**Startdocument werkgroep archiveringsvraagstukken datawarehouses**

1. **Inleiding**

In de Gemeente Amsterdam is behoefte aan eenduidig beleid voor het beheer in de zin van duurzame toegankelijkheid (zoals selectie, vernietiging en permanente bewaring) ten aanzien van datawarehouses en gegevensleveringen. De werkgroep buigt zich over hieraan gerelateerde vraagstukken. Het beoogde resultaat van de werkgroep is een set aan kaders en inrichtingsprincipes voor de duurzame toegankelijkheid van gegevens in datawarehouses en gegevensleveringen alsook de output van gegevensanalyses.

1. **Afbakening**

Het rood omkaderde deel van onderstaande praatplaat geeft de afbakening weer van de vraagstukken waar de werkgroep zich mee bezig houdt: het beheer van gegevens die in een datawarehouse zijn opgeslagen en van daaruit worden gebruikt door afnemende processen.

Buiten scope zijn vragen die zich primair richten op het beheer van gegevens en documenten in bronsystemen en/of DMS’en.



**• Figuur 1: Weergave van de afbakening van de vraagstukken.**

1. **Definities**

De werkgroep hanteert de volgende 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.
* 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 geagregeerde klancontactinformatie.
  + Bron: https://www.noraonline.nl/wiki/Klantbeeld.
* Rapportagetooling
  + Zie: beheren van managementinformatie.

1. **Onderzoeksvragen**

Voor het categoriseren van de onderzoeksvragen, 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 datawarehousebeheerder) en de context van de (verrijkte) gegevens (het perspectief van de afnemer van (verrijkte) gegevens).

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

* 1. **Vragen vanuit de samenhang tussen de verschillende contexten (zie plaatje).**
* Wie draagt welke verantwoordelijkheid voor het beheer van gegevensverzamelingen in een datawarehouse: er is een bronhouder, een beheerder en een afnemer. Wie doet wat qua archivering?
* Welke typen gegevenssets moeten we onderscheiden om het gepaste beheerregime te kunnen bepalen?
  + Bijvoorbeeld kern-/basisregistraties, opendatasets, sets met/zonder persoonsgegevens, origineel/kopie, geanonimiseerde data zoals onderzoeksdata.
  1. **Vragen vanuit de context van gegevenssets**
* Is er sprake van nieuwe c.q. andere informatie nadat gegevens vanuit een bron naar een datawarehouse worden gekopieerd?
* In hoeverre moeten gegevens in een datawarehouse worden beheerd (in de zin van duurzame toegankelijkheid), wat is daarbij de samenhang met de corresponderende gegevens in de bron?
* Moet een dataset vernietigd worden als de brongegevens vernietigd zijn? Met andere woorden: mag je na vernietiging van de brongegevens nog een kopie in een datawarehouse houden?
  + Zo ja, onder welke voorwaarden?
  + Zo nee, hoe kunnen we dit in de praktijk organiseren?
* Kunnen gegevenssets in een datawarehouse voor blijvende bewaring in aanmerking komen?
  + Zo ja, waar en hoe bewaar je dergelijke gegevenssets?
* Welke (voorname) datawarehouses zijn er in de stad?
  + Welke data bevatten deze?
  + Voor welke doelen worden deze gebruikt?
  1. **Vragen vanuit de context van (verrijkte) gegevens**
* Moeten acties ten behoeve van gegevensanalyse (query’s/algoritmes, analyses) reconstrueerbaar zijn?
  + Welke voorwaarden (o.a. bewaartermijnen) zijn hierop van toepassing?
  + Welke vormen/methodieken zijn mogelijk?
  + Welke impact hebben veranderende datasets op het beheerregime?
* Moeten resultaten (rapportages, dashboardgegevens etc.) van gegevensanalyse bewaard worden of reconstrueerbaar zijn?
  + Welke voorwaarden (o.a. bewaartermijnen) zijn hierop van toepassing?
  + Welke vormen/methodieken zijn mogelijk?

1. **Backlog vervolgvragen**

Uit de kick-off-sessie is een aantal vragen naar voren gekomen die de werkgroep als minder urgent heeft ingeschat of die nog nadere uitwerking behoeven. Deze vragen zijn opgenomen in de backlog met vervolgvragen. De backlog wordt periodiek nagelopen. Het gaat om de volgende vragen:

* Hoe kunnen we de waarde van data bepalen?
* Hoe garandeer je dat een datawarehouse compliant is met wet- en regelgeving?
  + Welke wet- en regelgeving is relevant (AVG, Archiefwet)?
* Hoe hanteren we bewaartermijnen in de geest van de oorspronkelijke regeling?
* Kunnen we, in plaats van bewaartermijnen voor data, toepassingstermijnen hanteren waarbij het doel bepaalt welke datasets je (nog) mag gebruiken?
* In hoeverre is het mogelijk data te vernietigen?
* Wat doet de Expertisepool Gegevensbeheer op dit vlak?