

OV campus sessie Data voor Dummies

22 november 2022

Jaap, Martin en Marloes



Samenwerkingsverband
van decentrale
OV-autoriteiten

Making Everything Easier™

(Big)Data

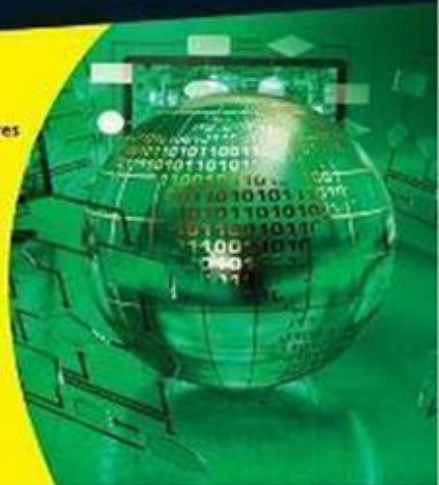
FOR
DUMMIES
A Wiley Brand

Learn to:

- Leverage big data tools and architectures
- Explore how big data can transform your business
- Integrate structured and unstructured data into your big data environment
- Use predictive analytics to make better decisions

DOVA Special edition

Marloes Verhoeven
Martin van Vuure
Jaap de Kleine



Hoofdstuk #1

Wat is data?

Hoofdstuk #2

Welke vormen van data zijn er?

Hoofdstuk #3

Welke datasets zijn er bij OV-data?

Hoofdstuk #4

Hoe maak je van data <chocolade> correctie; informatie?

Hoofdstuk #5

Hoe gebruik je informatie voor beleid of kennis?

Hoofdstuk #6

Benodigdheden

DOVA

Hoofdstuk #1 Wat is data?

Dataset

Een **gegevensverzameling** of **dataset** is een **verzameling** van **gegevens** (data), meestal gepresenteerd in **tabelvorm**. Elke **kolom** vertegenwoordigt een bepaalde **variabele**. De dataset representeert dan een verzameling **rijen** (ook *records* genoemd), elk voor één object (in de ruime zin van het woord, het kunnen bijvoorbeeld ook personen zijn). Een rij bevat de waarden voor elk van de variabelen, of attributen, voor het betreffende object, bijvoorbeeld lengte en gewicht.

Een verzameling rijen is op zichzelf niet geordend. Een weergave met de rijen in een andere volgorde is dan niet de weergave van een andere dataset. Ordening kan op basis van de waarde van een variabele. De rijen kunnen ook geordend zijn, los van de waarden van de variabelen.

Als de datatabel van de al of niet geordende dataset maar één kolom heeft wordt deze wel vereenzelvigd met de betreffende (respectievelijk niet of wel) geordende multiset (die eventueel een verzameling is).

Een gegevensverzameling niet in tabelvorm kan de vorm hebben van een **karacterstring**, zoals een **XML**-bestand.

Bron; wikipedia

Hoofdstuk #2 Welke vormen van data zijn er?

Datatype

Een **datatype**, ook wel **gegevenstype** genoemd, is in de **informatica** een specifiek soort gegevens, zoals **integers**, **booleans**, **reals**, **karakters**, **strings**. In een **programmeertaal** wordt met iedere **variabele**, en meer in het algemeen met iedere **expressie**, een datatype geassocieerd. Dit datatype bepaalt welke waarden de variabele of de expressie kan aannemen, hoe deze waarden in het geheugen worden opgeslagen en welke bewerkingen op de variabele of de expressie uitgevoerd kunnen worden.

Bij elk gegevenstype hoort een verzameling van de mogelijke waarden die een variabele of expressie van dat type kan aannemen. Ook hoort er een systeem bij voor de codering van de waarden. Een waarde kan alleen al binnen één programmeertaal soms door meerdere datatypes gerepresenteerd worden: zo kan het getal 5 gerepresenteerd worden door diverse typen integer en real (er zijn vaak meerdere typen van elk, met verschillende aantallen bits).

Bron; wikipedia

Hoofdstuk #2 Welke vormen van data zijn er?

Indeling [\[bewerken | brontekst bewerken \]](#)

Gegevenstypes kunnen worden onderscheiden in primitieve (*primitive*), enkelvoudige (*simple*) en samengestelde (*complex*) types. Primitieve datatypes vormen de basis voor de definities van andere gegevenstypes.

Primitief type [\[bewerken | brontekst bewerken \]](#)

Een *primitief type* wordt door de taal zelf gedefinieerd en kan niet beschreven worden in termen van een ander datatype. In de programmeertaal C, bijvoorbeeld, zijn *char*, *int* en *float* primitieve types.

Algemeen voorkomende primitieve typen zijn:^{[1][2]}

- *boolean*, ook bekend als *bool*, *flag* of *logic*. Kan de waarde *ja* of *nee* bevatten. Andere benamingen voor deze waarden zijn *waar* en *onwaar*, of, gebruikelijker, het Engelse *true* en *false*.
- *karakter*, ook bekend als *character* of *char*. Deze kan precies één ANSI- of EBCDIC- of Unicode-teken bevatten. Het aantal bytes dat dit type inneemt hangt af van de taal. Historisch was dat meestal 1 byte, maar tegenwoordig ondersteunen veel talen Unicode en worden er meer bytes gereserveerd voor een variabele van type *char*.
- *integer*, ook bekend als *int*, *short*, *long*, *signed* is de representatie voor gehele getallen, hoewel moet worden benadrukt dat de twee niet hetzelfde zijn, een *integer* heeft nu eenmaal een beperkt bereik, terwijl een geheel getal dat niet heeft. Meestal worden definities zo gekozen dat een integer in een register past, maar dit is geen wet van Meden en Perzen, aangezien bewerkingen op een 64-bit integer vrij gemakkelijk kunnen worden verdeeld in twee 32-bit bewerkingen of zelfs 4 16-bit bewerkingen.

Hoofdstuk #2 Welke vormen van data zijn er?

Indeling [\[bewerken | brontekst bewerken \]](#)

- *real*, ook bekend als *float*, *single*, *double*; alle niet gehele getallen. Voor het bereik van een *real* geldt hetzelfde als dat van een integer.
- *decimal*, om decimale breuken zonder afrondingsfouten te representeren
- *void*, ook wel bekend als *null* of *unit*. Het type *void* heeft geen waarde. Dit type duidt het ontbreken van een waarde aan.^{[3][4]}

Bron; wikipedia

Hoofdstuk #3 Welke datasets zijn er bij OV-data?

Voorbeelden van datasets binnen DOVA

- Facturen overzicht OV-data (Excel bestand)
- Dienstregeling van Vervoerders volgens KV1 (type)
- Dataset voor de Drukke indicator (type)

Waarvoor wordt dit gebruikt?

Intern gebruik

Reisinformatie voor DRIS panelen en open data loketten

Reisinformatie in planners via open data loketten EN voor datakwaliteitscontrole en vergelijking

Hoofdstuk #3 Welke datasets zijn er bij OV-data?

	A	B	C	D	E	F	G	H	I
1	organisatie	excl BTW	BTW	incl BTW	datum	factuurnummer	project	kenmerk	project
2	Cyclomedia				31-12-2021	20211855	71000 BD1 D022		Licentievergoeding beeldmateriaal en software 2021 (50% na acceptatie beeldmateriaal)
3	Cyclomedia				31-12-2021	20211880	71000 BD1 D022		Hosting en gebruik beeldmateriaal 1-31 december 2021
4	Fitch Data				31-12-2021	2021061		202107-01	Diensten Lisa Tostrams december 2021
5	Rootnet				31-12-2021	202104359	71000 BD10 D057	202011-02	Hosting CDD
6	Flowerbed				1-1-2022	V220029	71000 BD1 D017	202201-06	Hosting Splunk
7	Flowerbed				1-1-2022	V220030	71000 BD1 D017	202102-03	KPI dashboard
8	Pro10				3-1-2022	2022/01/862/01	202109-03	GRB 4054	aanbesteding DRIS server
9	Supp-IT				7-1-2022	SI2201	71000 BD10 D057	202102-04	koppelvlak Open DRIS
10	Flowerbed				14-1-2022	V220059	71000 BD1 D021	02110-02	Data Analytics Q4-2021
11	Flowerbed				14-1-2022	V220060	71000 BD10 D057	202110-07	CDD realisatie
12	Walta consulting				19-1-2022	211201	70000 ONTW4 D003	202102-05	projectleiderschap ICT vervangingsinvestering
13	Walta consulting				19-1-2022	211202	70000 ONTW4 D003	202109-04	Ondersteuning aanbesteding DRIS server
14	Alex van Wezel Beheer BV				20-1-2022	2022_002	70000 ONTW4 D003	202102-09	uren clockwise
15	de Roostermakers				20-1-2022	2022-006	70000 ONTW8 D054	202101-08	OV ladekast
16	Translink				25-1-2022	11822			Informatiehuishouding jaar 4
17	Flowerbed				27-1-2022	V220100	71000 BD1 D017	202201-05	Splunk licentie
18	Ivengi				31-1-2022	20220037	202102-06	71000 BD6a D030	Drismelding.nl
19	Staac				31-1-2022	2022-001	71000 BD3	20191201	DOVA schrijfwijzer
20	Easyhosting				1-2-2022				Easy 6 en Easy 1
21	Flowerbed				1-2-2022	V220129	71000 BD1 D017	202201-06	Hosting Splunk
22	Walta consulting				2-2-2022	220102	70000 ONTW4 D003	202109-04	Ondersteuning aanbesteding DRIS server
23	Irias				4-2-2022	2022-03	75000 OV83 D082	202111-03	doorontwikkeling laadkaart bus
24	Pro10				11-2-2022	2022/02/862/01	202109-03	GRB 4054	aanbesteding DRIS server
25	Irias				13-2-2022	2022-01	70000 ONTW7 D051	202101-07	doorontwikkeling basisregistratie OV
26	Irias				14-2-2022	2022-05	70000 ONTW7 D051	202201-01	doorontwikkeling basisregistratie OV
27	Irias				14-2-2022	2022-06	70000 ONTW5 D014	202201-02	ontwikkeling dashboard OverOV
28	Irias				14-2-2022	2022-07	70000 ONTW7 D051	202201-03	ontwikkeling CHB
29	Irias				14-2-2022	2022-08	70000 ONTW5 D014	202201-04	ontwikkeling analysetool Sociale Veiligheid
30	Flowerbed				15-2-2022	V220167	71000 BD10 D057	202110-07	CDD realisatie/CHB check en KV8 aflevercheck gemaakt
31	Intraffic				16-2-2022	115	3500_000043		Hosting OV data systeem en JAVA DRIS
32	Intraffic				16-2-2022	116	3500_000075		GOVI-ICT Beheer en onderhoud 2021-2022
33	Intraffic				16-2-2022	117	3500_000076		OV data fase 2, 2e termijn
34	Intraffic				16-2-2022	118	3500_000074		GOVI hosting 2021-2022
35	Intraffic				16-2-2022	121	3500_000074		GOVI hosting 2021-2022

Hoofdstuk #3 Welke datasets zijn er bij OV-data?

Koppelvlak	Toelichting	Van	Naar	Standaard
PSA	Haltenummer	Samenwerkingsverband DOVA	Samenwerkingsverband DOVA	https://bison.SamenwerkingsverbandDOVA.nu/standaarden/psa-passengerstopassignment
NeTEx	Dienstregeling	Vervoerders	Samenwerkingsverband DOVA	https://bison.SamenwerkingsverbandDOVA.nu/standaarden/nederlands-netex-profiel
KV6	Actuele ritpunctualiteit	Vervoerders	Samenwerkingsverband DOVA	https://bison.SamenwerkingsverbandDOVA.nu/standaarden/kv6-actuele-ritpunctualiteit-en-voertuiginformatie
KV15	Vrije teksten	Vervoerders	Samenwerkingsverband DOVA	https://bison.SamenwerkingsverbandDOVA.nu/standaarden/kv15-haltegerelateerde-mededelingen-en-vrije-teksten
KV17	Mutaties in operatie	Vervoerders	Samenwerkingsverband DOVA	https://bison.SamenwerkingsverbandDOVA.nu/standaarden/kv17-mutaties-op-het-operationeel-proces
KV7	Planning per halte	Samenwerkingsverband DOVA	DRIS-leveranciers	https://bison.SamenwerkingsverbandDOVA.nu/standaarden/kv7-8-geplande-en-actuele-reisinformatie-op-halteniveau
KV8	Actuele reisinformatie per halte	Samenwerkingsverband DOVA	DRIS-leveranciers	https://bison.SamenwerkingsverbandDOVA.nu/standaarden/kv7-8-geplande-en-actuele-reisinformatie-op-halteniveau
KV7/8 turbo	Planning en actuele reisinformatie	Samenwerkingsverband DOVA	NDOV-loketten	https://bison.SamenwerkingsverbandDOVA.nu/standaarden/kv78turbo-effici%C3%ABnte-reisinformatie-op-halteniveau



Home

Nieuws

Agenda

Over BISON

BISON

Het platform BISON heeft als functie het opstellen, beheren, harmoniseren en bewaken van alle informatiestandaarden die de informatie-uitwisseling binnen het openbaar vervoer faciliteren. BISON fungeert eveneens als kenniscentrum voor het vastleggen van IT-beleid en voornemens in ondermeer concessies. In BISON zijn zowel provincies, vervoerregios, rijksoverheden, infrabeheerders, vervoerders, marktpartijen als reizigers vertegenwoordigd.

Hoofdstuk #3 Voorbeeld van Dienstregeling (KV1)



Versie: 8.3.0.2

Datum: 8 januari 2020

Status: release

Bestand: tmi8 dienstregeling (kv 1) v8.3.0.2, release.docx

© Platform Beheer Informatie Standaarden OV Nederland (BISON), 2020

Op dit werk is de Creative Commons Licentie/ by-nd/3.0/nl van toepassing.
<http://creativecommons.org/licenses/by-nd/3.0/nl/>

Specificatie TMI8

Dienstregeling *Koppelvlak 1*

2.5.10 Dataset levering en bestandsnamen

Een dienstregeling dataset wordt aangeleverd als een aantal bestanden. De bestanden bevatten de inhoudelijke planningsdata. De opbouw van (delen van) de bestandsnamen kunnen tussen vervoerder(s) en integratoren in onderling overleg worden afgestemd.

Hoofdstuk #3 Voorbeeld van Dienstregeling (KV1)

2 KV 1: Planning dienstregeling

2.1 Datamodel

Het datamodel planning bevat de volgende entiteiten:

De beschrijving van de verschillende attributen per object/record zijn vastgelegd in afzonderlijke .DEF files. De volgende recordtypes worden gebruikt:

Object	Naam	Omschrijving
Organisatie-eenheden (lijnen, vestigingen) van de aangeleverde gegevens		
ORUN	Organizational Unit	Een eenheid binnen een openbaar vervoerbedrijf die een bepaald aantal verantwoordelijkheden heeft met betrekking tot de exploitatie. <i>Voor de publieksdienstregeling wordt een lijn als kleinste organisatie-eenheid gekozen</i>
ORUNORUN <i>Niet verplicht</i>	Hogere organisatie eenheid	De in hiërarchie hoger liggende organisatie eenheid, t.b.v. o.a. vastleggen van (afwijkende) geldigheden. <i>niet verplicht</i>
Concessiegegevens		
CONFINREL	Concessie financier relatie (veelal perceel).	Kleinste eenheid waarvoor gegevens van een concessie worden vastgelegd in relatie tot de opdrachtgevende OV autoriteit.
FINANCER <i>Niet verplicht</i>	Financier	Financier van het perceel
CONAREA	Concession Area	Concessie gebied

Netwerkgegevens		
USRSTAR	STOP AREA = Knooppunt volgens codering vervoerder	Een StopArea is een bundeling van haltes die voor de reiziger eenzelfde naam hebben en logisch bij elkaar horen. (bijv. een busstation of overstappunt). Ook tegen over elkaar liggende haltes kunnen een StopArea vormen. <i>T.b.v eenduidige naamgeving en presentatie naar reizigers.</i> <i>Alle Planpunten kunnen de minimale set StopAreas vormen.</i> <i>Niet alle haltes worden in een STAR aangeleverd.</i>
USRSTOP	Halte/tijdpunt volgens codering vervoerder	Halte of passagepunt waarvoor tijden worden vastgelegd in het planningsysteem van de vervoerder. Coördinaten van een User Stop Point worden vastgelegd als Point. Bij de keuze van UserStopPoints is het van belang dat de coördinaten eenduidig en verifieerbaar worden vastgelegd. Een USRSTOP heeft een UserStopType; hiervoor zijn drie typen mogelijk: PASSENGER, BRIDGE of FINANCIAL. Bij een passagiershalte de locatie van de haltepaal (of als deze er niet is de kop van de halte waar de voorkant van de bus normaliter stopt), bij een brug de stopstreep van de meest rechter rechtdoor strook voor de brug (per richting).

Hoofdstuk #3 Voorbeeld van Dienstregeling (KV1)

2.6 Tabel definities

2.6.1 Interpretatie van de beschrijving objecten en attributen

Hieronder worden de verschillende recordtypes beschreven die in de interface worden gebruikt. De tabellen moeten als volgt worden geïnterpreteerd:

Naam	Omschrijving
Attribuut	korte omschrijving van het veld
Aard	x = verplicht (header), # = sleutelveld (verplicht), + = verplicht attribuut, o = optioneel attribuut/relatie
Type	A = Alfnumeriek, N = Numeriek
Lengte	Aantal posities
Geldige waarden/vulling	Omschrijving/toegestane waarden. Voor domeinspecifieke types (in hoofdletters, eindigend op TYPE), zie "BISON Enumeraties en Tabellen" document.

Omschrijving van enkele gebruikte typen:

Type	Omschrijving
Boolean	De waarde '1' is gelijk aan waarde 'true', '0' is gelijk aan 'false'. Toegestane waardes zijn: 0,1, false, true
Datum	Alfanumerieke string van max 10 karakters. Toegestaan formaat: 'YYYY-MM-DD'
Tijd	Alfanumerieke string van max 8 karakters. Toegestaan formaat: 'HH:MM:SS'. Geldige tijdstippen bevinden zich tussen 00:00:00 en 31:59:59u

2.6.2 Stamgegevens

Die gegevens van de planning welke gemeenschappelijke zijn voor alle varianten.

Organizational Unit Een verzameling van ritten met dezelfde geldigheidskenmerken. Een organisatorische eenheid kan deel uitmaken van een 'hogere' eenheid.
Een organisatorische eenheid wordt gedefinieerd als de eenheid waarvoor een planning van ritten wordt samengesteld. Bij het definiëren van organisatorische eenheden is het van belang dat alle ritten binnen het pakket een homogene geldigheid hebben (schoolvakanties, koopzondagen, buitenlandse feestdagen).

Attribuut	Aard	Type	Lengte	Geldige waarden / vulling:
Recordtype	x	A	10	ORUN
Version number	x	N	2	1
Implicit/Explicit	x	A	1	I

Attribuut	Aard	Type	Lengte	Geldige waarden / vulling:
DataOwnerCode	#	A	10	Vervoerder (E1)
OrganizationalUnitCode	#	A	10	
Name	+	A	50	
OrganizationalUnitType	+	A	10	Volgens ORUNTYPE
Description	o	A	255	

Tabel 1 Organization Unit

Hoofdstuk #3 Voorbeeld van Dienstregeling (KV1)

DATASET KV1 EBS (Vervoerder) Waterland (Concessie)

roseboom 15 - 172.16.14.1 - verbinding met extern bureaublad

IZArc - C:\Users\m.vanvuure\Downloads\EBS.20221210-20230107.WTL.zip

File View Actions Tools Options Help

New Open Add Extract Test Exit

C:\Users\m.vanvuure\Downloads\EBS.20221210-20230107.WTL.zip

File Name	File Type	Modified	Size	Ratio	Packed	CRC	Attributes	Folder
CONAREAXXX.TMI	TMI-bestand	17-11-2022 14:23:58	162	22%	127	E5436632	A	
CONFINRELX.TMI	TMI-bestand	17-11-2022 14:23:58	150	23%	115	3776CFB2	A	
DESTXXXXXX.TMI	TMI-bestand	17-11-2022 14:23:58	6,221	83%	1,038	68F2F302	A	
FINANCERXX.TMI	TMI-bestand	17-11-2022 14:23:58	128	12%	113	4AA475EF	A	
JOPATILDXX.TMI	TMI-bestand	17-11-2022 14:23:58	336,713	92%	28,165	888AD881	A	
JOPAXXXXXX.TMI	TMI-bestand	17-11-2022 14:23:58	5,302	86%	733	42118F81	A	
LINEXXXXXX.TMI	TMI-bestand	17-11-2022 14:23:02	2,349	73%	629	F1C3764E	A	
LINXXXXXX.TMI	TMI-bestand	17-11-2022 14:23:58	51,635	81%	9,623	196DAF60	A	
NOTICEXXXX.TMI	TMI-bestand	17-11-2022 14:23:58	214	26%	159	797592AE	A	
NTCASSGNMX.TMI	TMI-bestand	17-11-2022 14:23:58	5,776	88%	671	B01C2BC9	A	
OPERDAYXXX.TMI	TMI-bestand	17-11-2022 14:23:02	34,957	93%	2,468	A68192DF	A	
ORLUNXXXXX.TMI	TMI-bestand	17-11-2022 14:23:02	1,957	76%	470	BACF187D	A	
POINTXXXXX.TMI	TMI-bestand	17-11-2022 14:23:58	68,439	79%	14,327	25D0D462	A	
POOLXXXXXX.TMI	TMI-bestand	17-11-2022 14:23:58	137,397	87%	17,767	DE8E780E	A	
PUJOPASSXX.TMI	TMI-bestand	17-11-2022 14:23:58	36,575,331	92%	2,909,573	A2172A18	A	
SCHEDVERSX.TMI	TMI-bestand	17-11-2022 14:23:02	16,454	94%	984	D4711200	A	
TILDXXXXXX.TMI	TMI-bestand	17-11-2022 14:23:58	43,920	82%	8,009	C8C92E6A	A	
USRSTARXXX.TMI	TMI-bestand	17-11-2022 14:23:58	18,560	79%	3,989	C887D46A	A	
USRSTOPXXX.TMI	TMI-bestand	17-11-2022 14:23:58	107,816	87%	14,212	D12B2E81	A	

Hoofdstuk #3 Voorbeeld van Databestand Drukte indicator

CONCEPT koppelvlak "Bezetting" (versie 1)

IN ONTWIKKELING

juli 2020



Datastructuur

Het koppelvlak "Bezetting" (versie 1) omvat de volgende elementen:

Attribuut	Aard	Type	Lengte	Geldige waarden / vulling
DataOwnerCode	#	A	10	Vervoerder (BISON Domeinwaarden, Tabel E1)
OperatingDay	#	D	10	YYYY-MM-DD
LinePlanningNumber	O	A	10	Uniek systeemlijnummer binnen domein van DataOwner Verplicht bij verwijzing naar KV1.LinePlanningNumber
JourneyNumber	#	N	8	[0 - 99999999] Verwijzing naar KV1.JourneyNumber (BTM) Verwijzing naar IFF.Treinnummer (conform ritinfo Trein)
ReinforcementNumber	#	N	2	Versterkingsrit indicator, 0 = geplande rit, >0 = versterkingsrit Verplicht voor BTM
TimingLinkOrder	#	N	3	Volgnummer van de verbinding binnen het ritpatroon
UserStopCodeBegin	+	A	10	Van halte/station van de verbinding (Timinglink)
UserStopCodeEnd	+	A	10	Tot halte/station van de verbinding (Timinglink)
Occupancy	+	N	1	Verwachte bezetting, enumeratie (zie onder)
VehicleType	O	A	20	Soort materieel, bijv. "SLT" Verplicht voor trein (i.c.m. TotalNumberOfCoaches)
TotalNumberOfCoaches	O	N	2	Totaal aantal rijtuigen (bakken) van het in VehicleType vermelde soort. Verplicht voor trein (i.c.m. VehicleType)

Legenda:

Kolom	Omschrijving
Attribuut	Korte omschrijving van het veld
Aard	X = verplicht (header), # = sleutelveld (verplicht), + = verplicht attribuut, O = optioneel attribuut/relatie
Type	A = alfanumeriek, N = numeriek, D = datum
Lengte	Aantal posities
Geldige waarden / vulling	Omschrijving / toegestane waarden. Voor domeinspecifieke types, zie het "BISON Enumeraties en Tabellen" document.

Materieelsamenstelling

De verwachte bezetting heeft betrekking op de *geplande* materieelsamenstelling (materieelsoort + aantal bakken), zoals beschreven in *VehicleType* en *TotalNumberOfCoaches*. Als de materieelsamenstelling in de uitvoering anders is dan het materieel dat is gebruikt om de verwachte bezetting uit te rekenen, geldt de waarde die volgens dit koppelvlak is geleverd niet meer.

De samenstelling in de uitvoering kan worden berekend m.b.v. een **vertaaltabel**, die als los bestand bij elke levering van het koppelvlak wordt meegeleverd, indien *VehicleType* en *TotalNumberOfCoaches* zijn gevuld. Dit bestand heeft de suffix "_RS" in de bestandsnaam en bevat de volgende elementen:

Attribuut	Aard	Type	Lengte	Geldige waarden / vulling
DataOwnerCode	#	A	10	Vervoerder (BISON Domeinwaarden, Tabel E1)
VehicleType	#	A	20	Soort materieel, bijv. "SLT" <i>Conform ritinfo.materieeldeel.materieelsoort (Trein)</i>
VehicleSubType	#	A	20	Aanduiding materieel, bijv. "6" of "4SA" <i>Conform ritinfo.materieeldeel.materieelaanduiding (Trein)</i>
NumberOfCoaches	X	N	2	Aantal rijtuigen (bakken) van het in <i>VehicleType</i> en <i>VehicleSubType</i> vermelde soort materieel.

Classificatie van de bezetting

De bezetting wordt weergegeven conform onderstaande enumeratie. Deze definieert bezetting zoals de reiziger deze beleeft (kan ik vrij zitplaatsen kiezen, is er ruimte naast/tegenover iemand, is de kans groot dat ik moet staan?). Dit wordt vertaald in een percentage vrije of bezette zitplaatsen.

Een vervoerder mag (in eerste instantie) zelf bepalen welk percentage vrije zitplaatsen als grenswaarden worden gehanteerd voor de classificaties. Ter illustratie is de classificering die NS hanteert toegevoegd.

Nummer	Enumeratie	Opmerkingen / interpretatie NS
0	No information	Geen informatie / onbekend. Deze waarde is t.b.v. toekomstig gebruik om de bezetting aan te vullen met real-time gegevens.
1	Empty	<i>Bij NS: Waarden 1 & 2 worden samengevoegd tot 2</i>
2	Many seats available	<i>Bij NS: <65% van zitplaatsen bezet</i>
3	Few seats available	<i>Bij NS: >65% en <100% van de zitplaatsen bezet</i>
4	Standing room only	<i>Bij NS: >100% van de zitplaatsen bezet</i>
5	Full	Deze waarde is alleen relevant voor real-time gegevens. Indien de verwachte bezetting 'vol' is, wordt extra capaciteit ingezet.

HoofdstuHoofdstuk #3 Voorbeeld van Databestand

CONCEPT koppelvlak "Bezetting" (versie 1)

IN ONTWIKKELING

juli 2020



Leveringsproces

- Er worden voor iedere dag verwachte bezettingsgegevens geleverd conform deze KV specificatie.
- De layout komt overeen met de voorbeelden in de bijlage.
- Uiterlijk om 20.00 uur dienen de gegevens van minimaal de 2 eerstvolgende operationele dagen bij de ontvangers (NDOV loketten en DOVA) beschikbaar te zijn.
- Na 20:00 uur en ook gedurende de operationele dag mogen leveringen met actuelere prognoses worden aangeleverd. Deze leveringen dienen minimaal 3 opeenvolgende operationele dagen te bevatten (de huidige dag + 2 opvolgende dagen).
- Een nieuwe levering overschrijft de vorige levering (voor operationele dagen die al eerder zijn geleverd).
- Een levering bevat altijd alle ritten van een operationele dag. Bij levering gedurende een operationele dag worden ook reeds uitgevoerde ritten alsnog aangeleverd.
- De naam van het bestand bevat de dataownercode en de datum van de eerste operationele dag.
- De bestanden worden aangeleverd in gzip formaat.

Bestandsnaam

De bestandsnaam is als volgt samengesteld: `OC_dataownercode_YYYYMMDD.csv`

De naam van het bijbehorende vertaaltabel is: `OC_dataownercode_YYYYMMDD_RS.csv`

Na zippen: `OC_dataownercode_YYYYMMDD.csv.gz` en `OC_dataownercode_YYYYMMDD_RS.csv.gz`

Dataownercode is de identificatie van het domein van de vervoerder (BISON enumeratie Tabel E1).

YYYYMMDD is de datum van de eerste operationele dag in het bestand.

Hoofdstuk #3 Welke datasets zijn er bij OV-data?

Voorbeelden

Voorbeeld BTM

Arriva levert de verwachte bezetting voor de eerstvolgende 3 operationele dagen.

Bestand `OC_ARR_20200708.csv` bevat de verwachte bezetting voor 8 juli, 9 juli en 10 juli (alleen de eerste regels worden getoond):

```
DataOwnerCode,OperatingDay,LinePlanningNumber,JourneyNumber,ReinforcementNumber,TimingLinkKorder,UserStopCodeBegin,UserStopCodeEnd,Occupancy,VehicleType,TotalNumberOfCoaches
ARR,2020-07-08,11401,1003,0,1,10009024,13908210,1,,
ARR,2020-07-08,11401,1004,0,1,13908210,10009024,1,,
ARR,2020-07-08,11401,1007,0,1,10009024,13908210,1,,
ARR,2020-07-08,11401,1008,0,1,13908210,10009024,1,,
ARR,2020-07-08,11401,1011,0,1,10009024,13908210,1,,
ARR,2020-07-08,11401,1012,0,1,13908210,10009024,1,,
ARR,2020-07-08,11401,1015,0,1,10009024,13908210,1,,
ARR,2020-07-08,11401,1016,0,1,13908210,10009024,1,,
ARR,2020-07-08,11401,1019,0,1,10009024,13908210,1,,
ARR,2020-07-08,11401,1020,0,1,13908210,10009024,1,,
ARR,2020-07-08,11401,1023,0,1,10009024,13908210,1,,
ARR,2020-07-08,11401,1024,0,1,13908210,10009024,1,,
ARR,2020-07-08,11401,1027,0,1,10009024,13908210,1,,
ARR,2020-07-08,11401,1028,0,1,13908210,10009024,1,,
ARR,2020-07-08,11401,1031,0,1,10009024,13908210,1,,
ARR,2020-07-08,11401,1032,0,1,13908210,10009024,1,,
ARR,2020-07-08,15020,8003,0,1,53603012,53553010,1,,
ARR,2020-07-08,15020,8003,0,2,53553010,53403010,1,,
ARR,2020-07-08,15020,8003,0,3,53403010,53443010,1,,
ARR,2020-07-08,15020,8003,0,4,53443010,53443110,2,,
ARR,2020-07-08,15020,8003,0,5,53343110,53223010,2,,
ARR,2020-07-08,15020,8003,0,6,53223010,53003010,1,,
ARR,2020-07-08,15020,8004,0,1,53443020,53403020,1,,
ARR,2020-07-08,15020,8004,0,2,53403020,53553020,1,,
ARR,2020-07-08,15020,8004,0,3,53553020,53603022,1,,
ARR,2020-07-08,15020,8007,0,1,53603012,53553010,1,,
ARR,2020-07-08,15020,8007,0,2,53553010,53403010,1,,
ARR,2020-07-08,15020,8007,0,3,53403010,53443010,1
```

Voorbeeld Trein

NS levert de verwachte bezetting voor de eerstvolgende 9 operationele dagen.

Bestand `OC_NS_20200709.csv` bevat de verwachte bezetting van NS voor 9 t/m 17 juli (alleen de eerste regels worden getoond):

```
DataOwnerCode,OperatingDay,LinePlanningNumber,JourneyNumber,ReinforcementNumber,TimingLinkKorder,UserStopCodeBegin,UserStopCodeEnd,Occupancy,VehicleType,TotalNumberOfCoaches
NS,2020-07-09,,6936,0,1,HT,ZDM,1,SLT,10
NS,2020-07-09,,6936,0,2,2EM,GDM,1,SLT,10
```

Bestand `OC_NS_20200709_RS.csv` bevat de gegevens om materieel om te rekenen naar aantal rijtuigen (bakken) voor dezelfde periode:

```
DataOwnerCode,VehicleType,VehicleSubType,NumberOfCoaches
NS,SLT,6,6
NS,VIRM,4,4
NS,SW7-25KV,2+7,7
NS,DDE,43R,4
```

Hoofdstuk #4 Hoe maakt je van data chocolade?

RUTGER
Bakt

Recepten

Bakhulp

Inspiratie

Feestdagen

Boeken

Bakcursussen

Over Rutger



Zelf chocolade maken – van cacaoboon tot chocoladereep

19 november 2018



Hoofdstuk #4 Hoe maak je van data chocolade?

Recept voor chocolade taart a la OV-data

Ingrediënten

- Zelfrijzend bakmeel
- Pure chocolade
- Witte basterd suiker
- Ongezouten roomboter
- Gecondenseerde volle melk
- Kraanwater
- Eieren
- Melkchocolade
- Slagroom

Ingrediënten

- xlsx files
- csv files
- JSON files
- XML files
- shape files
- META data

Benodigheden; springvorm, pan, mixer, oven en geduld juiste tools en mensen (kennis en capaciteit)

Hoofdstuk #4 Hoe maak je van data chocolade?

Recept voor chocoladetaart met slagroom à la OV-data Recept voor reisinformatie à la OV-data

- 1) Verwarm de oven, vet bakvorm in en hak chocolade grof
 - 2) Doe suiker, boter, ½ chocolade, melk en kokend water in pan. Verwarm al roerend op laag vuur tot gladde massa.
 - 3) Voeg toe aan bakmeel in de kom.
 - 4) Klop de eieren los en voeg ze toe en roer ze erdoor
 - 5) Schenk het mengsel in de springvorm
 - 6) Bak circa 55 minuten in de oven
 - 7) Kan ondertussen de melkchocolade fijn en doe met ½ van pure chocolade in een kom
 - 8) Klop slagroom met de mixer
 - 9) Neem taart uit de oven en laat afkoelen in 90 minuten
 - 10) Snijd taart door de midden en bestrijk bovenkant beide helften met chocolademengsel en leg weer op elkaar.
- 1) beschik over juiste tools = computerprogramma's en kennis
 - 2) Check de data kwaliteit, voldoet dit aan de (gestelde) eisen?
 - 3) Zo niet, schoon of bewerk dan eerst de (input) data
 - 4) Zet waar nodig data om in het juiste formaat voor import
 - 5) Importeer de data in de tool (of DOVA server)
 - 6) Binnen enkele seconden heeft de DOVA server de bestanden (KV1, KV6, KV15 en KV17) omgezet in real-time reisinformatie (KV7/8)
 - 7) Deze real-time reisinformatie wordt naar DRIS displays verstuurd bij Bus, Tram en Metro haltes
 - 8) Deze real-time informatie wordt naar de NDOV loketten gestuurd en is beschikbaar als OPEN DATA; reisplanners

Plannen & Boeken – Betalen – Reizen maar ;)

SMULLEN MAAR

Hoofdstuk #4 Hoe maak je van data chocolade... (d)uh informatie?

Stap 1: Gegevensverzameling

Verschillende vormen;

Vragenlijst (papier, digitaal, nieuwe innovaties en technieken)

Interview (telefonisch, face-to-face)

Observeren (turven, camera en algoritmes,)

Sensor meetings (boardcomputers, OV-chipkaart, lus data, iVRI's, KV6, tracking,)

Register data

Stap 2: Gegevensverzameling komt in Databestand

Databestand heeft Metadata en Lexicon (vertaling getal/klasse naar taal)

Stap 3: bewerken data

Transformeren data (bewerken en verrijken) → **Informatie** → **Inzicht** (kaart, grafiek, tabel en dashboards)

Hoofdstuk #4 Hoe maak je van data chocolade... (d)uh informatie?

The screenshot shows a journey planner interface with a yellow header. The origin is 'Eindhoven Centraal' and the destination is 'Heerlen'. The date is 'do 24 nov' and the time is '08:00'. A 'Plannen' button is visible. Below the header, there are options for 'Vertrek', 'Aankomst', 'Nu', and 'Reisopties'. A purple tooltip says 'Kies met welk vervoer je reist.' The main area displays several train options, with the 08:08 → 09:11 IC Intercity option highlighted. A detailed view of this option shows the route: Eindhoven Centraal (08:08) to Heerlen (09:11) via Spoor 1 and Spoor 5. The cost is €10,90. A tilted card in the foreground shows a list of train segments with their departure and arrival times and icons for services like IC Intercity, NS Intercity, and Arriva Stoptrein.

A vertical timeline showing a journey from Utrecht to Diemen. The timeline starts at 09:25 with 'Vertrek Utrecht' and a walking icon for 'Lopen' (1 minuut). At 09:26, there is an 'Aankomst' at 'Treinstation Utrecht Centraal' (Vertrek: Spoor 5) and a train icon for 'NS Intercity' (Richting Amsterdam RAI). At 09:42, there is an 'Aankomst' at 'Spoor 2' and a 'Vertrek' at 'Treinstation Amsterdam Bijlmer Arena'. At 09:42, there is a walking icon for 'Lopen' (4 minuten). At 09:46, there is an 'Aankomst' at 'Busstation Station Bijlmer Arena, Amsterdam'. At 09:47, there is a 'Vertrek' at 'Perron E' and a bus icon for 'GVB Bus 44' (Richting Diemen Noord). At 10:05, there is an 'Aankomst' at 'Bushalte Station Diemen, Diemen'.

Hoofdstuk #5 Hoe gebruik je informatie voor Beleid & Kennis?

Waarom is er data nodig?

Inzichten

Antwoorden op beleidsvragen (OV-Netwerk)

Antwoorden op kennisvragen (CROW-KpVV)

Inzichten in trends (onder andere Staat van het OV van CROW)

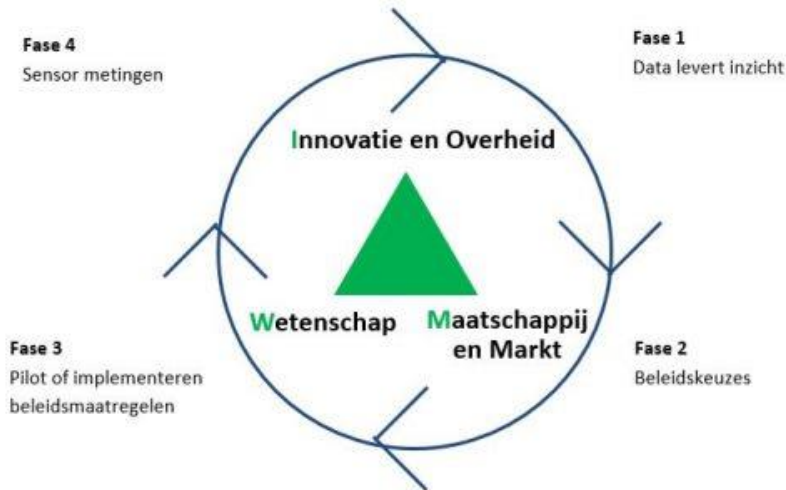
Modelleren en simuleren van scenario's met modellen

Voor het nemen van besluiten (Rijksoverheid, Decentrale OV Autoriteiten, Regio's en Gemeentes en Politiek)

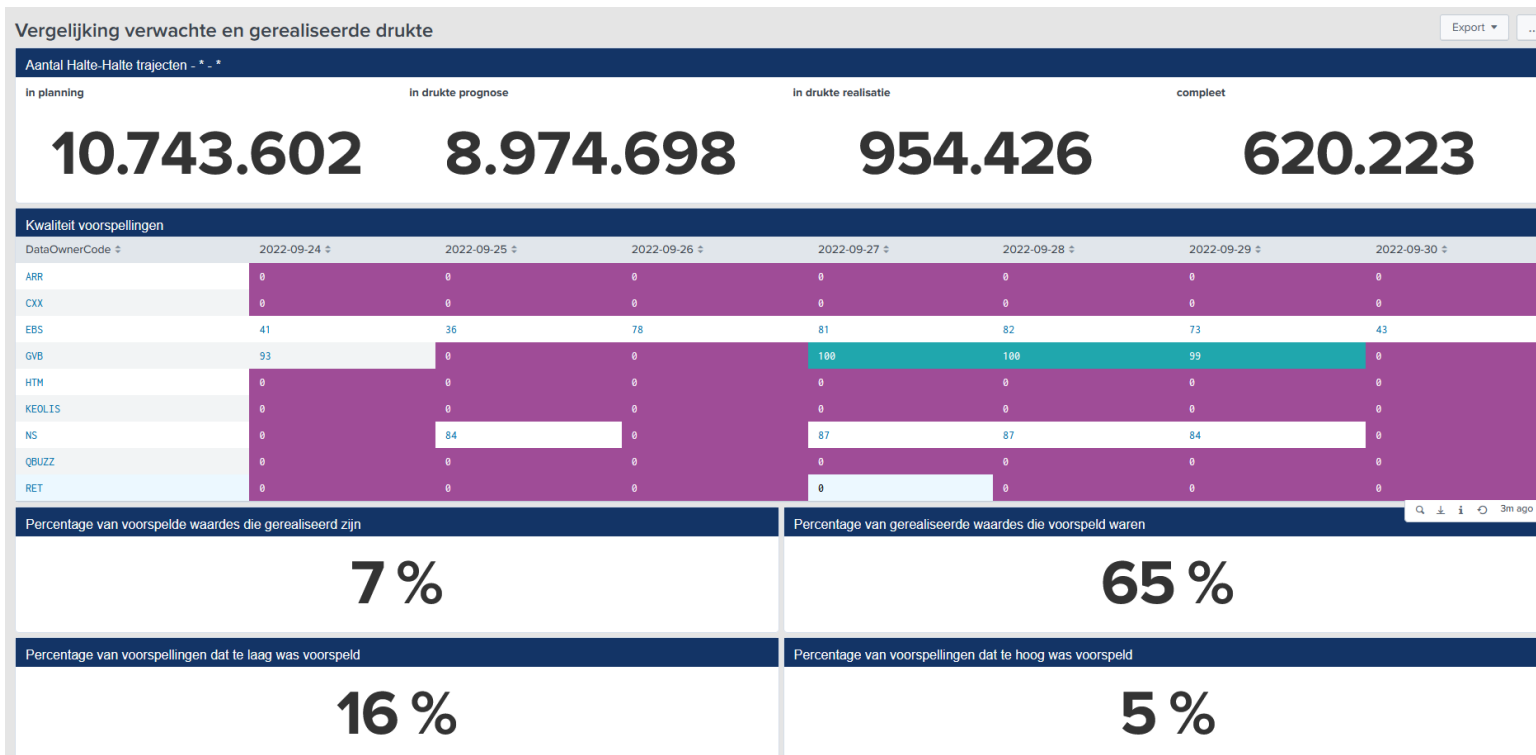
METEN is WETEN ;)

Hoofdstuk #5 Hoe gebruik je informatie voor Beleid & Kennis?

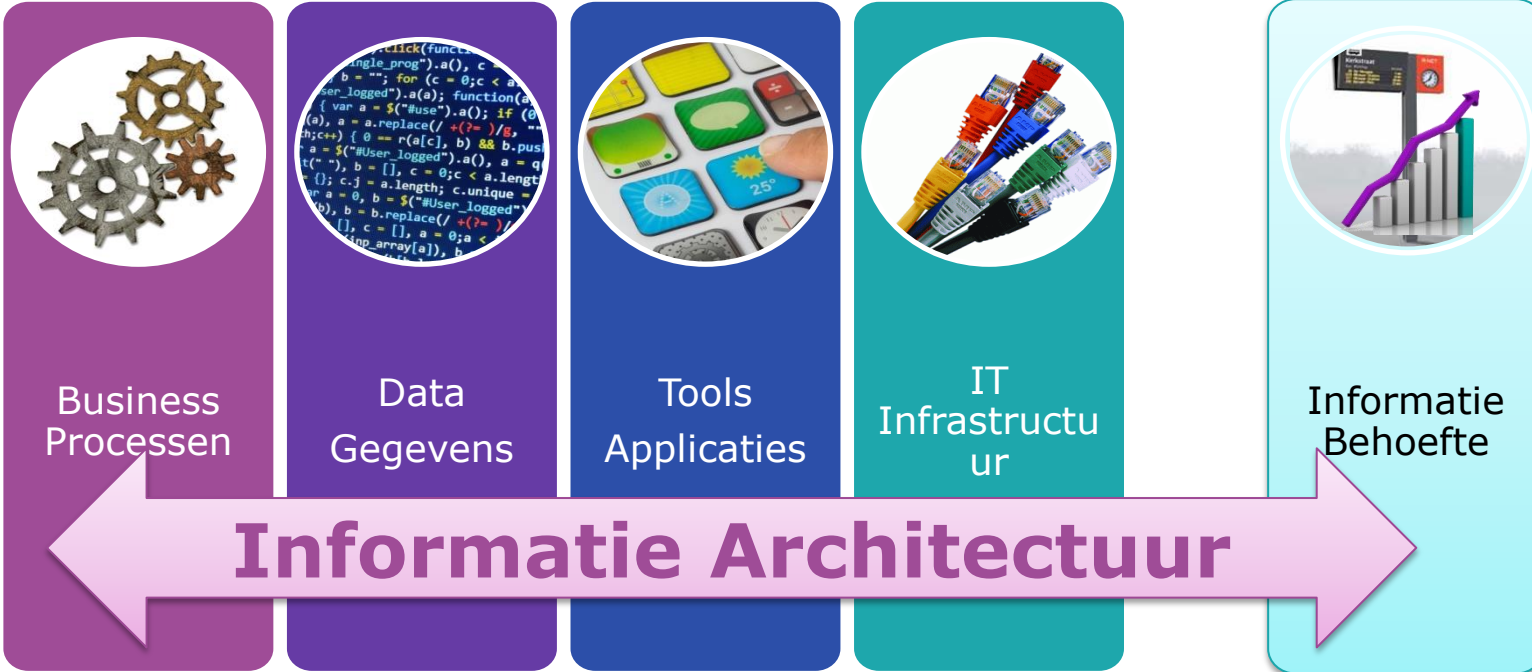
Conclusie: WIM en 4 fases



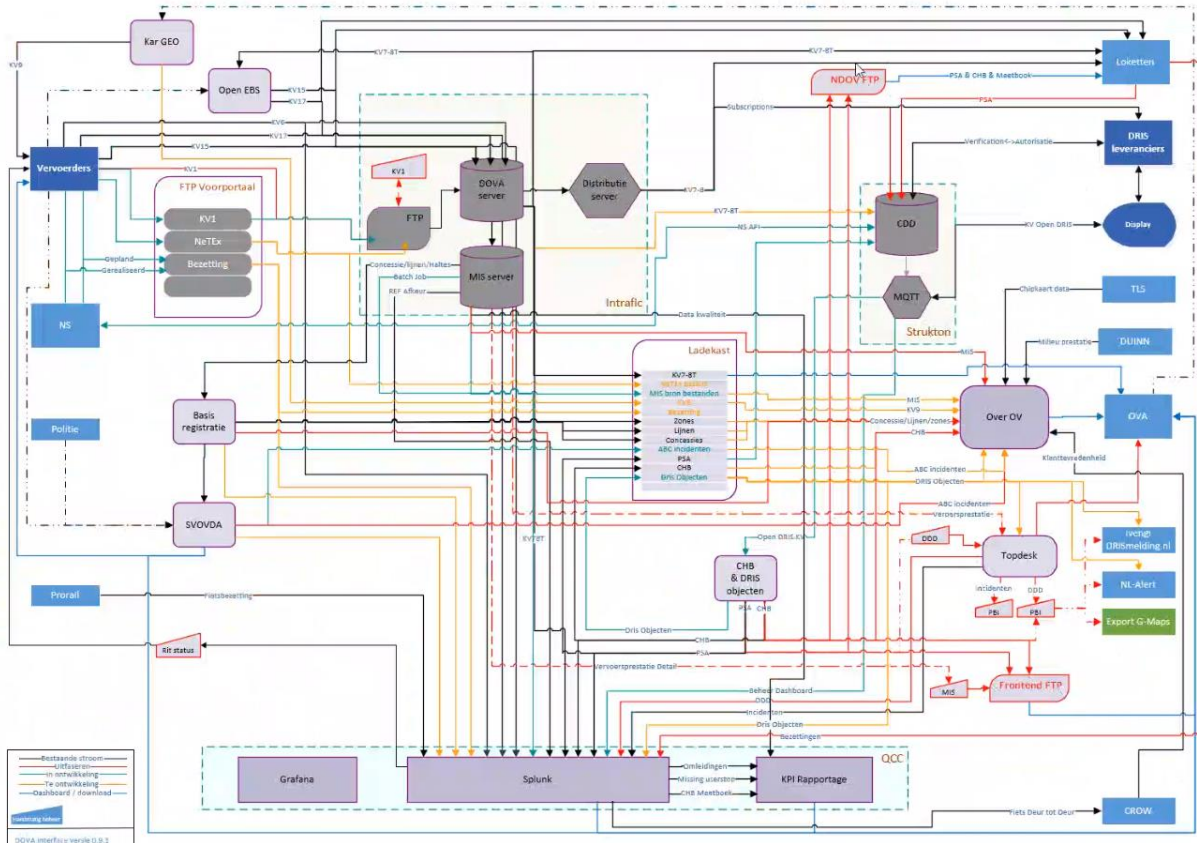
Hoofdstuk #5 Hoe gebruik je informatie voor Beleid & Kennis?



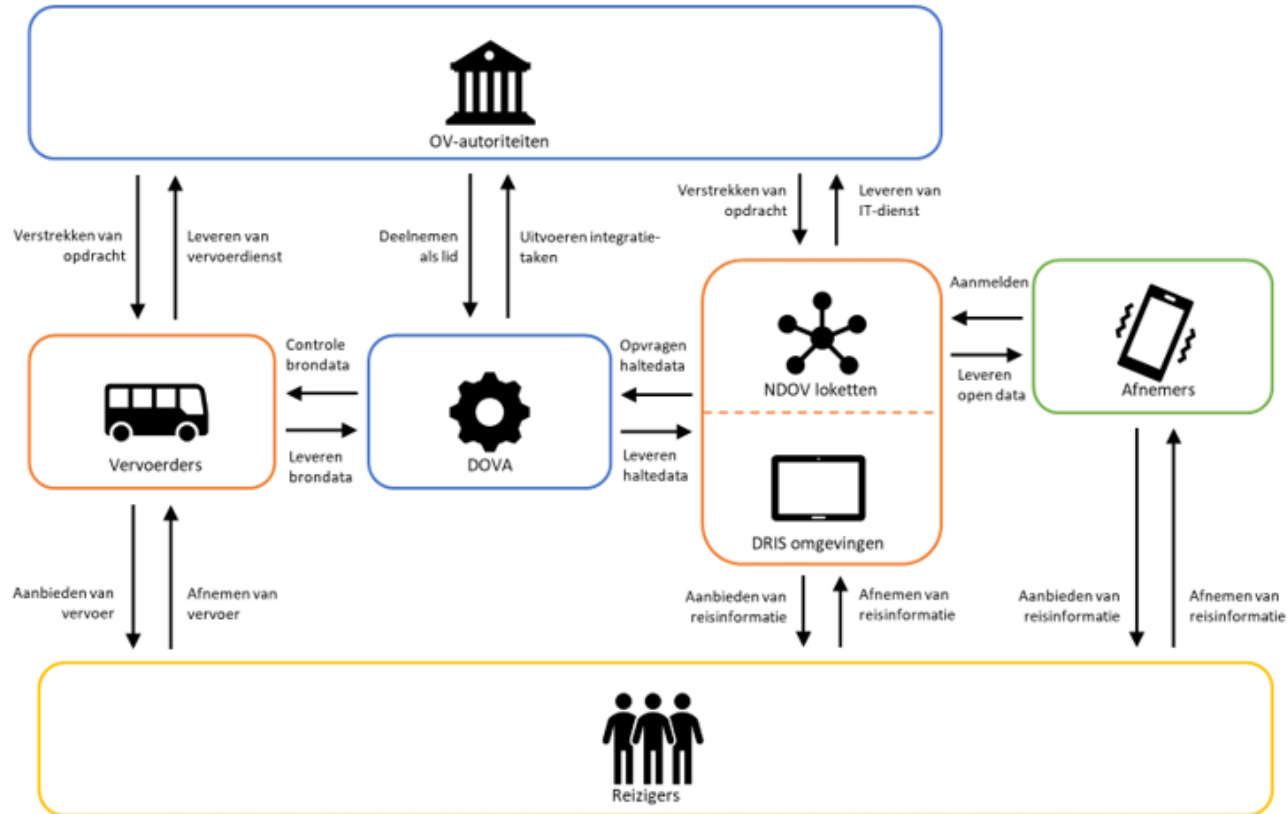
Hoofdstuk #6 Benodigdheden



Hoofdstuk #6 Benodigdheden: Data architectuur (in ontwikkeling)



Hoofdstuk #6 Benodigheden: Reisinformatieketen en actoren



Figuur 1. Schematisch overzicht realiseren reisinformatie.



Samen beter OV

Discussie en vragen?

de expert

= iemand die ergens veel van weet



De expert vertelde dat je onder water prachtige dingen kan zien.

de leek

= iemand die niet deskundig is op een bepaald gebied



De leek wist niet wat je onder water kon zien.

