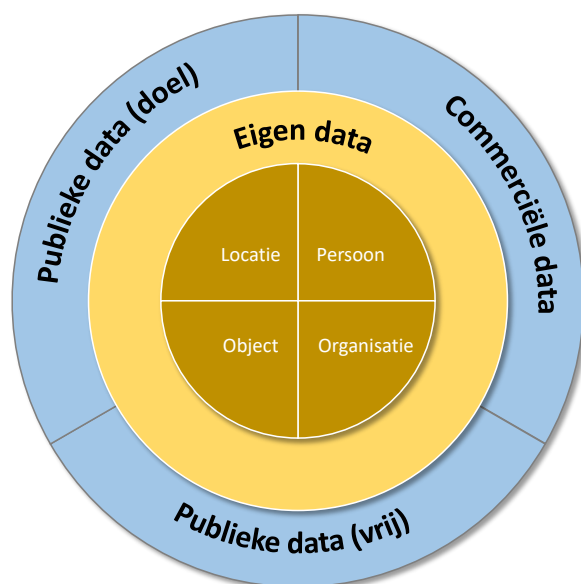




Toelichting SIVI Catalogus Datasets



Inhoud

1	Inleiding.....	2
1.1	Aanleiding.....	2
1.2	Uitgangspunt SIVI Catalogus Datasets	2
1.3	Opzet notitie.....	2
2	Context.....	2
2.1	Belang van data.....	2
2.2	Externe data	3
3	SIVI Catalogus Datasets.....	4
4	Prioriteiten 2020	5

1 Inleiding

1.1 Aanleiding

In de laatste 20 jaar nam het belang van data in zijn algemeenheid maar zeker ook binnen verzekeringsprocessen sterk toe. In de “Blik op 2019-2020”¹ signaleert SIVI een aantal ontwikkelingen rond een meer data gedreven invulling van verzekeringsprocessen. Dit betreft onder meer het creëren van inzicht (rapportage), productontwikkeling, risicoanalyse en acceptatie. Het combineren van interne eigen data met extern beschikbare data is voor de verdere uitbouw van de benutting van data randvoorwaardelijk.

Steeds meer data komt vrij beschikbaar onder de noemer Open Data². Overheden, semioverheden en zelfs commerciële organisaties bieden gratis of tegen beperkte vergoeding datasets aan. Veel (semi-)overheidsinstellingen leveren vanuit de eigen processen naast open data ook data tegen betaling. Het verschil zit deels in de actualiteit en toegankelijkheid van de data. Daarnaast zijn er meer specialistische datasets of datasets met een specifieke doelbestemming waarop toezicht moet plaatsvinden. Commerciële datasets kunnen als aanvulling op bovenstaande datasets zeer waardevol zijn. Het is vanuit het perspectief van deze waarnemingen belangrijk om inzicht en overzicht op te bouwen rond beschikbare datasets, de kwaliteit en de opties rond het ontsluiten hiervan.

1.2 Uitgangspunt SIVI Catalogus Datasets

De SIVI Catalogus Datasets levert een overzicht van voor het verzekeringsdomein relevante datasets, de wijze van beschikbaarheid, de onderliggende data-definities en kwaliteitskenmerken.

De catalogus is functioneel van opzet en daarmee toegespitst op beleidsmakers/adviseurs en informatie- en procesdeskundigen. Partijen actief in de verzekeringsindustrie kunnen SIVI verzoeken de catalogus met specifieke bronnen uit te breiden. SIVI neemt dit dan op in haar planning voor de catalogus.

1.3 Opzet notitie

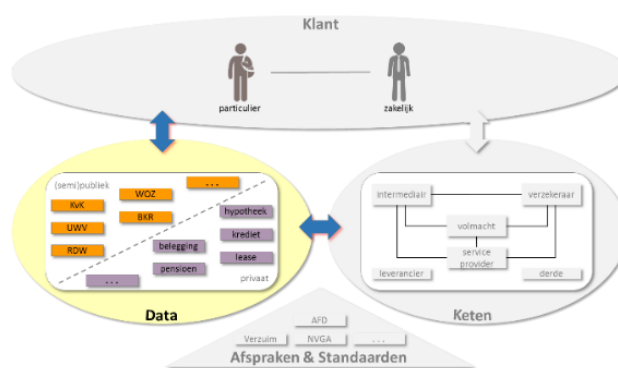
Na deze inleiding wordt ingegaan op de context waarbinnen de SIVI Catalogus Datasets wordt ontwikkeld. Daarna volgt een beschrijving van de catalogus in zijn huidige staat. Besloten wordt met een hoofdstuk rond de geplande vervolgstappen voor 2020.

2 Context

2.1 Belang van data

In de “Blik op 2019 – 2020” staat SIVI stil bij de centrale plaats van data in het digitaal ecosysteem voor verzekeringen (figuur 1). Het gaat daarbij om data rond objecten, locaties, organisaties en personen in de meest brede zin. Binnen processen is de inzet van data noodzakelijk voor geautomatiseerde verwerking (STP) en het inrichten van self service functies met minimale drempels. Ten aanzien van analyse is er toenemend een noodzaak voor de beschikbaarheid van data ten behoeve o.a. actuariële processen, premieberekening, acceptatie en het voorkomen / onderkennen van fraude.

Beschikbare eigen data binnen de organisatie verstrekt door de klant, afkomstig uit monitoring of afgeleid is belangrijk voor bijvoorbeeld het vaststellen van risico's, marketing acties of klantbediening. Met name eigen afgeleide data (bijv. risico indicatoren) zijn steeds vaker mede afgeleid van externe data.



Figuur 1: Digitaal ecosysteem (“Blik 2019-2020”)

¹ www.sivi.org/actueel/sivi-voortuitblik-2019-2020

² Open data is data die vrij gebruikt kan worden, hergebruikt kan worden en opnieuw verspreid kan worden door iedereen - onderworpen enkel, in het uiterste geval, aan de eis tot het toeschrijven en gelijk delen. (bron: http://opendatahandbook.org/guide/nl_BE/what-is-open-data/ - Status september 2019)

Praktijk voorbeelden rond de inzet van externe data:

- Gemak: Er wordt steeds minder data gevraagd aan de klant (bijv. gewicht voertuig).
- Risico: Toetsen omdat het kan (eigendom pand) of moet (UBO).
- Verkoop: houdt de ontwikkeling van het zoekvolume bij Google Trends voor 'premie AOV' gelijke tred met de activiteiten op de eigen website of het eigen extranet?
- Schade: sluiten de schademeldingen rond hagelschade aan bij de neerslagrapportage van het KNMI?
- Product management: sluit de herbouwwaarde van een opstalportefeuille aan bij een berekening op basis van externe datasets?

Het belang van het combineren van eigen data met externe bronnen (figuur 2) neemt verder toe binnen alle fasen van het verzekeringsproces. Dit maakt het belangrijk een goed beeld op te bouwen rond de beschikbaarheid, de kwaliteit en de ontsluiting van die datasets.

2.2 Externe data

Overheden, semioverheden en zelfs commerciële organisaties bieden gratis of tegen beperkte vergoeding datasets aan. Deze data is openbaar en vrij te gebruiken. Voorbeelden zijn data.overheid.nl (economie en samenleving), KNMI (weer en klimaat), RDW (voertuigen) en KvK (bedrijven).

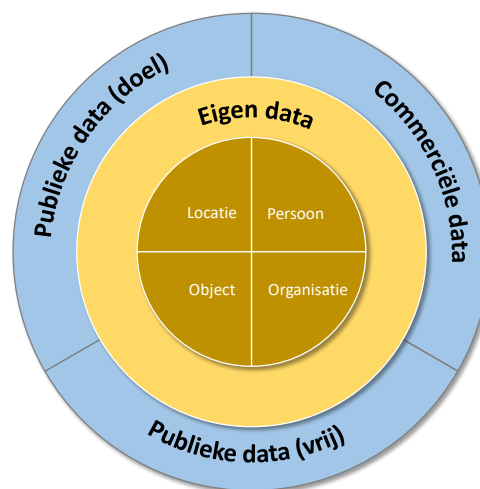
Veel (semi)overheidsinstellingen leveren vanuit de eigen processen naast open data ook data tegen vergoeding. Het verschil zit deels in de actualiteit en toegankelijkheid van de data. Daarnaast gaat het om meer specialistische datasets of datasets met een specifieke doelbestemming waarop toezicht moet plaatsvinden. Deze (semi-)overheidsinstellingen en de verzekeringsindustrie hebben een gezamenlijk belang rond niet-commerciële thema's zoals bijvoorbeeld fraudeonderzoek of algemene risicopreventie. Het is belangrijk om als sector oog te hebben voor de mogelijkheden voor samenwerking bij het ontsluiten van specifieke data rond thema's die de BV Nederland raken.

Commerciële datasets kunnen in aanvulling op bovenstaande datasets zeer waardevol zijn. Sectorbrede normen rond te leveren data-elementen en datakwaliteit kunnen een level playing field creëren voor leveranciers en daarmee een incentive om te investeren in diensten.

Bij de toenemende inzet van externe datasets ontstaan nieuwe risico's als sectorbreed clustering plaatsvindt rond het gebruik van één specifieke dataset of van datasets. Deze risico's nemen toe als partijen aanvullend op basis van deze data convergeren naar één systematiek voor bijvoorbeeld een specifieke risicobeoordeling of element van premieberekening. Ook ontstaan risico's als autonome processen (bijv. acceptatie van verzekering) datasets gebruiken die – ongezien – fouten kunnen bevatten. Fouten in datasets kunnen door automatische verwerking leiden tot onjuiste uitkomsten / beoordelingen met in potentie grote impact. Normen voor en inzicht in de kwaliteit van datasets en de transparantie rond het proces van totstandkoming van gebruikte datasets worden belangrijk.

De toegankelijkheid van datasets verschilt:

- Data is open en gratis toegankelijk voor iedereen (als de voorraad woningen bij CBS).
- Data is toegankelijk tegen betaling (als basisregistratie personen (BRP)).
- Data is toegankelijk met doelbestemming (als tenaamstelling voertuig bij RDW).
- Data is alleen toegankelijk na tussenkomst van de feitelijke eigenaar (als data uit MijnPensioenOverzicht).



Figuur 2: Datasoorten

Gebruikte terminologie:

Dataverancier	De RDW is een dataverancier.
Databron	Het Kentekenregister van de RDW is een databron.
Dataset	Deelverzameling binnen de databron: <ul style="list-style-type: none"> • Voertuigdata op basis van kenteken (vrij beschikbaar) • Eigendomsdata op basis van kenteken (doelbestemming)

Met name voor de laatste twee punten geldt dat deze data niet zonder drempels beschikbaar is. Punt van aandacht is de wijze waarop deze data toegankelijk gemaakt kan worden binnen o.a. inventarisatie- en adviessoftware.

Met de SIVI Catalogus Datasets wil SIVI uiteindelijk een compleet overzicht bieden van voor het verzekeringsdomein relevante datasets, de wijze van beschikbaarheid, de onderliggende data-definities en kwaliteitskenmerken. De catalogus zal aanvullend de basis zijn voor eventuele afstemming met dataleveranciers rond sectorspecifieke dataleveringen en eventuele sectorspecifieke kwaliteitseisen.

3 SIVI Catalogus Datasets

De catalogus van dataleveranciers wordt breed opgezet enerzijds met datasets waarvan evident is dat ze een rol spelen binnen verzekeringsprocessen anderzijds –waar dat logisch lijkt– ook met datasets waarvan op voorhand niet direct duidelijk is in welke mate er toegevoegde waarde is voor de verzekeringsmarkt.

De SIVI Catalogus Datasets bevat een functionele beschrijving van de data zodanig dat partijen deze informatie kunnen gebruiken in analyse- en onderzoekstrajecten. In onderstaande tabel is een overzicht opgenomen van de informatie die is vastgelegd in deze eerste versie van de catalogus.

Rubriek	Omschrijving
Titel	De naam van de dataset.
Basissleutel	Een gegeven waarmee een dataset direct kan worden benaderd, bijvoorbeeld een kenteken, of een combinatie van postcode en huisnummer waarmee een auto of een pand direct kan worden gevonden.
Leverancier	De partij die de dataset beschikbaar stelt.
Conditie	Onder welke voorwaarden de gegevens kunnen worden verkregen of gebruikt. Wordt de informatie bijvoorbeeld gratis, tegen vergoeding of op basis van wederkerigheid ter beschikking gesteld.
Informatietype	Aanduiding of sprake is van: <ul style="list-style-type: none"> • “brongegevens”; gegevens uit de eerste hand. • “verzamelde gegevens”; gegevenssets die met een bepaald doel zijn samengesteld uit verschillende brongegevens. • “afgeleide informatie”; (in een proces ingesloten) informatie gebaseerd op brongegevens en/of verzamelde gegevens.
Doelbinding	Aanduiding of de gegevens slechts raadpleegbaar zijn door personen/bedrijven die voldoen aan vooropgezette criteria of door iedereen.
Omschrijving	Verdere uitleg over de dataset.
Soort	Aanduiding of de dataset kan worden gedownload, getoond of (als service) bevroegd kan worden.
Bron	Een verwijzing naar de bron (de vindplaats) van de dataset.
Update frequentie	Duidt aan hoe vaak de gegevens worden vernieuwd.

4 Prioriteiten 2020

Gedurende 2020 wordt de catalogus in een 3-tal releases uitgebreid en aangevuld, de prioriteiten zijn:

1. Verbreding van de catalogus (meer datasets) en in het verlengde hiervan starten met de opbouw van een relatiernetwerk van dataleveranciers.
2. Doornemen van de catalogus met branchepartijen voor accenten bij verdere uitbreiding en verdieping.
3. Leveranciers van datasets uitvragen over (het toezicht op) kwaliteit van de datasets, gericht op het vermijden van systeemfouten.
4. Onderzoeken op welke wijze de sector kwaliteitsnormen kan ontwikkelen voor specifieke datasets.