

Register- en sleutelbeleid

Register- en sleutelbeleid

Bert Dingemans

Abstract

Bij een toenemende volwassenheid van data-architectuur neemt het gebruik van generieke gegevensverzamelingen toe. Deze gegevens worden veelal beheerd in registers waarbij men werkt volgens het principe eenmalig beheer en meervoudig gebruik.

Om generieke gegevens te kunnen onderscheiden in registers dient beleid rond sleutels opgesteld te worden. Dit document beschrijft een aantal aspecten van dit register- en sleutelbeleid.

Register- en sleutelbeleid

Inleiding

Binnen vrijwel alle (informatieverwerkende) organisaties is het gebruik van gegevens binnen de werkprocessen een alledaags gegeven. Daarbij zie je dat bepaalde gegevens in meerdere processen ingezet worden. Dit hergebruik kan zelfs over de grenzen van de organisatie heen gaan.

Deze generieke gegevensentiteiten zijn in te delen naar niveaus van genericiteit en herbruikbaarheid. De informatiesystemen waarin deze entiteiten beheerd worden zijn in te delen naar deze genericiteit in de vorm van registers. Hoe hoger in de indeling hoe meer eisen er gesteld worden aan de inrichting- en beheerprocessen .

Door deze registerindeling ontstaan er eisen aan de herbruikbaarheid en de identificeerbaarheid van de afzonderlijke entiteiten binnen de gegevensverzamelingen. Dit wordt beschreven als sleutelbeleid. Deze sleutels identificeren de entiteiten en daarmee stellen deze sleutels eisen aan de beschrijving, metadata en inrichting.

Voor deze registers en sleutels zijn een aantal ondersteunende applicatie functies zoals componenten en services te onderkennen. Deze worden in het hoofdstuk voorzieningen beschreven.

Registers

Binnen alle bedrijfsprocessen is er behoefte aan gegevens, deze gegevens zijn de basis voor informatie en op basis van informatie worden beslissingen genomen. Bijvoorbeeld binnen een subsidieproces van de overheid zullen aanvraaggegevens van de aanvrager geanalyseerd worden in combinatie met de criteria op basis waarvan een subsidie wordt toegekend.

Naast de gegevens die de bedrijfsprocessen ondersteunen zijn er ook gegevens te onderkennen die door de organisatie beheerd worden en voor de relaties (of afnemers) van de organisatie relevant zijn binnen hun bedrijfsprocessen. Denk hierbij bijvoorbeeld aan het beschikbaar stellen van gegegevens door overheden die ingezet kunnen worden binnen de bedrijfsprocessen van natuur- en milieuorganisaties en het bedrijfsleven. De ontwikkelingen op het vlak van open data zijn een voorbeeld hiervan.

Bovenstaande uitwerkingen zijn voorbeelden van de herbruikbaarheid van gegevens. Nu is het ene gegeven meer herbruikbaar dan het andere, er is dus een gradatie in de herbruikbaarheid.

Bijvoorbeeld de gegevens van een bedrijf zijn herbruikbaar voor de KvK, meerdere overheidsinstellingen en het bedrijfsleven. Gegevens omtrent de medewerkers van de eigen organisatie worden hergebruikt binnen meerdere interne bedrijfsfuncties. Gegevens over de bedrijfsprocessen binnen een afdeling zullen in veel gevallen alleen relevant zijn binnen deze afdeling en daardoor een lage herbruikbaarheid kennen.

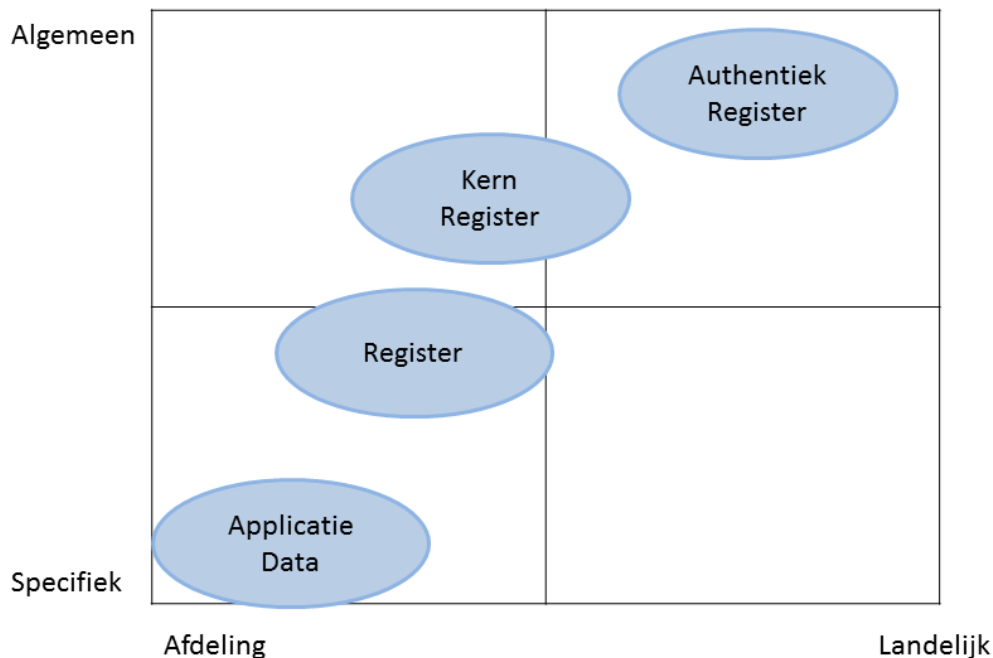
Bij traditionele automatisering werd het onderscheid tussen generieke- en specifieke gegevens nauwelijks gemaakt. Ieder informatiesysteem had bijvoorbeeld een eigen registratie (en opslag) van relatiegegevens. Vanuit perspectief van efficiëntie, actualiteit en correctheid van gegevens is dit ongewenst. Daarom wordt er een niveauindeling gemaakt van gegevensopslag.

Indeling

Onderstaand indeling geeft een indeling van gegevensopslag:

- **Authentieke registers**, opslag van zeer generieke gegevens die veelal organisatieoverstijgend gebruikt worden. Voorbeeld zijn de basisregistraties van de overheid zoals GBA, NHR en BAG. Ook andere gegevens kunnen in een authentiek register worden opgeslagen, bijvoorbeeld gezamenlijke geo databanken die door externe partijen geraadpleegd worden.
- **Kernregisters**, opslag van gegevens die gebruikt worden door meerdere afdelingen in de organisatie. Een voorbeeld zijn de personeelsgegevens die gebruikt worden voor weergave in de wieswie maar ook voor toegang tot de verschillende applicaties in de organisatie.
- **Registers**, opslag van gegevens die hergebruikt worden maar binnen een beperkte scope, bijvoorbeeld binnen één afdeling. Een voorbeeld is de data in de CMDB van Topdesk
- **Domeinspecifieke data**, opslag van gegevens die specifiek zijn voor de ondersteuning van één werkproces of bedrijfsfunctie. Deze gegevens zijn niet relevant voor een andere werksoort m.u.v. bijvoorbeeld rapportages. Denk hierbij aan procesgegevens bijvoorbeeld de opslag van servicedesk gegevens binnen Topdesk.

In onderstaande afbeelding is dit uitgewerkt in een kwadrant



Logischerwijs zijn de gegevens in de authentieke- en kernregisters het meest interessant om onder te brengen in een inrichting van “enkelvoudig beheer en meervoudig gebruik”. In

onderstaande opsomming zijn een aantal voorbeelden van gegevensentiteiten opgenomen die via deze registers beheerd en ontsloten kunnen worden:

Authentieke registers

- Bedrijf of Instelling
- Persoon
- Adres
- Gebouw
- Geo objecten

Kernregisters

- Medewerker
- Functie en rol
- Project
- Relatie
- Zaak

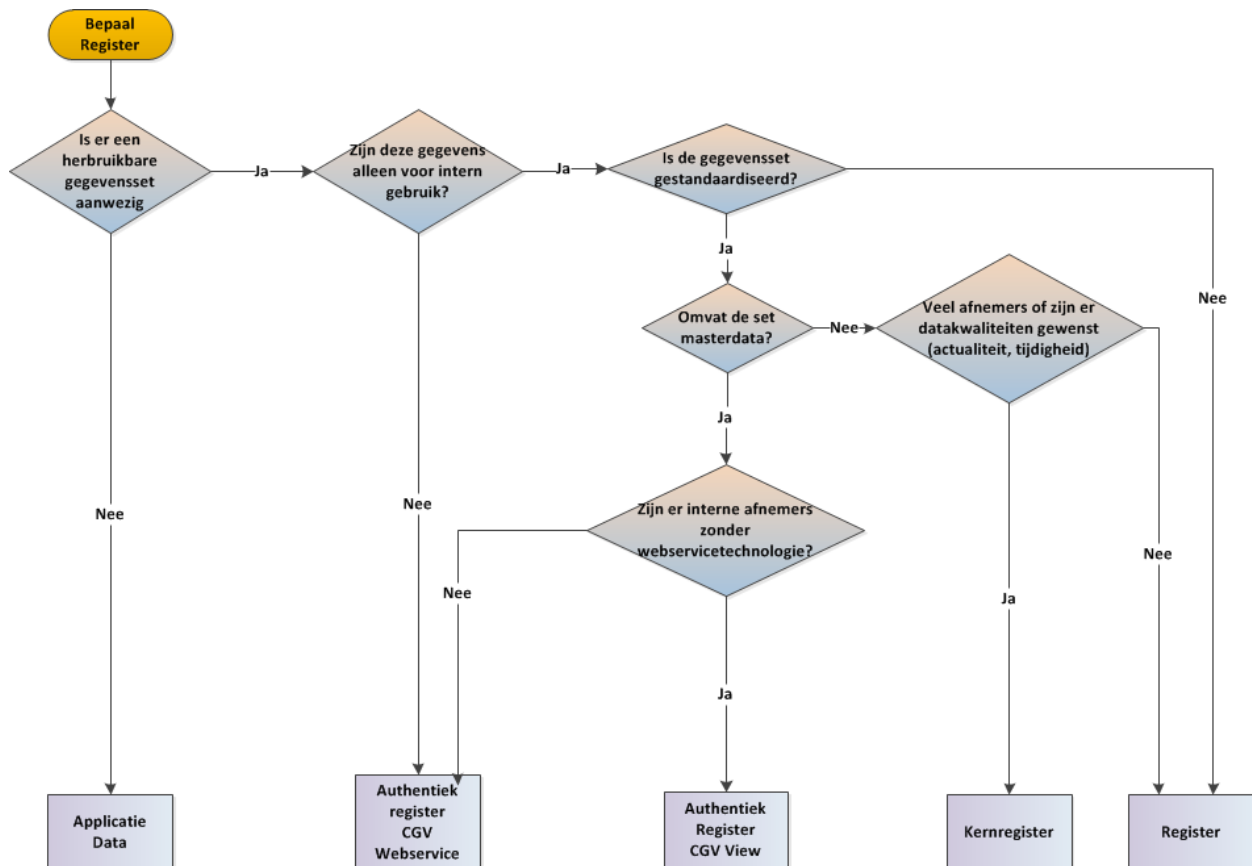
Registers

- Financiële gegevens zoals budgetten e.d.
- Organisatie indeling (afdelingen/diensten)

Registerbepaling

Voor het inrichten en ontsluiten van een register, en zeker van een authentiek register is een grotere inspanning vereist dan voor het inrichten van een applicatie die alleen applicatiedata gebruikt. Denk hierbij aan voorzieningen zoals platform- en implementatie afhankelijkheid of genericiteit en standaardisatie van gegevensverzamelingen en interfaces. In de volgende paragraaf wordt hierop ingegaan.

In deze paragraaf gaan we in op het bepalen van het registertype in combinatie met de ontsluitingswijze. Zie hiervoor onderstaande beslisboom.



Data kwaliteiten

Op basis van de data kwaliteiten zoals gespecificeerd in de architectuurrichtlijn over data kwaliteiten wordt in onderstaande matrix aangegeven welke data kwaliteiten relevant zijn. In deze architectuurrichtlijn wordt vervolgens aangegeven met welke maatregelen (zie de richtlijn) deze kwaliteiten te realiseren zijn.

Omschrijving	Toelichting	Authentiek register	Kern register	Register
Accuraatheid	Accuraatheid heeft betrekking op de mate waarin een data entiteit de werkelijkheid weergeeft. Accuraatheid kan bepaald worden door een data entiteit te vergelijken met de entiteit in de werkelijkheid. Voorbeeld uit het verleden waarbij een baby een stemkaart ontving op basis van een lage accuraatheid binnen de GBA.	X		
Compleetheid	Dit heeft betrekking op de mate waarin bepaalde attributen binnen een data entiteit	X		

	aanwezig zijn. Daarnaast geldt de compleetheid ook voor het altijd voorkomen van een bepaalde set van entiteiten (rijen) in een data set. Bijvoorbeeld een persoon kan enkel de eigenschap naam hebben, maar ook bestaan uit roepnaam, voornamen, tussenvoegsel, achternaam en meisjesnaam. In het laatste geval is er sprake van een hogere compleetheid			
Consistentie	Dit heeft betrekking op het feit dat de ene data set van een bepaalde entiteit gelijk is aan een andere dataset. Met andere woorden is een data entiteit onafhankelijk van de bron altijd dezelfde. Een voorbeeld van een lage consistentie is als er verschillen zitten tussen datasets van dezelfde entiteiten afkomstig uit verschillende bronnen. Replicatie van gegevens is vaak een bron van lage consistentie.	X	X	
Actualiteit	Mate waarin een data entiteit de actuele situatie van de werkelijkheid weergeeft. Veelzeggende voorbeelden zijn overleden mensen die een brief krijgen op basis van een verouderde dataset. Replicatie van gegevens is vaak een bron van lage actualiteit.	X	?	
Precisie	Mate van detail waarin een data entiteit de werkelijkheid weergeeft. Dit heeft bijvoorbeeld betrekking op de precisie van getallen e.d. Opslag van getallen en datums kunnen onvoldoende nauwkeurig zijn omdat afronding bij opslag nodig is. Ook domeinen op eigenschappen kunnen onvoldoende precisie hebben (denk aan een nederlandse postcode in een internationale data-opslag	X	X	
Privacy	Voor sommige data entiteiten is toegangscontrole (autorisatie en authenticatie) of monitoring van gebruik nodig. Denk hierbij aan eisen die gesteld worden aan de ontsluiting van vertrouwelijke gegevens. In het GBA kent men bijvoorbeeld meerdere niveaus van vertrouwelijkheid. Zo worden bevragingen	X		

	van ambtenaren wel gelogd en weergegeven aan de burger maar voor opsporingsambtenaren worden wel gelogd maar niet getoond.			
Redelijkheid	Heeft vooral betrekking op verwachtingen binnen een bepaalde operationele context. Denk bijvoorbeeld aan het accepteren van een lagere performance bij piekbelasting of aan het langer moeten wachten op een resultset van gearchiveerde data-entiteiten	X	?	
Referentiele integriteit	Dit is de situatie waarbij verwijzigingen vanuit de ene data entiteit altijd correct verwijzen naar de gerelateerde data entiteiten. Voorbeelden zijn dubbele sleutels in een dataset waardoor de gekoppelde entiteiten niet kunnen bepalen wat de ouder entiteit is. Ook dangling references waarbij de ouder niet meer bestaat hoort hiertoe	X	X	
Tijdigheid	Is een dataset tijdig beschikbaar binnen de gestelde verwachtingen. Het is het verschil tussen het moment van behoefte en beschikbaarheid. Bijvoorbeeld het opvragen van gegevens in het KCC stelt hoge eisen. Vijf minuten wachten op een response van de gegevensopslag is dan niet acceptabel.	X	X	
Uniekheid	Uniekheid van een data entiteit is gericht op het feit dat er geen andere entiteiten zijn met dezelfde gegevens. Voorbeeld uit de praktijk was een tweeling met dezelfde voorletters, achternaam en geboortedatum. Het onderscheid was daardoor niet te maken door een te lage compleetheid.	X	X	X
Validiteit	Dit is de mate waarin een data entiteit bij opslag en uitwisseling voldoet aan het gewenste formaat. Denk hierbij bijvoorbeeld aan het domein maar ook het datatype van de attributen van een data entiteit. Binnen ketenuitwisseling is dit bijvoorbeeld van het grootste belang. Men wil niet dat aan het einde van de keten blijkt dat de gegevens niet valide zijn. Dit dient in een vroeg stadium geconstateerd te worden. Iedereen kent de voorbeelden uit het verleden waarbij bij webapplicaties	X		

Sleutels

Bij het werken met gegevens is tot op heden de opslag in relationele databases nog steeds de meest toegepaste vorm. Een kenmerkend verschijnsel van gegevens in relationele databases is dat deze afzonderlijk en uniek worden geïdentificeerd door het gebruik van sleutels. Inrichting van relationele databases is vergaand geoptimaliseerd om met sleutels om te gaan. Bijvoorbeeld om snel te kunnen zoeken maar ook om verschillende entiteiten met elkaar te kunnen verbinden. Sleutels zijn daarom ook bij het uitwisselen en gebruiken van gegevens belangrijk omdat hiermee bepaald wordt welke gegevens in gebruik cq uitgewisseld worden zijn. Ook bij gegevens die niet afkomstig zijn uit relationele databases wordt bij uitwisseling of gebruik vaak gebruik gemaakt van sleutels om onderscheid te kunnen maken tussen de verschillende onderdelen.

Bij het gebruik van sleutels kunnen combinaties van eigenschappen gecombineerd worden. Veelal wordt ervoor gekozen om aan de eigenschappen van een entiteit een extra eigenschap toe te voegen waarmee iedere entiteit altijd uniek geïdentificeerd kan worden. Meestal is dat een nummer dat door de database zelf gegenereerd wordt.

Sleutels en registers

Bij het inrichten van applicaties die alleen gebruik maken van applicatiedata is het sleutelbeleid minder van belang, veelal kan volstaan worden met het kiezen van de sleutels (op basis van database sleutels) en het correct inrichten van verwijzende sleutels binnen de applicatie. Complexer wordt het wanneer er gebruik gemaakt gaat worden van gegevens afkomstig uit andere applicaties of uit de verschillende registersoorten. Zeker bij kern- en authentieke registers is sleutelbeleid noodzakelijk. In onderstaande opsomming een overzicht:

- **Register**, deze hebben veelal een beperkte hoeveelheid afnemers en aan de sleutels worden daarom beperkte eisen gesteld. Complexer is het in de situatie waarbij er meerdere registers zijn die dezelfde gegevensentiteiten ontsluiten. In die situatie is het uitwerken van de sleutelinrichting noodzakelijk in verband met het handhaven van de uniekheid van de gegevens.
- **Kernregister**, hierbij is sleutelbeleid van belang, met name vanwege het feit dat het aantal afnemers omvangrijker is en dat er een wisselende technologie is toegepast. Dat stelt eisen aan de platform-, technologie en inrichtingsonafhankelijkheid. Meestal wordt er in deze situatie gekozen voor het inzetten van webservices. Daarnaast is door intensief gebruik van de gegevens door meerdere afnemers performance, actualiteit en beschikbaarheid een vereiste.
- **Authentiek register**, in deze situatie is het sleutelbeleid het meest stringent. Zeker in situaties waarbij een deel van de afnemers zich buiten de eigen organisatie bevinden. Het kiezen van interne (database) sleutels is in die gevallen veelal ook niet meer voldoende. Vaak wordt er een uniek identificerende sleutel gekozen die voor alle afnemers van de gegevens inzetbaar is. Denk bijvoorbeeld aan het BurgerServiceNummer binnen de Nederlandse overheid. Daarnaast worden er hoge eisen aan de platform-, technologie en inrichtingsonafhankelijkheid. Meestal wordt er in deze situatie gekozen voor het inzetten van in vergaande mate gestandaardiseerde en beschreven webservices, bijvoorbeeld door het gebruik van digikoppling. Daarnaast is door

intensief gebruik van de gegevens door meerdere afnemers performance, actualiteit en beschikbaarheid een vereiste.

Sleutelbeleid

Sleutelbeleid kent een aantal niveaus waarbij een gestapelde uitwerking geldt. Dus applicatiebeleid geldt in combinatie met algemeen beleid en eventueel specifiek beleid indien relevant. Sleutelbeleid kan als volgt geformuleerd worden:

Applicatieniveau

- Maak indien mogelijk gebruik van de sleutel implementatiemogelijkheden binnen relationele databases. Dus inrichting op basis van primaire en verwijzende sleutels op basis van een door de database ingerichte sleutel.
- Een sleutel maakt een entiteit in de gegevensverzameling uniek
- Sleutels mogen bij voorkeur geen betekenis geven
- Sleutels mogen niet veranderen
- Sleutels mogen niet herleidbaar zijn naar een object (zoals BSN)

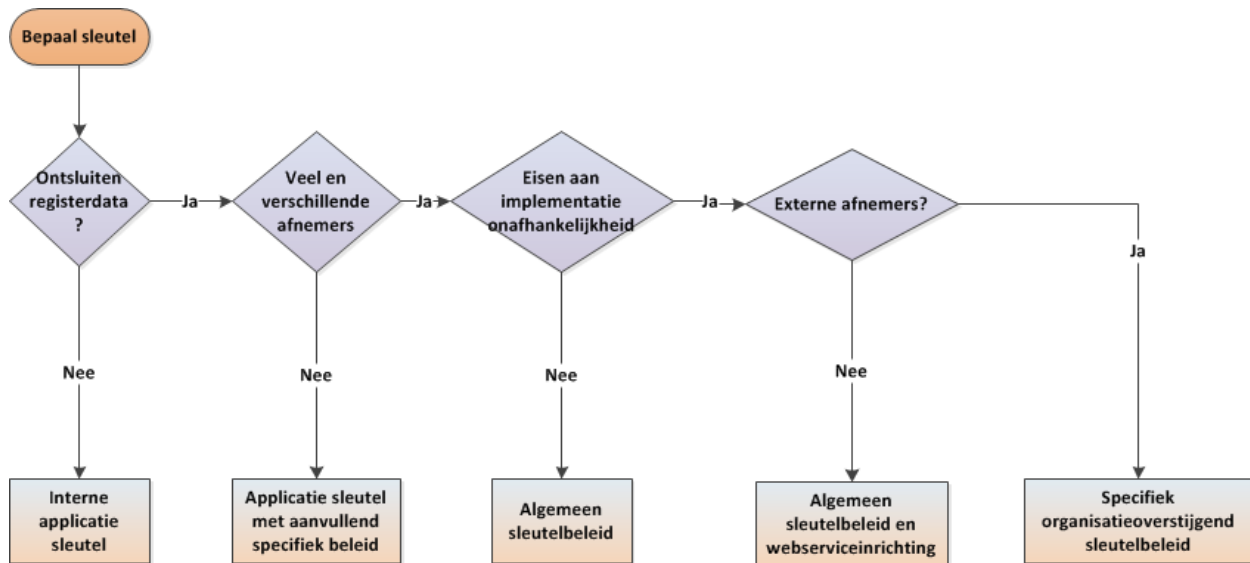
Algemeen beleid

- Bepaal unieke sleutel op basis van de database inrichting, dit kan de database sleutel of een sleutel in combinatie met een andere kolom
- Verrijk de sleutel; Is met name van de toepassing indien er meerdere registers met dezelfde gegevensentiteiten zijn
- Maak gebruik van de sleutelkast werkwijze en registreer de sleutels in het sleutelregister
- Bepaal of webservicetechnologie een vereiste is (zie onderstaande beslisboom)
- Verrijk de ontsluiting van een gegevensentiteit met alle unieke sleutelcombinaties, maak in de views en webservices eventueel gebruik van de sleutelkastfunctionaliteit (zie paragraaf 5.3).

Specifiek beleid

- Bepaal een organisatie overstijgende sleutel
- Leg het beheer van deze sleuteldefinitie, beschrijving en metadatering bij een eigenaar (meestal de eigenaar van de bron)
- Beschrijf de sleutel in het register in detail.

In onderstaande beslisboom wordt aangegeven welk beleid wanneer moet worden toegepast



Voorzieningen voor register en sleutelbeleid

Rond register en sleutelbeleid wordt aan projecten en de beheerorganisatie een aantal voorzieningen geboden waarmee het implementeren van dit beleid eenvoudig te realiseren is. In onderstaande paragrafen worden deze voorzieningen toegelicht

Centrale Gegevens Ontsluiting

De centrale gegevens ontsluiting CGO (ook bekend onder de term operational data store) draagt bij aan het implementeren van met name het registerbeleid, omdat hiermee enkelvoudig beheer en meervoudig gebruik geïmplementeerd wordt. Enerzijds worden registers via het CGO ontsloten zodat de bronnen geen performance issues ervaren door het gebruik van afnemers. Daarnaast is het mogelijk om registergegevens te combineren (logische views) binnen het CGO. Anderzijds biedt het register een aantal voorzieningen op het vlak van sleutels zoals het component voor de sleutelkast.

Sleutelconventie en sleutelregister

Voor het uniek identificeren van elementen binnen de informatievoorziening zijn sleutels essentieel. Dit kan geïmplementeerd worden in een sleutelconventie die bestaat uit de volgende onderdelen:

- Voor iedere entiteit kan een unieke sleutel bepaald worden
- Op basis van de unieke sleutel kan indien wenselijk een andere unieke sleutel bepaald worden
- De unieke sleutel wordt uitgebreid met een unieke beschrijvende naam.
- Voor het leggen van relaties kan bij gerelateerde entiteiten een verwijzende sleutel opgenomen worden met verwijzing naar de unieke sleutel in de primaire gegevensset.

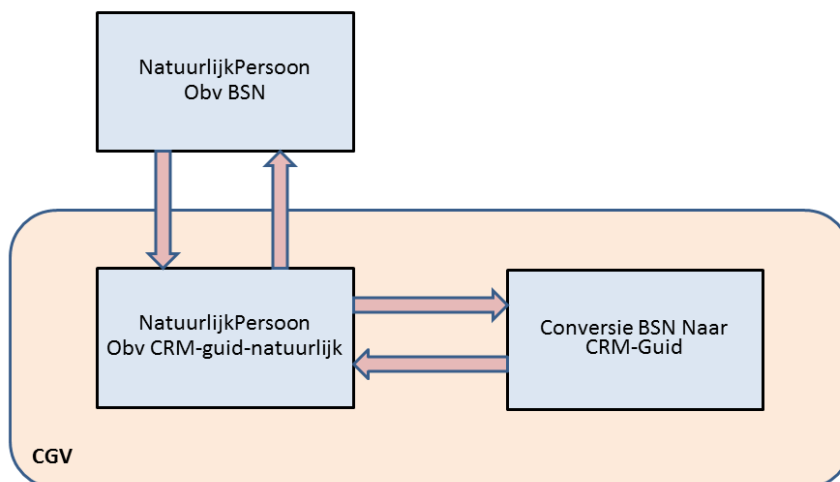
- De beschrijving van de sleutels wordt beschreven in de meta data van een gegevensset, dit kan in een register binnen het CGO of in de architectuur repository..
- De beschrijving van de sleutels wordt beheerd door de data-architect.

Sleutelkast

De werking van de sleutelkast is het beste toe te lichten aan de hand van een voorbeeld:

- De gegevens van natuurlijke personen (burgers) worden binnen de organisatie geregistreerd in een CRM systeem.
- CRM systeem maakt gebruik van guids voor het bepalen van de unieke sleutel
- Deze gegevens worden ontsloten via het CGO waarbij de guid en de omschrijving CRM-guid-natuurlijkpersoon opgenomen.
- Stel nu dat een afnemend systeem de beschikking heeft over het BSN van de natuurlijk persoon.
- De sleutelkast draagt zorg voor voor de conversie van BSN naar CRM-guid-natuurlijkpersoon en draagt zorg voor het ontsluiten van de juiste entiteit.

Onderstaande afbeelding licht dit toe.



Ontsluiting sleutelkast

De functionaliteit van de sleutelkast dient geïmplementeerd te worden binnen de technische infrastructuur. De functionaliteit vraagt het ontsluiten van meerdere alternatieve sleutels en een omzetroutine voor deze sleutels. Dit kan het beste geïmplementeerd worden binnen de databases omdat hier de sleutels opgeslagen zijn en er hierdoor minimale performance verliezen plaatsvinden.

Voorstel is om deze functionaliteit in de vorm van een aantal stored procedures en – functions te implementeren binnen het CGO. Hier zijn enerzijds de alternatieve primaire sleutels ontsloten en zal het meest gebruik gemaakt worden van deze gewenste functionaliteit.

Functionaliteit wordt geïmplementeerd als volgt:

- Webservice inrichting van de conversie routine (welke intern onderstaande views oproept)
- Ontsluiting via databaseviews waarbij een aantal stored functions opgenomen wordt voor de conversie

Tot slot

Register- en sleutelbeleid is bij een groeiende volwassenheid van de data-architectuur en het data-management steeds belangrijker. Door toenemend hergebruik ontstaan aanvullende eisen. In een vroeg stadium nadenken over de sleutels, de inzet van componenten, maar ook aan requirements van aan te schaffen informatiesystemen kan bijdragen aan een volwassen en stabiel gegevenslandschap.