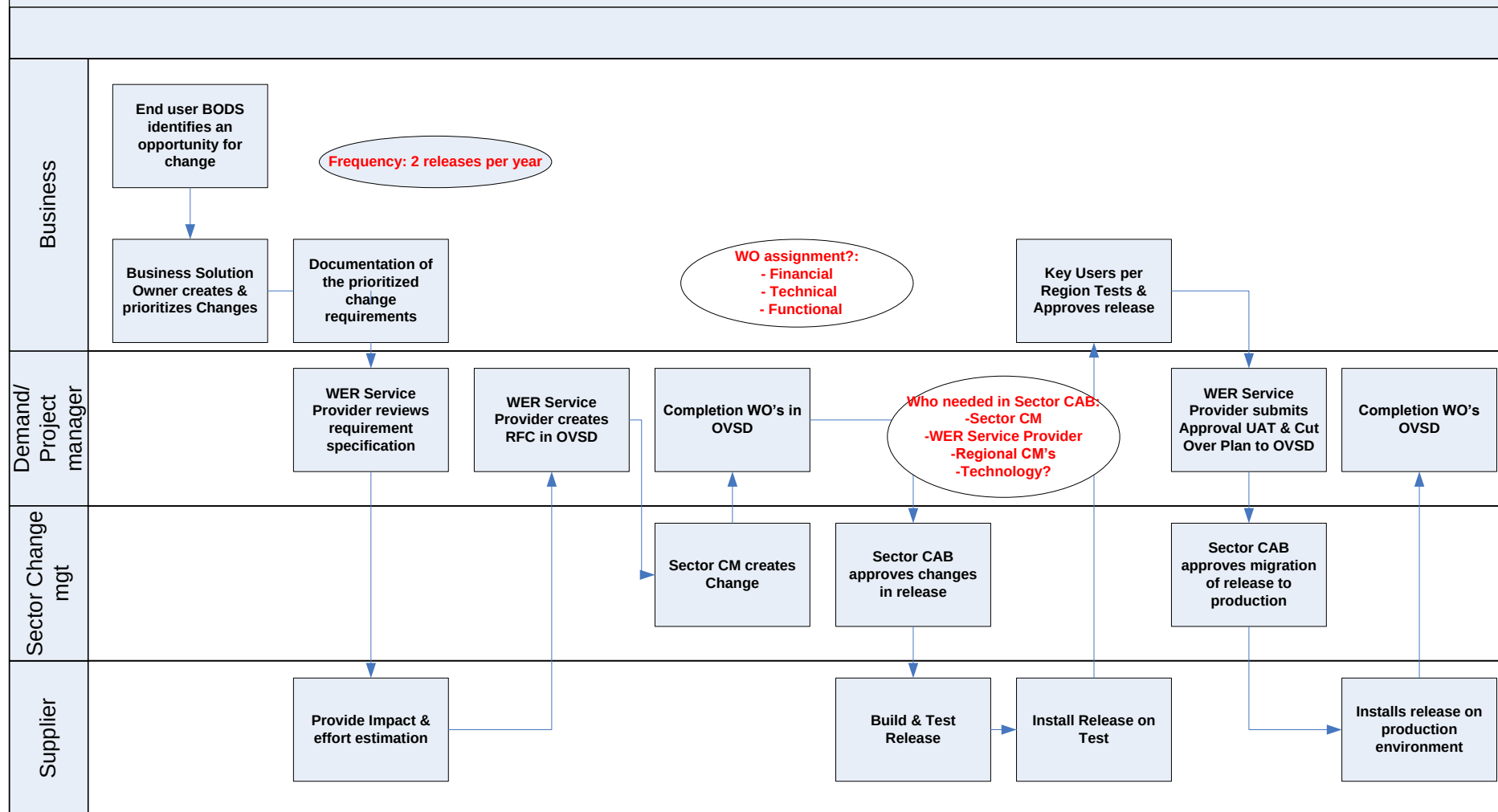
De version van het ANY bestand moet op major=2 geplaatst worden. Een encrypted file wordt niet meer gebruikt/ondersteund.

Bijlage 12 Incident and change management





Release management – Non Standard Changes



Incident Management in OVSD

- 1 Gebruiker meldt issue bij de lokale Service Desk (SD).
- 2 Lokale SD registreert een Service Request (SR) in OVSD (applicatie).
- 3 Lokale SD kent de juiste Service toe aan het Service Request: **PEU-SH-BODS**.
- 4 Lokale SD kent de juiste CI toe (configurable item,)
- 5 Het BIS team lost het probleem op, eventueel ondersteund door een externe partij
- 6 Lokale SD sluit de ticket en stuurt een mail naar de gebruiker

Change management

Hierin wordt onderscheid gemaakt in:

- Standard change, voornamelijk voor het verlenen van gebruikersrechten
Hierbij gaat het om één RFC en een pre-approved change door het CAB.
- Non standard change, voor releases (een verzameling van changes)
Hierbij gaat het om meerdere RFC's en twee goedgekeurde changes door de CAB. Eén om een release te starten (op basis van technische, functionele en financiële specificaties) en één om de test door te zetten naar productie (stage/gate).

Standard change (User Access management)

- 1 Gebruiker vult voor toegang tot de applicatie de Excel template in en laat deze tekenen door manager voor goedkeuring.
- 2 Gebruiker stuurt het verzoek en de Excel sheet met goedkeuring naar de lokale SD
- 3 Lokale SD maakt een RFC, met de juiste service (PEU-SH-BODS) en toegewezen CI
- 4 Het WER Solutions team selecteert de juiste template: BODS User Access (PEU-SH)
- 5 Zij relateert de Licentie CI aan de gebruiker
- 6 Na het opslaan van de change worden er automatisch 3 WO's aangemaakt.
- 7 Als alle WO's klaar zijn, wordt de Change op gereed gezet, waarna degene die het RFC gemaakt heeft wordt geïnformeerd (de lokale SD) en deze vervolgens de gebruiker informeert.

Non standard change (releases)

- 1 Gebruiker identificeert een mogelijkheid voor verandering
- 2 De applicatie eigenaar maakt een change en kent prioriteit toe
- 3 IT demand/project manager beoordeelt de specificaties en stuurt een verzoek naar de software leverancier
- 4 Leverancier komt met een voorstel, inclusief benodigde tijd en kosten
- 5 IT demand/project manager maakt een RFC
- 6 Sector change manager creëert een change
- 7 Als de WO's gereed zijn moet de release goedgekeurd worden door de CAB
- 8 Software leverancier bouwt en test de release en installeert deze
- 9 Gebruiker test de release en geeft goedkeuring
- 10 IT demand/project manager geeft goedkeuring
- 11 CAB geeft goedkeuring om de nieuwe release naar productie te zetten.
- 12 Leverancier installeert de release in productie

AD	Active Directory
CAB	Change Advisory Board
CI	Configurable Item, zoals hardware, software, server
RFC	Request For Change (Verzoek)
OVSD	Open View Service Desk (HP applicatie)
SD	Service Desk
SR	Service Request