

Datakwaliteit belangrijkste voorwaarde voor deduplicatie

Consumenten Data Integratie

Sjoerd Hobo en Rob Peters

Hoe leren we de klant kennen? Hoe komen we tot een eenduidig profiel van onze klant? Het zijn slechts twee voorbeeldvragen die consumentenorganisaties zich stellen in het kader van hun CRM-strategie. Alleen als een organisatie haar consument kent, zal ze beter in staat zijn deze consument voor de organisatie te behouden of zelfs de relatie verder uit te breiden.

In dit artikel staan we stil bij hoe een consumentenorganisatie al haar beschikbare consumenteninformatie kan integreren en alle uitdagingen die hierbij komen kijken. Het gaat om integraties die leiden tot een eenduidig consumentenbeeld met als doel het mogelijk maken van consumentencontact.

Consumentengegevensbronnen

We beginnen bij de te onderscheiden bronsystemen die consumenten- en productenregistraties vastleggen en wat hierbij de essentiële gegevenselementen zijn. Vervolgens wordt het deduplicatieproces onder de loep genomen, het proces dat dubbele consumentenregistraties weer tot een eenduidige registratie terugbrengt. Belangrijke randvoorwaarde voor het kunnen dedupliceren, maar ook voor het juist kunnen gebruiken van de gegevens zoals deze door de consument zijn achtergelaten, is datakwaliteit. Daarom worden de gegevens geschoond en verrijkt daar waar nodig en mogelijk. In dit artikel zal duidelijk worden dat een consumenten database (CDB) vraagt om bidirectionele interfaces met de aanleverende gegevensbronnen. Tot slot wordt stilgestaan bij hoe de door de consument al dan niet gegeven permissies kunnen worden geïmplementeerd binnen het CDB. Consumenten leren kennen, begint bij het verzamelen van consumentengegevens. Een organisatie kan op verschillende manieren, via diverse kanalen, deze informatie verkrijgen. In dit artikel beperken we ons tot vier manieren.

Registratie websites. Meer en meer wordt aan consumenten gevraagd zich via internet te registreren (consumentenregistratie). Veelal is dit een vrijblijvend verzoek, al dan niet aangemoedigd door het aangeven van de toegevoegde waarde van de registratie. Iets minder vrijblijvend wordt het indien wordt gevraagd een product te registreren om de garantieperiode kenbaar te maken (productregistratie). In de consumentenelektronica markt (bijvoorbeeld streamium-producten) werken steeds meer producten in

combinatie met een on-line internetverbinding, waarvoor registratie vereist is.

Call Centers. Een andere manier om een consument te leren kennen is via het vastleggen van een vraag, probleem of klacht over het product dat is aangeschaft. Veelal bieden organisaties twee opties. Een vraag, probleem of klacht kan worden geregistreerd via het internet (zie punt 1, wederom een registratie-site) of een call center kan worden gebeld. In dit laatste geval zal de call operator de consument te woord staan en de consument een aantal vragen stellen. Deze gegevens worden vastgelegd in de call center-applicatie.

eShops. De verkoop van producten via internet is een volgende mogelijkheid voor het vastleggen van klantgegevens. Om juiste betaling en aflevering te kunnen garanderen, dient de consument zijn basisgegevens te registreren.

Overige manieren. Uiteraard zijn er nog veel meer manieren om consumenten te leren kennen of om basisgegevens te verkrijgen. Denk aan het aankopen van externe bestanden, aankopen van aangeboden CRM services, gebruiken van interne versplinterde bedrijfsinitiatieven (zogenoemde *shoebboxes*) of het laten uitvoeren van survey's op consumenten die daarvoor in een eerder stadium hun toestemming hebben gegeven.

Belangrijke gegevenselementen

Hoe meer consumentendetails, hoe beter? Misschien is dit toch niet helemaal waar. Immers, als een consument vaker en om meer informatie wordt gevraagd, dan wordt de kans groter dat de consument zich niet kenbaar maakt aan de organisatie of, verveeld door de vele vragen, onjuiste (misleidende) informatie opgeeft.

Een organisatie doet er verstandig aan te bepalen wat de basisset aan consumentengegevens moet zijn teneinde de consument te leren kennen. Deze basisset wordt afgeleid uit de CRM-strategie,

eventueel ondersteund met een campagnekalender. Waarom een consument vragen naar bepaalde interesses als er verder niets mee wordt gedaan? Waarom zou men niet een consument vragen naar het geslacht, als er wel plannen zijn producten onder de aandacht te brengen in campagnes die specifiek zijn voor mannen of vrouwen?

Het is de uitdaging voor iedere organisatie om te bepalen wat de basisset aan belangrijke gegevenselementen is en hoe de consument ertoe aan te sporen deze kenbaar te maken. Uiteraard zijn NAW-gegevens, geslacht, leeftijd, e-mailadres belangrijk binnen een klantenprofiel. Met geregistreerde producten (van het type eigenaar of geïnteresseerde) kan een veel specifiekere klantenprofiel worden bepaald en zelfs up-selling en cross-selling worden overwogen. Overigens, het vragen naar AW-gegevens lijkt in veel gevallen een automatisme binnen registratie-sites. Echter, consumentencontacten vinden steeds vaker elektronisch plaats. Hoe belangrijk is het dan nog om de AW te vragen en op te slaan? In ieder geval zal altijd de consument expliciet gevraagd moeten worden om aan te geven of hij of zij 14 jaar of ouder is en dat ze al dan niet toestemming geeft voor toekomstige vormen van klantencontact.

Deduplicatie

Het is gunstig dat een consumentenorganisatie in staat is via meerdere kanalen (en bijbehorende systemen) consumentenregistraties te verzamelen. Meer mogelijkheden, betekent meer consumenten- en productregistraties. Echter, meer dan één registratiesysteem brengt onmiddellijk een probleem met zich mee. Namelijk, een consument kan zich via meer dan één kanaal kenbaar maken. Als gevolg is het dus mogelijk dat voor een en dezelfde consument, meerdere profielen beschikbaar zijn. Dit is uiteraard niet wenselijk, iedere organisatie zal ten alle tijden willen streven naar een globaal en eenduidig consumentenprofiel. Om dit probleem op te lossen, is het in eerste instantie noodzakelijk alle consumentenregistraties te verzamelen in één omgeving, de consumenten-database (CDB). Vervolgens zal binnen de CDB een proces moeten plaatsvinden die verschillende registraties voor één en dezelfde consument verenigd tot slechts één registratie. Dit proces wordt ook wel aangeduid met het deduplicatieproces. Deduplicatie is functioneel gezien een lastig proces. Hoe bepaal je nu of je te maken hebt met dezelfde consument? Diverse algoritmen zijn denkbaar. Deduplicatie begint met te trachten te identificeren welke consumentenregistraties nu tot dezelfde consument toe behoren. Dit kan door in eerste instantie de meest



DataMirror[®]
SUCCESS COMES IN REAL-TIME

When information is accurate, available and in sync across the enterprise, organizations profit.

DataMirror Transformation Server
Catch the real-time wave.

For more information, please call us at **+32 (0)3 206 1882** or e-mail us at **beneluxoffice@datamirror.com**

www.datamirror.com/bi



Who knew data could be so accessible?

unieke elementen uit een profiel te proberen te matchen. Een e-mailadres en een telefoonnummer zijn doorgaans uniek (behoort toe aan slechts een consument). Het is verstandig met deze twee elementen te starten in het deduplicatieproces. Mocht dit niet mogelijk zijn, kunnen meer elementen worden toegevoegd aan de zoektocht (NAW, geboortedatum, etcetera).

Nadat is bepaald welke profielen behoren tot eenzelfde consument, staat een volgende uitdaging te wachten. Welk profiel, of deel van een profiel, is nu het beste en correct? Wederom zijn diverse opties mogelijk. Een voor de hand liggende manier is te stellen: des te nieuwer het profiel, des te beter, dus dit profiel gebruiken! Helaas gaat dit niet altijd op en moeten er meer elementen betrokken worden in dit proces. Bijvoorbeeld, afhankelijk van het kanaal kan een profiel meer of minder foutgevoelig zijn.

De verschillende soorten bronsystemen zullen doorgaans ook een verschillende mate van datakwaliteit laten zien. Een eShop en Call Center website leveren doorgaans de betere gegevens op dan een registratie-site. Bij het aanschaffen van een product of het melden van een probleem, is een consument gebaat bij het geven van juiste informatie. Een registratie-website is veel vrijblijvender. Het is een optie om een eShop-registratie boven een registratie website te plaatsen. Maar, om het toch weer ingewikkeld te maken, een actueel registratie-website profiel kan weer beter zijn dan een drie jaar oude Call Center-registratie. Kortom, het bepalen van het deduplicatieproces is een ingewikkeld proces. Alle relevante elementen moeten worden meegenomen in de beslisboom.

Bij vergelijking van twee profielen, P1 en P2, zal het nooit zo zijn dat één profiel volledig beter is dan het andere profiel. Het is verstandig deelprofielen te definiëren: bijvoorbeeld deel a = consumentnaam en eigenschappen; deel b = consumentadres; deel c = consument e-mail. Het uiteindelijke profiel wordt dan opgebouwd uit het beste van de twee profielen P1 en P2. Dus een eindprofiel

P zou er uiteindelijk als volgt kunnen uitzien: $P = P1.deel a + P1.deel b + P2.deel c$.

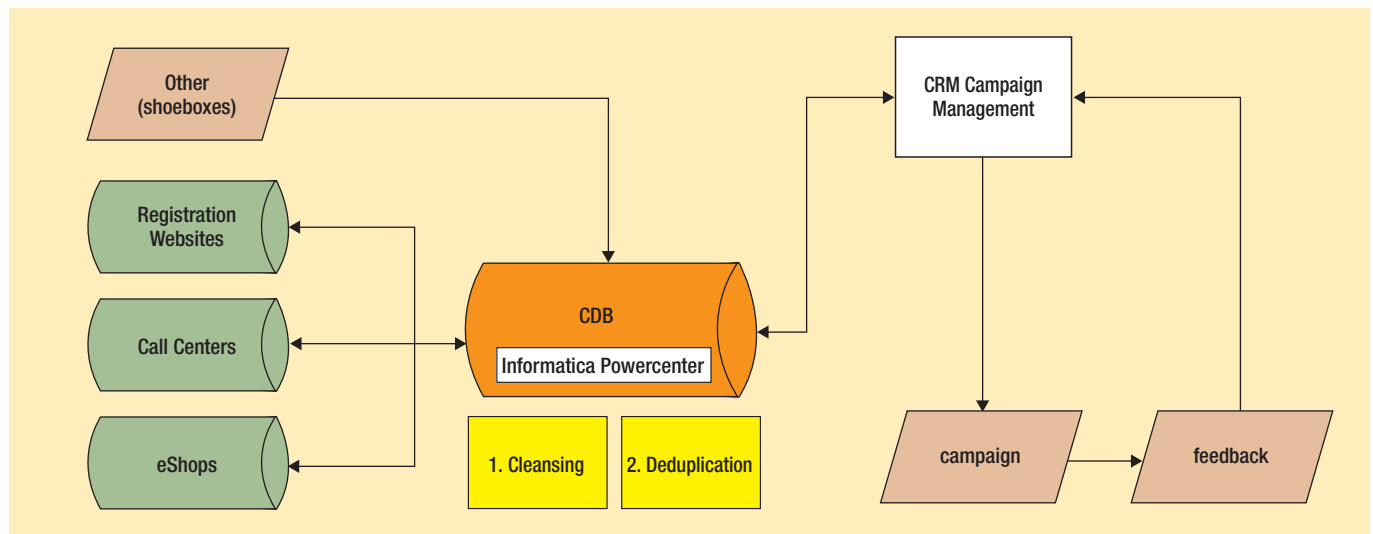
Data schonen en verrijken

Consumentengegevens verzamelen en integreren is geen doel op zich voor een consumentenorganisatie. Het is de consumentenorganisaties enerzijds te doen om inzicht (het leren kennen van de consument en zijn gedrag), anderzijds om (indien er expliciet toestemming voor is gegeven) de consument op de hoogte te houden van nieuws, promoties, etcetera met behulp van 'white mail' of e-mail campagnes.

Deduplicatie is functioneel gezien een lastig proces

Essentieel voor consumentendata-analyse en campagnes is dat de consumentengegevens zo accuraat mogelijk zijn. Daar waar mogelijk en noodzakelijk, worden consumenten gegevens geschoond en verbeterd met behulp van data quality-producten (zoals bijvoorbeeld Trillium Software). Het schonen van data richt zich bijvoorbeeld op het corrigeren van NAW-gegevens, het doorvoeren van verhuizingen, aangeven dat een consument is overleden. Daarnaast is het binnen de CDB omgeving ook belangrijk profielen te verrijken, bijvoorbeeld door toevoeging van geslacht en geboortedatum. Voor het gericht kunnen sturen van campagnes naar de consumenten in het CDB, kan het belangrijk zijn te weten wie er man of vrouw is (indien een campagne omtrent een nieuwe Ladyshave wordt overwogen) of een indicatie te hebben van de leeftijd van een consument (de groep van 14 tot 18 jaar is wellicht niet de groep om te benaderen over de nieuwste high-end flatscreens).

Data schonen en verrijken staat overigens niet los van deduplica-



Afbeelding 1: Bidirectionele architectuur.

tie. Teneinde betrouwbare deduplicatie mogelijk te maken, dienen de data eerst geschoond en verrijkt te worden, het vergroot de kans van succesvolle deduplicatie en dus het bepalen van het eenduidige eindprofiel uit de eventueel diverse aangeleverde consumentenprofielen.

Bidirectionele bron-interfaces

Uit de besproken aspecten van consumentendata-integratie (deduplicatie en data schonen/verrijken), volgt dat in het CDB een eenduidig en zo accuraat mogelijk profiel van de consument wordt vastgelegd. Hierdoor is het dus mogelijk dat het consumentenprofiel in het CDB niet meer in lijn is met de aangeleverde gegevens uit de diverse bronsystemen. Daarnaast is het heel gebruikelijk dat een consument in een ontvangen e-mail van de campagne, in de gelegenheid wordt gesteld zijn gegevens aan te passen. Deze zullen moeten worden verwerkt in het CDB. Ook de feedback op eventueel gehouden survey's kan belangrijke nieuwe consumentengegevens opleveren.

Echter, als de bronsystemen die het CDB voeden met consumentenprofielen, unidirectioneel zijn, zal de consument zijn doorgegeven wijziging niet terugzien in het user interface van deze bronnen (bijvoorbeeld in de registratie-website) of de call operator niet in het bezit zijn van bijvoorbeeld de nieuwe NAW-gegevens van een verhuisde consument. Kortom, dataschoning en -verrij-

king alsmede ontvangen feedback zullen teruggestuurd moeten worden naar die bronsystemen die ooit het gewijzigde consumentenprofiel hebben aangeleverd.

De vraag is dan welke elementen van een profiel moeten worden teruggestuurd naar de bronsystemen? Basiselementen als: een consument wil niets meer ontvangen of juist wel (opt-out of opt-in) of juist meer of misschien minder (interest area's), zijn voor de hand liggend. Maar wat nu als een consument aangeeft man te zijn, maar in het CDB is vastgesteld dat het een vrouw betreft. Zou de consument ervan gediend zijn dat in het geval moedwillig verkeerde informatie is verstrekt, dat dit vanuit het CDB gecorrigeerd wordt in het bronsysteem en dus daarmee zichtbaar voor de consument veranderd is? Dit gaat niet over goed of fout maar over hoe de consumentenorganisatie met haar klanten omgaat. In ieder geval dient binnen de CDB architectuur rekening te worden gehouden met bidirectionele interfaces, zie afbeelding 1.

Consumentpermissies

Consumenten- en productregistraties mogen niet zonder meer door organisaties worden opgeslagen en gebruikt binnen de CDB omgeving. Een eerste belangrijke randvoorwaarde waaraan voldaan moet zijn voor het mogen opslaan van een registratie in het CDB, is de garantie dat het een consument betreft die 14 jaar of ouder is. Het aanleverende bronsysteem zal dit gegeven voor

Jong ICT bedrijf zoekt BI & CMS Young Professionals met ambitie



Jong

De gemiddelde leeftijd ligt bij ons rond de 30 jaar. Hierdoor ontstaat een jonge honden cultuur: veel enthousiasme en een grote betrokkenheid. We werken hard, maar vergeten niet onze successen te vieren. We leren veel, maar zorgen ook voor ontspanning. Dit alles resulteert in een erg plezierige werksfeer.

Young Professionals met ambitie

Op 1 januari 2007 gaat VLC's Young Professional Academy van start. In zes weken tijd word je opgeleid tot junior Business Intelligence & Content Management consultant. Vervolgens ga je onder begeleiding van een senior consultant aan de slag bij een van onze klanten. Een ambitieuze start van een ambitieuze carrière.

VLC zoekt

Starters met een afgeronde HBO/academische IT opleiding, goede communicatieve vaardigheden, enthousiasme, ambitie en leergierigheid.

VLC biedt

Een ambitieuze start van je carrière, jonge collega's, een goed salaris, een leaseauto, een vaste maandelijkse onkostenvergoeding en een laptop van de zaak. En dit alles vanaf de eerste dag dat je bij ons in dienst bent, dus ook tijdens je opleiding.

Interesse?

Stuur dan een reactie naar mathijs.kreugel@vlc.nl of kijk op www.vlc.nl.



Age >= 14	No	Yes		
Permission asked	Not Processed	No		Yes
		Existing Consumer	New Consumer	
		Permissions Unchanged	Set Opt-Out = Yes	
Opted-Out	Not Processed	Not Processed		No
				Process Individual Interest Areas **
Interest Area 1	Not Processed	Not Processed		Set to value of Interest Area 1
Interest Area 2	Not Processed	Not Processed		Set to value of Interest Area 2
Interest Area N	Not Processed	Not Processed		Set to value of Interest Area N

Afbeelding 2: Consumentenpermissies.

iedere consument moeten aangeven. Indien dit niet gebeurt of de consument niet 14 jaar of ouder is, zal de consument niet opgeslagen mogen worden in het CDB, zie afbeelding 2. Voor consumenten van 14 jaar of ouder is de volgende controle na te gaan of de consument gevraagd is toestemming te geven voor iedere vorm van communicatie tussen de consument en de organisatie. Indien dit niet het geval is, is het van belang na te gaan of de consument al reeds bekend is in het CDB. Mocht dit het geval zijn, dan wordt de consument geladen in het CDB, echter met de aantekening dat deze consument niet benaderd mag worden in welke communicatievorm dan ook. Deze registratie zal dus enkel en alleen nuttig zijn voor consumentanalyses. Indien de consument al wel bekend is in het CDB, dan wordt zijn profiel wel aangeboden aan het CDB (en dus gededuplicateerd) echter zijn reeds aanwezige instelling met betrekking tot toestemming voor communicatie blijft ongewijzigd.

Ook feedback op survey's kan belangrijke nieuwe consumentengegevens opleveren

Als tijdens het registratieproces wel toestemming is gevraagd aan de consument voor toekomstige communicatie tussen de consument en de organisatie, dan wordt bij geen toestemming de consument geladen in het CDB met de aantekening dat de consument niet benaderd mag worden. Indien het antwoord positief is, dan kan het voor de organisatie meest waardevolle

consumentprofiel worden geladen en in de toekomst worden benaderd voor promoties, nieuwsbrieven etcetera (afhankelijk van de opgegeven interest area's).

Conclusie

Consumenten Data Integratie (CDB) kent een aantal specifieke uitdagingen. Het CDB zal in de praktijk diverse bronsystemen kennen die consumenten- en productenregistraties opleveren. Het integreren van deze profielen in het CDB kan dubbele registraties opleveren. Om deze dubbele registraties terug te brengen naar een eenduidig consumentenbeeld zal er een deduplicatieproces geïmplementeerd moeten worden. De kwaliteit van de profielen zal niet altijd even goed zijn, het hangt af van het aanleverende kanaal. Als een organisatie voordeel wilt halen uit de registraties, zal ze de profielen moeten schonen en verrijken. Dit is alleen al noodzakelijk om het deduplicatieproces een kans van slagen te geven.

Het schonen en verrijken van data alsmede het deduplicatieproces zullen ertoe leiden dat het uiteindelijke consumentenprofiel in het CDB niet meer in lijn is met de bronsystemen. Binnen de CDB omgeving zijn bidirectionele interfaces noodzakelijk om het aangepaste profiel terug te koppelen aan de aanleverende bronsystemen. Overigens is het een organisatie niet zonder meer toegestaan consumentenprofielen op te slaan en te gebruiken. Binnen het CDB moet een duidelijke permissiestructuur worden geïmplementeerd.

Sjoerd Hobo en Rob Peters

zijn senior consultant bij QNH Enterprise Intelligence b.v.