Handreiking datawarehouses (DWH) en archivering

17 juni 2019
Versie 1.0
Expertisepool Recordmanagement en Stadsarchief

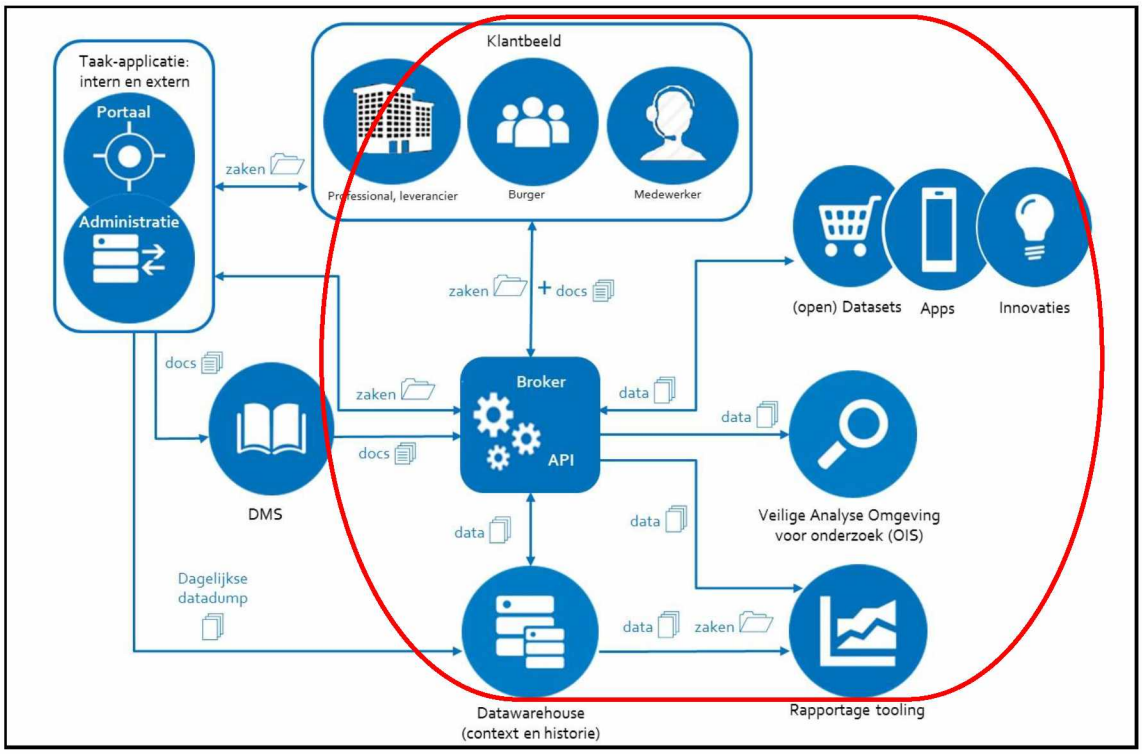
Inleiding

In de Gemeente Amsterdam is behoefte aan eenduidige afspraken voor het beheer in de zin van duurzame toegankelijkheid (zoals selectie, vernietiging en permanente bewaring) ten aanzien van datawarehouses (DWH's) en gegevensleveringen. Dit document biedt een kader en inrichtingsprincipes, voor het beheer van gegevens in datawarehouses en gegevensleveringen, alsook de output van gegevensanalyses. Deze principes helpen de IB-adviseurs bij de advisering van vraagstukken op dit vlak.

Afbakening

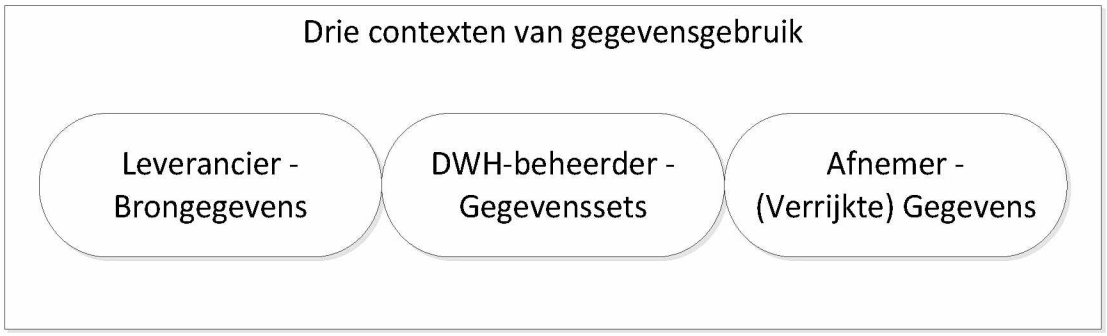
Het rood omkaderde deel van onderstaande praatplaat geeft de afbakening weer van dit document: het beheer van gegevens die in een datawarehouse zijn opgeslagen en van daaruit worden gebruikt door afnemende processen.

Buiten scope is het beheer van gegevens en documenten in bronsystemen en/of DMS'en.



• **Figuur 1: Afbakening.**

In dit kader maken we onderscheid tussen een drietal contexten. Namelijk de context van brongegevens (het perspectief van de leverancier van brongegevens), de context van gegevenssets (het perspectief van de DWH-beheerder) en de context van de geleverde/verrijkte gegevens (het perspectief van de afnemer van gegevens).



Figuur 2: Weergave van de drie contexten van gegevensgebruik en de bijbehorende rollen.

Eigenaarschap

Eigenaarschap en daarmee samenhangende verantwoordelijkheden (zoals die voor vernietiging) zijn in de regel duidelijk belegd voor bronsystemen en afnemende processen. Voor data in de omgeving daartussen, i.c. DWH's, is dat vaak niet het geval. Toepassing van de hiernavolgende pr principes begint met het beleggen van dat eigenaarschap.

Principe 1

De afnemer is verantwoordelijk voor de archivering van diens informatieproducten

De afnemer van gegevens kent aan die gegevens (nieuwe) betekenis toe en creëert daarmee nieuwe informatie. Op deze informatie is de Archiefwet van toepassing. Dit betekent dat de nieuwe context bepalend is voor het toekennen van bewaartermijnen.

De afnemer:

- Kent betekenis toe;
- Hanteert bewaartermijn;
- Treft maatregelen voor bewaring en vernietiging.

Wie?

- Afnemer: organisatieonderdeel binnen gemeente Amsterdam dat gegevens uit andere processen afneemt en hergebruikt ten behoeve van de eigen processen.

Wat?

- De afnemer is verantwoordelijk voor het beheer van informatieproducten die zijn gecreëerd met behulp van de afgenomen data;
- De afnemer is niet verantwoordelijk voor het beheer van de data in de oorspronkelijke context bij de bron.

Hoe?

Het beheer kan volgens drie patronen verlopen:

- Beschikbaarheid regelen in bron;
  - Procesafspraken door de afnemer met bronhouder;
- Beschikbaarheid in DWH;
  - Procesafspraken met DWH-beheerder;
- Beschikbaarheid bij afnemer;
  - Gegevens naar eigen beheeromgeving kopiëren.

Principe 2

Voor gegevens in een datawarehouse worden bewaartermijnen gehanteerd

DWH heeft als doel het verzamelen en beschikbaar stellen van gegevens. Bij het toekennen van bewaartermijnen wordt rekening gehouden met de werkprocessen die het DWH ondersteunt. Indien er geen bewaartermijn kan worden vastgesteld aan de hand van de afnemende processen, is het mogelijk het DWH te beschouwen als een hulpadministratie met een generieke bewaartermijn.

De DWH-gegevens:

- Krijgen een generieke bewaartermijn;
- Krijgen een bewaartermijn op basis van de processen die het DHW ondersteunt.

Wie?

- DWH-beheerder: Legt een DWH aan waarin hij gegevens beschikbaar stelt voor afnemers;
- DWH-beheerder: zorgt samen met een IB-adviseur voor een generieke bewaartermijn voor de gegevens in het DHW waarin de belangen voor de bedrijfsvoering, verantwoording en cultureel erfgoed wordt geborgd.

Wat?

- Wanneer gegevens in een DWH worden opgeslagen, dienen deze gegevens conform Archiefwet beheerd te worden;
- Een DWH moet beheersbaar blijven. Een generieke bewaartermijn voor de gegevens draagt daaraan bij.

Hoe?

- Er wordt nagegaan of aan de hand van de werkprocessen die het DWH ondersteunt een bewaartermijn kan worden bepaald. Indien er geen bewaartermijn kan worden vastgesteld, is het mogelijk het DWH te beschouwen als een hulpadministratie;
- Een afnemer kan, in overleg met een IB-adviseur, afspraken maken met de DWH-beheerder om gegevens langer in het DWH te bewaren als dit vanuit de verantwoordelijkheid van de afnemer gewenst is (zie principe 1 en bijlage 2);
- Gegevens worden na afloop van de generieke bewaartermijn conform protocol vernietigd.

Principe 3

In een datawarehouse worden geanonimiseerde gegevens opgeslagen

De persoonsgegevens in het bronsysteem worden verzameld ter uitvoering van het primaire proces. Als persoonsgegevens worden hergebruikt in een context waarin die oorspronkelijke doelbinding niet van toepassing is, dan is er geen grondslag voor verwerking van deze gegevens.

Persoonsgegevens:

- Worden conform de AVG beheerd, in de context van het proces waarvoor ze zijn verzameld (doelbinding);
- Ze zijn niet beschikbaar voor hergebruik als dit niet ondersteunend is aan de oorspronkelijke doelbinding;
- Voor andere typen gegevens geldt dit niet.

Wie?

- Bronhouder: is ervoor verantwoordelijk dat de gegevensset die aan het DWH wordt geleverd wordt geanonimiseerd of gepseudonimiseerd;
- DWH-beheerder: accepteert alleen geanonimiseerde of gepseudonimiseerde gegevens;
- Privacy-officer: levert bij wens tot afwijking van het principe een advies (risicoanalyse) aan de afnemer.

Wat?

- Het primaire proces geeft de doelbinding om persoonsgegevens te verwerken;
- Een DWH ondersteunt in principe niet dit proces, en ook niet de archivering ervan, zodat de doelbinding niet van toepassing is en daarmee is de verwerking niet rechtmatig;
- Afwijkingen hiervan zijn mogelijk na advies (risicoanalyse en beheersmaatregelen) van de privacy-officer.

Hoe?

Er is een aantal manieren om dit te regelen:

- Beperking van de toegang: de gegevens blijven beschikbaar, maar zijn via autorisatiebeheer alleen te benaderen door een wel omschreven groep, voor een wel omschreven doel;
- Anonimisering: de gegevens zijn niet meer tot een natuurlijke persoon te herleiden, dit is onherstelbaar;
- Pseudonimisering: de gegevens zijn niet meer tot een natuurlijke persoon te herleiden, dit is herstelbaar;
- Dataset met persoonsgegevens niet opnemen;
- Bij afwijkingen van dit principe is het advies (risicoanalyse) van de privacy-officer nodig, en is de eigenaar van het risico verantwoordelijk voor de beheersmaatregelen.

Principe 4

De afnemer heeft inzicht in de kwaliteit van de door hem gebruikte gegevens

Voor de afnemer is het van belang om inzicht te hebben in de kwaliteit van de gebruikte gegevens. Die bepaalt namelijk de betrouwbaarheid van diens informatieproducten.

Kwaliteit van gegevens:

- Bepaalt deels de relevantie van gegevens voor een afnemer;
- Moet inzichtelijk zijn;
- Bestaat uit verschillende kwaliteitsaspecten.

Wie?

- Bronhouder: De bronhouder is verantwoordelijk voor het borgen van kwaliteit in de bron en het inzichtelijk maken van het kwaliteitsniveau;
- DWH-beheerder: De DWH-beheerder voert regie op de kwaliteit van alle gegevenssets in het DWH door generieke kwaliteitseisen te formuleren en het gebruik van een dataregister voor vastlegging van de kwaliteit te stimuleren;
- Afnemer: Neemt kennis van de kwaliteit van de afgenomen gegevens om de bruikbaarheid te bepalen.

Wat?

Kwaliteit bestaat uit een aantal deelaspecten (bron: NEN 27001):

- Volledig (er ontbreken geen relevante gegevens en de gegevens zijn correct);
- Actueel (gegevens zijn up-to-date);
- Betrouwbaar (beveiligd tegen onbedoelde of ongeautoriseerde wijzigingen).
- Toegankelijk (gegevens zijn interpreteerbaar).

Hoe?

- Kwaliteitseisen zijn generiek benoemd en vastgelegd in datacatalogus;
- Geconstateerde gebreken in de kwaliteit worden gemeld aan de bronhouder, hiervoor is een proces ingericht;
- Kwaliteitsverbetering vindt plaats in de bron, te denken valt aan:
    - Metadatering gebeurt conform TMLO en/of vakspecifieke standaard;
    - In de bron wordt zoveel mogelijk met gecontroleerde tabellen gewerkt en waarden zijn waar mogelijk identiek aan die in basis- en kernregistraties;
    - Wijzigingen worden gelogd;
    - Frequentie van gegevenslevering aan DWH sluit aan op de veranderlijkheid in de bron;
    - Er zijn maatregelen getroffen om historische gegevens te kunnen blijven gebruiken zo lang de Selectielijst dat voorschrijft.

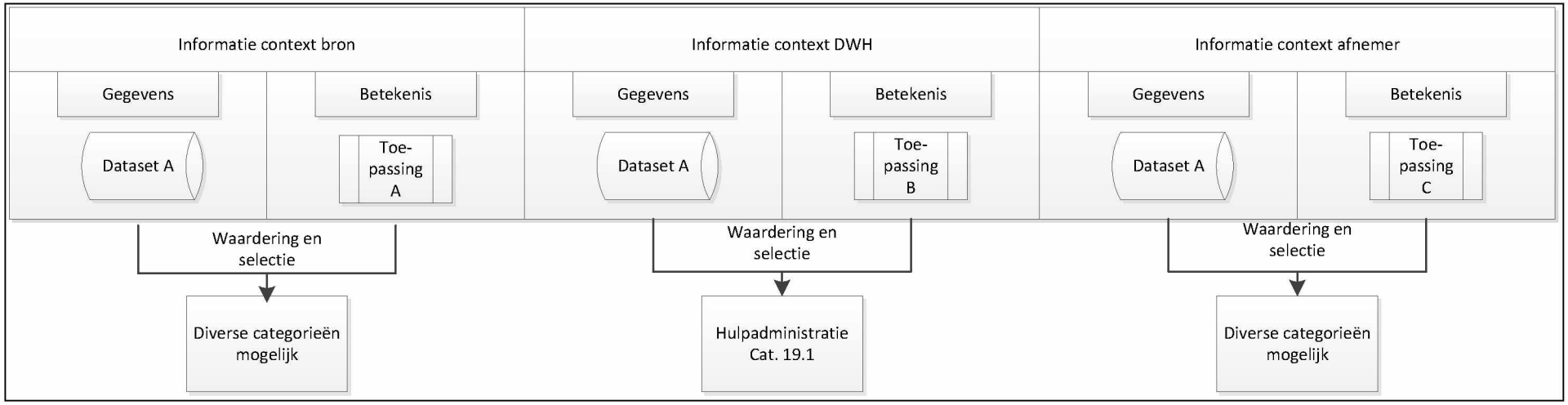
Bijlage 1: Definities

- Beheren van managementinformatie
  - Functionaliteit voor het creëren en tonen van managementinformatie. Hieronder valt functionaliteit voor het integreren van gegevens en het maken en tonen van standaard en custom rapportages of trendanalyes.
  - Bron: https://www.gemmaonline.nl/index.php/GEMMA2/0.9/id-8280288e-da88-11e1-7023-0050568a1905.
- BI-tooling
  - Zie: beheren van managementinformatie.
- Broker
  - Zie: gemeentelijke servicebus.
- Data
  - Zie: gegevens.
- Datawarehouse
  - In een datawarehouse worden actuele en historische gegevens opgeslagen voor het uitvoeren van trendanalyses en verantwoordingsrapportages. Een datawarehouse (vaak afgekort tot DWH) is een gegevensverzameling die in een dusdanige vorm is gebracht dat terugkerende en ad-hoc vragen relatief in korte tijd beantwoord kunnen worden zonder dat de bronsystemen zelf daardoor overmatig belast worden.
  - Bron: https://www.gemmaonline.nl/index.php/GEMMA_1/6/id-e2f5dffa-a445-11e3-67ab-0050568a6153).
- Gegevens
  - De weergave van een feit, begrip of aanwijzing, geschikt voor overdracht, interpretatie of verwerking door een persoon of apparaat. Synoniem meervoud: data (gegevens).
  - Bron: https://archiefwiki.org/wiki/Gegeven (19122017).
- Gemeentelijke servicebus
  - Systeem waarmee koppelingen tussen gemeentelijk systemen gerealiseerd worden. Een gemeentelijke servicebus is in de basis een generieke Enterprise Service Bus (ESB) waarmee gemeente specifieke koppelingen worden gerealiseerd. Een ESB biedt minimaal functionaliteit voor het versturen en beheren van elektronische berichten. Tevens kan een ESB services aanbieden voor het routeren, transformeren en eventueel orchestreren van het berichtenverkeer. De component die deze services aanbiedt, wordt ook een Integratie server of een broker genoemd.
  - Bron: https://www.gemmaonline.nl/index.php/GEMMA_1/6/id-599ddcd1-8861-11e3-67ab-0050568a6153.
- Generieke bewaartermijn
  - Bewaartermijn die standaard wordt toegepast op typen datasets van hetzelfde type, indien die niet in de context van een afnemend proces zijn voorzien van een specifieke bewaartermijn.
  - In de selectielijst staan generieke procestypen die hiervoor kunnen worden gebruikt. In de selectielijst die vigeert ten tijde van opstelling van dit stuk, kan bijvoorbeeld worden gedacht aan 19.1 (hulpadministratie zonder financiële gegevens) of 19.2 (niet-verwerkte gegevens) als bijvoorbeeld ruwe data wordt opgeslagen die enkel in geabstraheerde vorm wordt gebruikt. Belangrijk is om zorgvuldig te kijken naar de aard en functie van een datawarehouse om hierin de juiste keuze te maken.
- Hulpadministratie
  - Een registratie waarvan de gegevens geheel of gedeeltelijk, al dan niet in geaggregeerde vorm, in andere primaire registraties worden opgenomen.

    - Bron: Zie: https://labyrinth.rienkjonker.nl/lexicon/terminologie#letter_h
    - In de context van dit document betekent deze definitie dat gegevens geheel of gedeeltelijk, al dan niet in geaggregeerde vorm, naar een aparte omgeving (datawarehouse) worden gekopieerd om van daaruit te worden hergebruikt in andere primaire registraties.
- Informatie
    - Betekenisvolle gegevens (data).
    - Bron: https://www.noraonline.nl/wiki/Informatie (19122017).
- Klantbeeld
    - Een overzicht van de voorkeuren en behoeften van de klant op basis van geaggregeerde klantcontactinformatie.
    - Bron: https://www.noraonline.nl/wiki/Klantbeeld.
- Rapportagetooling
    - Zie: beheren van managementinformatie.

Bijlage 2: Waarderingsmechanisme per informatieperspectief

Kenmerkend voor het gebruik van datawarehouses, is dat dezelfde dataset meerdere toepassingen heeft. Onderstaande afbeelding geeft dit schematisch weer. Aan gegevenssets in een datawarehouse kan een default-bewaartermijn worden toegekend, die wordt overruled door een bewaartermijn vanuit een inhoudelijke toepassing.



• **Figuur 3: Schematische weergave van de drie informatieperspectieven met bijbehorend waarderingsmechanisme.**